

Volume 19
2012

Journal
of
Natural Sciences
Sookmyung
Women's
University

Nano-physics
Chemistry
Medical & Pharmaceutical
Mathematics
Statistics
Computer Science
Multimedia Science



Published by
The Research Institute of Natural Sciences
Sookmyung Women's University
Seoul 140-742, Korea

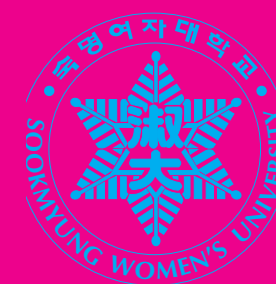
자연과학연구소 2012년도 연구 결과 백서

제 19 호 (2 0 1 2 년 호)

숙명여자대학교 자연과학연구소

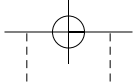
제19호
(2012년호)

자연과학연구소 2012년도 연구 결과 백서



나노물리학
화학
의약과학
수계학
통계학
컴퓨터과학
멀티미디어학

숙명여자대학교 자연과학연구소



자연과학연구소 2012년도 연구 결과 백서

편집위원회 위원장	황 선 영 숙명여자대학교 이과대학 통계학과	Editor-in-Chief	Sun-Young Hwang Department of Statistics Sookmyung Women's University
간사	김 선 홍 숙명여자대학교 이과대학 수학과	Excutive Editor	Seon-Hong Kim Department of Mathematics Sookmyung Women's University
위원	김 한 철 숙명여자대학교 이과대학 물리학과		Hanchul Kim Department of Nano-Physics Sookmyung Women's University
	이 한 길 숙명여자대학교 이과대학 화학학과		HanGil Lee Department of Chemistry Sookmyung Women's University
	장 민 선 숙명여자대학교 이과대학 의약학과		Minsun Chang Department of Medical & Pharmaceutical Science Sookmyung Women's University
	강 시 호 숙명여자대학교 이과대학 수학과		Si Ho Kang Department of Mathematics Sookmyung Women's University
	황 선 영 숙명여자대학교 이과대학 통계학과		Sun-Young Hwang Department of Statistics Sookmyung Women's University
	김 윤 희 숙명여자대학교 이과대학 컴퓨터학과		Yoonhee Kim Department of Computer Science Sookmyung Women's University
	박 화 진 숙명여자대학교 이과대학 멀티미디어학과		Hwajin Park Department of Multimedia Science Sookmyung Women's University

교육이념 우리 대학교의 창학 이념은 정숙, 현명, 정대를 지향하는 여성교육과 민족의 전통성 및 주체성을 확립하는 민족교육의 실현이다. 우리 대학교는 이러한 여성교육과 민족교육을 실현하기 위하여 시대가 요청하는 여성을 양성하고, 더 나아가 국가 민족과 세계 인류 발전에 기여하는 전인적이고 지도적인 여성을 배출하고자 한다.

자연과학연구소 2012년도 연구 결과 백서

2013년 2월 8일 인쇄

2013년 2월 10일 발행

발행처 : 숙명여자대학교 자연과학연구소

발행인 : 황 선 혜

편집인 : 황 선 영

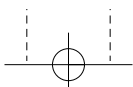
서울시 용산구 청파동 2가 53-2

인쇄소 : 한림원(주)

Tel : 02)2273-4201, FAX : 02)2179-9083

E-mail : hanrim@hanrimwon.com

◆ 숙명여자대학교 자연과학연구소 2012년도 연구 결과 백서는 숙명여자대학교 자연과학연구소에서 출판한 것이며, 아울러 본 논문집은 2012년도 숙명여자대학교 재정지원에 의하여 출판되었음.



자연과학연구소 주최 2012 학술행사

- [1] 2012.04.17. 멀티미디어과학과. 음성도서 전자책 표준 : DAISY (연사 : 주진규 박사).
- [2] 2012.04.17. 의약과학과. 국내·외 제약 산업 현황 및 신약개발 활성화를 위한 정부의 투자 현황 (연사 : 정재훈 박사).
- [3] 2012.04.28. 수학과. 2012 SPRING MEETING OF THE KOREAN MATHEMATICAL SOCIETY.
- [4] 2012.05.08. 의약과학과. 여성의 전문경력 관리를 위한 사람-사람 관계 형성의 중요성 (연사 : 박수정 박사).
- [5] 2012.05.09. 화학과. 이차원 적외선 분광학을 이용한 응축상에서의 화학교환 반응연구 (연사 : 곽경원 박사).
- [6] 2012.05.15. 멀티미디어과학과. DRM 이슈 (연사 : 강호갑 박사).
- [7] 2012.05.16. 화학과. APPLICATION OF NMR SPECTROSCOPY TO PROBE BIOMOLECULAR INTERACTIONS (연사 : 서정용 박사).
- [8] 2012.05.23. 화학과. EXPLORING NOVEL NANO (연사 : 함문호 박사).
- [9] 2012.05.29. 의약과학과. 바이오 의약 연구동향 및 미래 (연사 : 박명옥 박사).
- [10] 2012.05.30. 화학과. NEW DRUG DELIVERY SYSTEM: ORGAN AND CANCER SPECIFIC TARGETING FLUORESCENT PROBE (연사 : 김종승 박사).
- [11] 2012.06.01. 컴퓨터과학부. WIRELESS ELECTRICITY (WITRICITY) CHARGING PROTOCOL IN WIRELESS SENSOR NETWORKS (연사 : 이화정 박사).
- [12] 2012.06.22. 의약과학과. ENGINEERING BIOMATERIALS TO DIRECT STEM CELL FATE (연사 : Anthony Conway 박사).
- [13] 2012.07.17. 통계학과. Bayesian Analysis of Randomized Response SumScore Variables (연사 : 김종민 박사).
- [14] 2012.09.12. 화학과. THE DEVELOPMENT OF AMBIENT PRESSURE XPS AND ITS APPLICATION TO MODEL SYSTEMS IN SURFACE SCIENCE (연사 : 문봉진 박사).
- [15] 2012.09.18. 나노물리학과. 나노소자에서의 전기적 접합 (연사 : 김규태 박사).
- [16] 2012.09.19. 화학과. DEVELOPMENT OF SIGNAL-RESPONSIVE MATERIALS (연사 : 이연 박사).
- [17] 2012.10.16. 통계학과. MODEL DIAGNOSTICS IN THE PRESENCE OF INCOMPLETE OBSERVATIONS (연사 : Hira L. Koul 박사).
- [18] 2012.10.17. 통계학과. GOODNESS-OF-FIT TESTS FOR LONG MEMORY MOVING-AVERAGE MARGINAL DENSITY (연사 : Hira L. Koul 박사).
- [19] 2012.11.02. 의약과학과. 화장품산업에 있어서 바이오기술응용사례 및 향후전망 (연사 : 강병영 박사).
- [20] 2012.11.05. 컴퓨터과학부. IBM WEBSHPERE를 활용한 웹서비스응용개발기법 (연사 : 조성균 박사).

- [21] 2012.11.06. 나노물리학과. 금융수학과 복잡계 (연사 : 신용현 박사).
- [22] 2012.11.28. 컴퓨터과학부. INTELLIGENT ASSISTANT FOR SOCIAL DIALOGUE AND COLLABORATION (연사 : 김지희 박사).
- [23] 2012.12.04. 나노물리학과. Colloquium (연사 : 제원호 박사).
- [24] 2012.12.08. 멀티미디어과학과. 한국디지털콘텐츠학회 학술발표대회 (연사 : 홍상표 박사, 김하진 박사, 조성갑 박사).
- [25] 2012.12.17. 통계학과. Clustering Curves based on Changepoint analysis: A Nonparametric Bayesian Approach (연사 : 임채영 박사).

자연과학연구소 연혁

(1) 설립 목적 : 본 연구소는 자연과학의 기초 및 응용분야에 관한 연구를 수행함으로써 자연과학의 발전에 기여함을 목적으로 한다.

(2) 사업

1. 자연과학의 기초 및 응용분야의 연구, 개발 및 조사
2. 학술 강연회와 연구발표회 및 공개 강좌 개최
3. 연구자료의 확보 및 연구 논문집 발간
4. 위탁된 연구조사 및 감정 분석 실험
5. 기타 연구소 설립 목적에 부합되는 사업
6. 자연과학연구소 NEWSLETTER 발간

(3) 현황

숙명여자대학교는 21세기를 향한 자연과학분야의 기초 및 응용 분야에 관한 연구내용의 질적 향상을 도모하고 특성화 분야 연구의 중심적 역할을 수행하여 사회발전에 공헌하기 위하여 1985년 자연과학연구소를 설립하였다. 연구소의 활동 및 참여 연구원은 이과대학의 교수진 뿐 아니라 교내 외의 인사들이 대상이었고 현재 물리학 연구부, 화학 연구부, 의약과학 연구부, 수학 연구부, 컴퓨터과학 연구부, 통계학 및 멀티미디어학 연구부 등에서 전임 교원 및 석·박사 학위 소지자를 주축으로 활발한 연구활동을 수행하고 있다.

앞으로 예측되는 기초과학의 중요성에 비추어 볼 때 기초과학연구지원의 증대는 본 연구소의 연구수행 및 연구지원 활동에 보다 크게 기여할 것이다. 본 숙명여자대학교는 창학 100주년을 기점으로 새로운 대학으로 거듭나고자 학내 제반 분야에서의 개혁에 강한 의지를 표명하고 지속적으로 학교의 모든 분야에 대한 과감한 투자 및 개혁을 가속화 해 온 바 있다. 특히, 연구 분야에 대한 과감한 체제 개혁 및 투자를 한 결과, 행정 조직으로서 연구지원팀을 신설하여 연구활동 및 국내외 연구교류를 지원토록 하고, 교내 예산의 일부를 학술연구비로 할당 배분하고, 연구업적 평가제를 도입하여 모든 연구관련 자료를 전산화하고 있다. 또한, 본 연구소는 1997년에 기자재 확충계획을 수립하여 추진하고 있어 지난 수년 동안 교비 및 교육부 등의 약 20여 억원의 재정지원을 받아 최첨단 연구기자재를 확충하여 활발한 연구 활동이 진행 중에 있다.

역대 소장

1985년 9월 19일 창립

1985년 11월 - 1988년 10월 : 초대 소장 : 박 애 주 교수

1988년 11월 - 1990년 10월 : 2 대 소장 : 박 영 자 교수

1990년 11월 - 1992년 10월 : 3 대 소장 : 장 윤 경 교수

1992년 11월 - 1993년 10월 : 4 대 소장 : 박 영 자 교수

1993년 3월 - 1995년 2월 : 5 대 소장 : 이 장 로 교수

1995년 3월 - 1996년 9월 : 6 대 소장 : 오 성 담 교수

1996년 10월 - 2000년 9월 : 7 대 소장 : 민 경 희 교수

2000년 10월 - 2004년 10월 : 8 대 소장 : 노 광 현 교수

2004년 10월 - 2008년 8월 : 9 대 소장 : 김 재 성 교수

2008년 9월 - 2010년 8월 : 10대 소장 : 정 혁 교수

2010년 9월 - 2012년 8월 : 11대 소장 : 이 원 춘 교수

2012년 9월 - 현 재 : 12대 소장 : 황 선 영 교수

Contents

◆ 나노물리학과

Perpendicular Magnetic Anisotropy of Amorphous Ferromagnetic CoSiB/[Pt,Au] Multilayer H. I. Yim	3
Glass Formation and Magnetic Properties in the Co and Fe-based Amorphous Alloy with Small Mo or Nb Additions Haerin CHOI-YIM	5
Thermal and Elastic Properties of Cu-Hf-Al Bulk Metallic Glasses Haerin CHOI-YIM	7
Synthesis and Characterization of Co-Fe-based Amorphous Ribbons with Small Nb Additions Haerin CHOI-YIM	9
Initial Oxidation Structure of Chlorinated Si(001) Hanchul KIM	11
Van der Waals Interaction between P ₄ Molecules: Density Functional Theory Calculations with Dispersion Correction Hanchul KIM	13
Precursor State of Oxygen Molecules on the Si(001) Surface during the Initial Room-temperature Adsorption Hanchul KIM	15
Optical and Electronic Properties of Hydrogenated Silicon Nanoclusters and Nitrogen Passivated Silicon Nanoclusters: A Density Functional Theory Study Hanchul Kim	17
Theoretical investigation on the magnetic phase stability of Fe-doped Bi tellurides Miyoungh Kim	19
Surface and Correlation Effect on the Half-metallicity of the Co ₂ FeSi Heusler Alloy Miyoungh KIM	21
Temperature and Carrier-concentration Dependences of the Thermoelectric Properties of Bismuth Selenide Dioxide Compounds Miyoungh KIM	23

◆ 화학과

Atomic-level investigations on the amyloid- β dimerization process and its driving forces in water Sihyun Ham	27
Structural and Thermodynamic Characteristics That Seed Aggregation of Amyloid- β Protein in Water Sihyun Ham	29
Metal ion shuttling mechanism through thiacalix[4]crown: a computational study Sihyun Ham	31
Component analysis of the protein hydration entropy Sihyun Ham	33
Impact of chemical heterogeneity on protein self-assembly in water Sihyun Ham	35
Elucidating the Molecular Origin of Hydrolysis Energy of Pyrophosphate in Water Sihyun Ham	37
Aqueous interaction site integral-equation theory that exactly takes into account intramolecular correlations Sihyun Ham	39
Human IgA-binding Peptides Selected from Random Peptide Libraries: affinity maturation and application in IgA purification Sihyun Ham	41
Role of Chain Length in the Adsorption Structures and Geometric Configurations of Phenylalanine Derivatives on Ge(100) Surfaces Hangil Lee	43
Comparison and Contrast Analysis of Adsorption Geometries of Phenylalanine versus Tyrosine on Ge(100): Effect of Nucleophilic Group on the Surface Hangil Lee	45
Bioinspired Morphogenesis of Highly Intricate and Symmetric Silica Nanostructures Jin Seok Lee	47
In Vitro Developmental Acceleration of Hippocampal Neurons on Nanostructures of Self-Assembled Silica Beads in Filopodium-Size Ranges Jin Seok Lee	49
Effect of Mode of Binding Linkage on Monolayer Assembly of Zeolite Jin Seok Lee	51

◆ 의학과학과

Retroviral integration profiles : their determinants and implications for gene therapy Kwang-il Lim	55
Genome Sequence of <i>Vibrio</i> sp. Strain EJY3, an Agarolytic Marine Bacterium Metabolizing 3,6-Anhydro-L-Galactose as a Sole Carbon Source Kwang-il Lim	57
Retroviral Infection of hES Cells Produces Random-like Integration Patterns Kwang-il Lim	59
Resistive switching memory properties of layer-by-layer assembled enzyme multilayers Kwang-il Lim	61
Astrocytes regulate adult hippocampal neurogenesis through ephrin-B signaling Kwang-Il Lim	63
Tamoxifen Resistance in Breast Cancer Minsun Chang	65
Preclinical metabolism of LB42908, a novel farnesyl transferase inhibitor, and its effects on the cytochrome P450 isozyme activities Minsun Chang	67
GATA-binding protein 1 is a novel transcription regulator of peroxiredoxin 5 in human breast cancer cells MINSUN CHANG	69

◆ 수학과

Harnack-type inequalities for the porous medium equation on a manifold with non-negative Ricci curvature Jinho Lee	73
Some properties of Chebyshev polynomials Seon-Hong Kim	75
On Two Sequences of Polynomials Satisfying Certain Recurrence Seon-Hong Kim	77
On some combinations of self-reciprocal polynomials Seon-Hong Kim	79

Eigenvalue estimates for stable minimal hypersurfaces Keomkyo Seo	81
Stable minimal hypersurfaces in a Riemannian manifold with pinched negative sectional curvature Keomkyo Seo	83
Isoperimetric inequalities for submanifolds with bounded mean curvature Keomkyo Seo	85
Relative isoperimetric inequalities for minimal submanifolds outside a convex set Keomkyo Seo	87
Delta sets for divisors supported in two points Seungkook Park	89
Finitely generated gr-multiplication modules Seungkook Park	91
Module-theoretic characterizations of generalized GCD domains, III Seungkook Park	93
Relation between the floor bound and the order bound Seungkook Park	95
Some properties of gr-multiplication modules Seungkook Park	97
Voluntary retirement and portfolio selection: Dynamic programming approaches Yong Hyun Shin	99
Portfolio selection with regime-switching dynamic programming approaches Yong Hyun Shin	101

◆ 통계학과

New Design of Choice Sets for Choice-based Conjoint Analysis Bu-Yong Kim	105
Analysis of Consumer Preferences for Cosmetic Essence-for-Men via Choice-Based Conjoint with New Design of Choice Sets Bu-Yong Kim	107
Preliminary test of fit in a general class of conditionally heteroscedastic nonlinear time series S.Y. Hwang	109

Asymmetric GARCH processes featuring both threshold effect and bilinear structure S.Y. Hwang	111
Asymptotics for a class of generalized multicast autoregressive processes S.Y. Hwang	113
Arbitrary initial values and random norm for explosive AR(1) processes generated by stationary errors S.Y. Hwang	115
Non-ergodic martingale estimating functions and related asymptotics S.Y. Hwang	117
재충전이 있는 연속시간 리스크 모형에서 파산확률 연구 이익용	119
Ruin Probabilities in a Risk Model with Two Types of Claims Eui Yong Lee	121
Development of differential criteria on tongue coating thickness in tongue diagnosis In-Kwon Yeo	123
Statistical Interpretation of Economic Bubbles In-Kwon Yeo	125
Analysis of Food Poisoning via Zero Inflation Models In-Kwon Yeo	127
식중독 발생 예측모형 여인권	129
Statistical Analysis of Bivariate Current Status Data with Informative Censoring Using Frailty Effects Yang-Jin Kim	131
Statistical Analysis of K-League Data using Poisson Model Yang-Jin Kim	133
Interactions of single nucleotide polymorphisms with dietary calcium intake on the risk of metabolic syndrome Kyunga Kim	135
Detection of Differentially Expressed Gene Sets in a Partially Paired Microarray Data Set Kyunga Kim	137
Permutation test for incomplete paired data with application to cDNA microarray data Kyunga Kim	139

Novel Loci for Adiponectin Levels and Their Influence on Type 2 Diabetes and Metabolic Traits: A Multi-Ethnic Meta-Analysis of 45,891 Individuals Kyunga Kim	141
Distribution-free tests of mean vectors and covariance matrices for multivariate paired data Kyunga Kim	143
Evaluating adherence to recommended diets among cancer patients Kyunga Kim	145
cuGWAM: Genome-wide Association Multifactor Dimensionality reduction using CUDA-enabled high-performance graphics processing unit Kyunga Kim	147

◆ 컴퓨터과학과

A State Preserving Approach to Recognizing Human Behavior using Wireless Infrared and Vibration Sensors Bonghee Moon	151
Tridex: A lightweight triple index for relational database-based Semantic Web data management Junho Shim	153
대수적 특성을 고려한 벡터 유사도 측정 함수의 고찰 심준호	155
The dynamic competitive recommendation algorithm in social network services Seok Jong Yu	157
클러스터 비디오 서버에서 버퍼 공유를 최대화하는 동적 요청 분산 기법 이기용	159
시맨틱 웹 데이터에서 키워드 검색을 위한 구조 질의 랭킹 기법 이기용	161
A Privacy Protection Method for Social Network Data against Content/Degree Attacks Ki Yong LEE	163

◆ 멀티미디어과학과

Designing a Model for Context Awareness Based Reliable Auction-Recommendation System (CARARS) by Utilizing Advanced Information Hwa-Jin Park	167
---	-----

동적 경로 선정을 위한 효율적인 탐색 기법 박화진, 박영호	169
장애우 및 고령자 측면에서의 모바일 SNS 앱접근성 평가 박화진	171
캐릭터스티커 기능을 활용한 창의력 증진용 모바일 e-동화 웹앱 구현 박화진	173
디지털 말하기 책을 위한 음성 주석달기 시스템 이종우, 임순범	175
내용 MathML에서의 수식 독음 규칙 및 변환 임순범	177
스마트폰 음악 플레이어에서의 주변환경을 고려한 멀티모달 자동제어기법 임순범	179
독서장애인 전자책을 위한 음성인식을 이용한 어노테이션 브라우징 기법 이종우, 임순범	181
국내 독서장애인을 위한 Math Expression Reader의 구현 및 사용성 평가 이종우, 임순범	183
대학 교과서 콘텐츠를 위한 사용성 중심의 UI 설계 방안 임순범	185
키네틱 타이포그래피의 활용을 위한 메신저 용 모션 라이브러리 설계 임순범	187
UX 기반 소프트웨어 UI 설계를 위한 교과목 개발 임순범	189
CSS3를 활용한 웹 콘텐츠의 3D 표현 기법 연구 임순범	191
인스턴트 메신저에서 키네틱 타이포그래피의 모션 라이브러리 및 인터페이스 설계 임순범	193
N-스크린 환경에서의 웹 콘텐츠 재구성을 위한 콘텐츠 분리 방법 임순범	195
스마트 교육환경을 위한 e-교과서 포맷 표준화 방안 연구 임순범	197
H5Station: HTML5-기반 효과적인 멀티미디어 파일 관리 시스템 이종우	199

독서장애인용 모바일 전자책뷰어 인터페이스 설계 이종우, 임순범	201
사진 콘텐츠의 분류를 위한 의미적 유사도 기반 구조적 태그 클러스터링 기법 이종우	203
안드로이드 기반 환경오염 신고 앱 설계 및 구현 이종우	205
Efficient techniques on retrieving bio-information for active U-healthcare Young-Ho Park	207
A Design and Implementation of A Real-time Recruiting System on Android Platform Young-Ho Park	209

나노물리학과

Perpendicular Magnetic Anisotropy of Amorphous Ferromagnetic CoSiB/[Pt,Au] Multilayer

Sol JEONG and H. I. YIM*

Department of Physics, Sookmyung Women's University, Seoul 140-742, Korea

(Received 7 December 2011)

Perpendicular magnetic anisotropy is being widely studied as a possible candidate for a high-density spin-transfer torque magnetic random access memory. The key issues of a high-density spin-transfer torque magnetic random access memory are decreasing the switching current and the high thermal stability. In order to solve these problems, two approaches are suggested: One is the development a new amorphous ferromagnetic material as a pinned layer for a multilayer with a low saturated magnetization (M_s) value because of the interface roughness between the two layers. The other is a search for the most suitable materials with high perpendicular magnetic anisotropy in order to have high thermal stability. In this work, we present an amorphous ferromagnetic $\text{Co}_{75}\text{Si}_{15}\text{B}_{10}$ material and compare the magnetic properties of a $[\text{CoSiB} (0.3, 0.4, 0.5 \text{ nm})/\text{Pt} (1.4 \text{ nm})]_5$ multilayer and new combinations $[\text{CoSiB} (0.3, 0.4, 0.5 \text{ nm})/\text{Au} (1.5 \text{ nm})]_5$.

PACS numbers: 75.70.Cn, 75.70.-I, 75.50.Kj

Keywords: Perpendicular magnetic anisotropy, CoSiB, Magnetic multilayer, Amorphous ferromagnetic materials, Coercivity, Thin film

DOI: 10.3938/jkps.60.450

I. INTRODUCTION

The phenomenon of a magnetic thin film being preferentially magnetized perpendicular to the film plane is called “perpendicular magnetic anisotropy” (PMA). This is currently of interest in relation to high-density magnetic recording and Co/Cr thin films in which the anisotropy has a magneto-crystalline origin are popular examples. In the search for other candidate recording media, multilayered thin films offer the possibility that the anisotropy is dominated by the existence of interface anisotropy. Magnetic multilayer films with perpendicular magnetic anisotropy have been intensively studied as possible candidates for spin-transfer torque magnetic random access memories (STT-MRAMs) [1–3].

Some materials, which include amorphous metal alloy and magnetic multilayer films based on Co/Pt and Co/Pd, have been investigated because exchange coupled super-lattice films of alternating ferromagnetic and nonmagnetic layers can be engineered to show strong perpendicular magnetic anisotropy [4–7]. Mixing of this kind had been demonstrated for the CoSiB/Pt multilayer by the Magnetic Materials Science Laboratory at Sookmyung Women's University. The amorphous metal alloy CoSiB/Pt multilayer was expected to be a potential candidate material for the free layer in a high-density magnetic random access memory with a magnetic tun-

nel junction (MTJ) element, and amorphous magnetic materials are required for high-density magnetic devices because such materials show better interface roughness than crystalline materials [8–12]. Therefore, we presented $\text{Co}_{75}\text{Si}_{15}\text{B}_{10}$ in a previous study, obtained data for that material and proposed that the CoSiB/Pt multilayer thin film had a strong perpendicular magnetic anisotropy in a previous letter [13].

A chemical element that has properties similar to these of Pt is Au. Both Au and Pt are face-centered cubic (fcc) structures, the atomic numbers of Au and Pt are 79 and 78 and the atomic radii of Au and Pt are 144 pm and 139 pm, respectively. From previous data, we present a new multilayer, CoSiB/Au, in this letter, and we compare the CoSiB/Pt multilayer and the CoSiB/Au multilayer for the PMA.

II. EXPERIMENTS

Thin films with thicknesses from 16 nm to 18 nm were produced by DC sputtering on Si-SiO₂ substrates. The chamber base pressure was less than 3.0×10^{-7} Torr, and the working pressure was 2×10^{-3} Torr in Ar gas. All thin films were deposited at room temperature, and size were 1.4 cm $\times$ 1.4 cm. The film stacks consisted of films were the substrate Si-SiO₂ (100 nm)/Pt (3 nm)/[CoSiB (0.3, 0.4, 0.5 nm)/Pt (1.4 nm)]₅/Ta (5 nm) and the substrate Si-SiO₂ (100 nm)/Au (3 nm)/[CoSiB (0.3, 0.4,

*E-mail: haein@sm.ac.kr; Fax: +82-2-710-9239

Table 1. Values of the coercivity (H_c) and the saturated magnetization (M_s) of [CoSiB (0.3, 0.4, 0.5 nm)/Pt (1.4 nm)]₅ multilayers.

Multilayer combination	Coercivity (H_c)	Magnetization (M_s)
[CoSiB (0.3 nm)/Pt (1.4 nm)] ₅	78.330 Oe	230.39×10^{-6} emu
[CoSiB (0.4 nm)/Pt (1.4 nm)] ₅	190.32 Oe	320.38×10^{-6} emu
[CoSiB (0.5 nm)/Pt (1.4 nm)] ₅	239.61 Oe	375.83×10^{-6} emu

0.5 nm)/Au (1.5 nm)]₅/Ta (5 nm). Pt and Au were the seed layers (buffer layers), and Ta were the capping layer [13,14].

A VSM (vibrating sample magnetometer) system was used for measuring Coercivity (H_c) and saturated magnetization (M_s). A magnetic field of 1k Oe was applied in-plane and out-of-plane. A SEM (scanning electron microscope) was used to view cross sections of the CoSiB/Au bi-layer and the interface between two layers, and MFM (Magnetic Force Measurement) was used to check a single domain or multi domain.

III. RESULT AND DISCUSSION

[Co/Pt]_n and [CoFe/Pt]_n multilayer films which have been extensively studied as candidate for PMA, are believed to mainly consist of a Co/Pt (CoFe/Pt) interface-type anisotropy [15–17], and CoSiB has been studied as a quite soft magnetic material. In a previous study, we studied the magnetic properties of CoSiB/Pt multilayers with various thicknesses of the CoSiB and the Pt layers. We obtained the H_c and the M_s of a CoSiB single layer as 1.6 Oe and 407 emu/cm³, respectively. In this study, we proposed a new combination CoSiB/Au multilayer and compared the magnetic properties of the CoSiB/Pt multilayer and the CoSiB/Au multilayers with various thicknesses of CoSiB for fixed on thicknesses of Pt and Au.

Figure 1 present MFM images of [CoSiB (0.3 nm)/Pt (1.4 nm)]₅ and [CoSiB (0.3 nm)/Au (1.5 nm)]₅. Pt (1.4 nm) and Au (1.5 nm) are 10-monolayer-thick. The MFM image shows whether a single domain structure or a multi domain structure exists. This multi domain structure has been well-understood to be a result of the competition between domain wall energy and the magneto-static energy. In magnetic resistance (MR) devices, this multi domain structure must be avoided [17–20]. The sample [CoSiB (0.3 nm)/Au (1.5 nm)]₅ also has a single domain structure with [CoSiB (0.3 nm)/Pt (1.4 nm)]₅ films, which means the CoSiB/Au multilayer shows a PMA like the CoSiB/Pt multilayer. However, we conclude that the CoSiB/Au multilayer does not show a PMA because the angle of image of the CoSiB/Au multilayer sample is larger than that of the CoSiB/Pt multilayer sample. The angle of zero (0°) means that the spin rotates about the z-axis (perpendicular to the

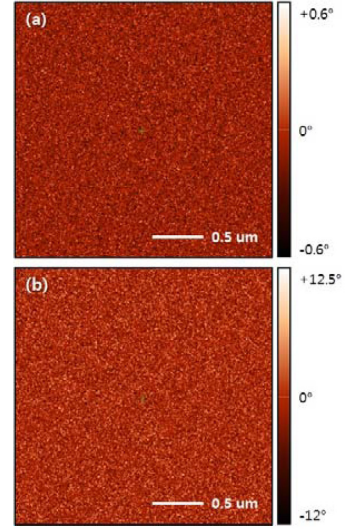


Fig. 1. (Color online) MFM images of the CoSiB/Pt multilayer and the CoSiB/Au multilayers: (a) [CoSiB (0.3 nm)/Pt (1.4 nm)]₅ and (b) [CoSiB (0.3 nm)/Au (1.5 nm)]₅ multilayers.

films plane). We did not observe a multi domain structure in any of the samples with three CoSiB/Pt multilayers and three CoSiB/Au multilayers, but the CoSiB/Au multilayer had a large angle (12.5°). Therefore, we speculate that the CoSiB/Au multilayer does not have any magnetic property with a PMA.

Figure 2 shows the dependence of the value of H_c on the Pt thickness in the CoSiB/Pt multilayer. The samples are [CoSiB (0.3, 0.4, 0.5 nm)/Pt (1.4 nm)]₅, and details are presented on Table 1. CoSiB (0.3, 0.4, 0.5 nm)/Au (1.5 nm)]₅ multilayer data are not shown in this letter because these samples tend to have non-perpendicular magnetic properties. The hysteresis loop of Fig. 2 shows good squareness, which means the CoSiB/Pt multilayer has a strong PMA. However, the CoSiB/Au multilayer does not have a magnetic property with a PMA because the hysteresis loop of the CoSiB/Au multilayer does not draw any shape like a square or a bowtie. These data signify that the CoSiB/Au multilayer does not have a PMA.

Table 1 shows the values of the H_c and the M_s for the CoSiB/Pt multilayer and the CoSiB/Au multilayers.

Glass Formation and Magnetic Properties in the Co and Fe-based Amorphous Alloy with Small Mo or Nb Additions

Hye-in JO and Haein CHOI-YIM*

Department of Physics, Sookmyung Women's University, Seoul 140-742, Korea

(Received 14 December 2011, in final form 23 December 2011)

The ferromagnetic metallic glass alloys in the Co-Fe-B-Si-M (M = Mo or Nb) system were investigated. Ingots of $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Mo}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) and $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Nb}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) alloys were cast into about 30 μm -thick ribbons by using a melt spinning method. As-spun ribbons were identified as being fully amorphous by using X-ray diffraction (XRD). The thermal stability parameters, such as the supercooled liquid region ($\Delta T_x = T_g - T_x$) and the exothermic heat (ΔH_{Exo}), were measured by using differential scanning calorimetry (DSC). For Co-Fe-B-Si-Mo alloys, ΔT_x and ΔH_{Exo} were in the range of 41.4 – 81.7 K and 23.40 – 104.21 J/g, and for Co-Fe-B-Si-Nb alloys, ΔT_x and ΔH_{Exo} were in the range of 60.4 – 88.0 K and 18.45 – 62.05 J/g, respectively. These glassy ribbons also exhibited semi-hard magnetic properties, *i.e.*, a coercive force (H_c) of 67.26 – 100.68 Oe and a saturation magnetization (M_s) of 341.288 – 746.176 emu-cm³.

PACS numbers: 81.05.Kf, 81.40.Rs, 81.70.Pg

Keywords: Co-based, Fe-based, Amorphous, Ferromagnetic, Semi-hard magnetic, Magnetic properties

DOI: 10.3938/jkps.60.477

I. INTRODUCTION

Since the first Fe-based metallic glasses of the Fe-P-C alloy was synthesized in 1967 by Duwez and Lin [1], ferromagnetic metallic glasses have attracted much attention from researchers and have been studied extensively [2]. In 1976, $\text{Co}_{70}\text{Fe}_{15}\text{B}_{10}\text{Si}_{15}$ was synthesized by a technique using melt spinning into rotating water. Subsequently, (Fe, Co)-Si-B amorphous alloys with engineering importance were developed in 1974, followed by the formation of (Fe, Co, Ni)-(Cr, Mo, W)-C, (Fe, Co, Ni)-(Zr, Hf) and then (Fe, Co, Ni)-(Zr, Hf, Nb)-B amorphous alloys [3]. After 1995, Fe-(Al, Ga)-(P, C, B, Si, Ge) [4], (Fe, Co, Ni)-(Zr, Nb, Hf, Ta)-B [5], (Fe, Co)-(Zr, Hf)-(Nb, Ta)-(Mo, W)-B [6], (Fe, Co)-Ln-B [7] (Ln = lanthanide metal), and (Nd, Pr)-Fe-Al [8] amorphous systems were reported with large supercooled liquid regions, low critical cooling rates below 100 K/s and high glass forming abilities. Developing glass forming ability and magnetic properties in Co- and Fe-based amorphous alloys has become the main topic because such alloys have potential possibilities for practical applications the magnetic heads, magnetic memories, electro-magnetic cores and sensors [9–11]. Especially, the soft magnetic properties have been characterized by the achievement of a high saturation magnetization for Fe-based alloys and high permeability and zero magnetostriction for Co-based al-

loys [12]. Also, the glass-forming ability (GFA) of the Fe-based alloy has been found to be improved by the addition of suitable quantities of metalloids [13]. With those kinds of methods, 20 – 30 at.% B and Si addition in Co-based alloys has extended the supercooling range and improved glass-forming ability [14,15]. However, B and Si decrease the magnetic moments of the material [16,17]. Thus, proper amounts of additive elements substituted for the metalloids (Si or B) in Co-based amorphous alloys might improve the glass-forming ability without decreasing the soft magnetic properties. Thus, Amorphous alloy CoFeBSi was chosen as a basic composition in this study, and we added 4 at.% of Mo or Nb to it, which was very effective for increasing the GFA and improving the thermal stability, with nearly no harmful effect on the soft-magnetic properties [18].

In this research, the thermal stability and magnetic properties of the series of $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Mo}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) and $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Nb}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) melt-spun amorphous alloys were investigated in detail. We aimed at obtaining metallic glasses with high glass forming ability and good magnetic properties and clarifying the effect of the replacement of Co by Fe and the role of small amount of Mo or Nb additions.

*E-mail: haein@sookmyung.ac.kr; Fax: +82-2-2077-7320

Table 1. Thermal and magnetic properties of $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Mo}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) and $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Nb}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) glassy alloys.

Alloys	Thermal Properties				Magnetic Properties	
	T_g (K)	T_x (K)	ΔT_x (K)	$\Delta H_{Exo.}$ (J/g)	H_c (Oe)	M_s (emu/dotcm ³)
$\text{Co}_{72}\text{B}_{19.2}\text{Si}_{4.8}\text{Mo}_4$	750.8	792.2	41.4	-23.40	67.26	416.711
$\text{Co}_{57.6}\text{Fe}_{14.4}\text{B}_{19.2}\text{Si}_{4.8}\text{Mo}_4$	787.5	845.1	57.6	-92.88	91.14	409.425
$\text{Co}_{36}\text{Fe}_{36}\text{B}_{19.2}\text{Si}_{4.8}\text{Mo}_4$	789.7	853.5	63.8	-52.25	68.74	518.864
$\text{Co}_{14.4}\text{Fe}_{57.6}\text{B}_{19.2}\text{Si}_{4.8}\text{Mo}_4$	792.2	870.6	78.4	-104.21	100.68	588.735
$\text{Fe}_{72}\text{B}_{19.2}\text{Si}_{4.8}\text{Mo}_4$	781.8	863.5	81.7	-96.73	69.54	746.176
$\text{Co}_{72}\text{B}_{19.2}\text{Si}_{4.8}\text{Nb}_4$	737.2	825.2	88.0	-18.45	84.73	341.288
$\text{Co}_{57.6}\text{Fe}_{14.4}\text{B}_{19.2}\text{Si}_{4.8}\text{Nb}_4$	797.7	858.1	60.4	-54.83	65.82	402.567
$\text{Co}_{36}\text{Fe}_{36}\text{B}_{19.2}\text{Si}_{4.8}\text{Nb}_4$	805.8	874.9	69.1	-54.23	89.01	617.989
$\text{Co}_{14.4}\text{Fe}_{57.6}\text{B}_{19.2}\text{Si}_{4.8}\text{Nb}_4$	808.4	888.5	80.1	-62.05	75.58	599.476
$\text{Fe}_{72}\text{B}_{19.2}\text{Si}_{4.8}\text{Nb}_4$	801.5	878.6	77.1	-49.76	72.93	597.443

II. EXPERIMENTS AND DISCUSSION

$[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Mo}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) and $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Nb}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) were prepared by using an arc-melting furnace under a Ti-gettered argon atmosphere. Each high purified ingot was remelted five times with high arc-power for compositional homogenization. Glass alloy ribbons with a thickness of 25 – 30 μm and a width of 3 mm were fabricated by using the single-copper-roller melt spinning process.

The glassy amorphous structure was identified by using X-ray diffraction (XRD) with Cu-K α radiation (Cu-K α , $\lambda = 1.54050$ Å). We measured the XRD patterns for angles of 2θ between the diffracted beam and the transmitted beam from 20° to 100° at a rate of $0.02^\circ/\text{s}$. The thermal stability values of the glass transition temperature (T_g), the crystallization temperature (T_x), the supercooled liquid region ($\Delta T_x = T_g - T_x$) and the exothermic heat ($\Delta H_{Exo.}$) were examined by using differential scanning calorimetry (DSC). The calorimetric measurement (DSC) were performed at temperatures from room temperature to 973 K at a heating rate of 0.67 K/s. The magnetic properties, such as the saturated magnetization and coercivity, were measured at room temperature by using a vibrating sample magnetometer (VSM). Based on the absence of crystal anisotropy in Co-Fe based amorphous materials, the samples were examined with an in-plane magnetic field in the range from -1.5 Tesla to 1.5 Tesla.

The structural, thermal and magnetic properties of $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Mo}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) and $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Nb}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) were examined. Figure 1 shows X-ray diffraction patterns show broad and halo peaks which characteristic of an amorphous sturcture. However, the $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Nb}_4$ ($x = 1.0$) samples has the sharp diffraction peak at 59° corresponding to a crystalline phase that has not been determined yet.

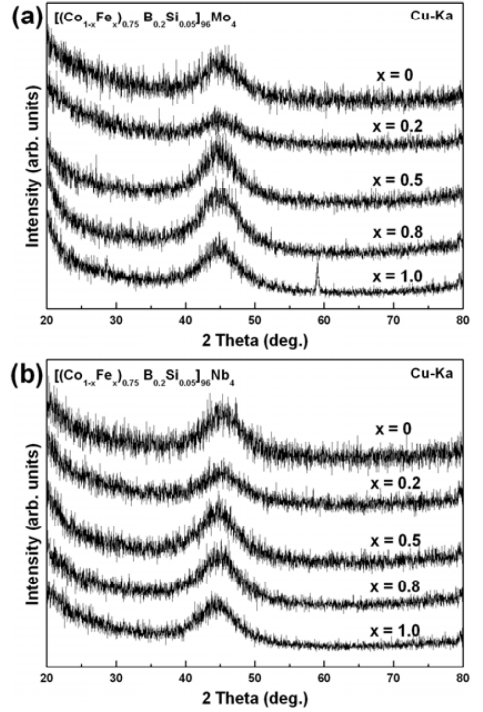


Fig. 1. XRD patterns of (a) $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Mo}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) and (b) $[(\text{Co}_{1-x}\text{Fe}_x)_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{96}\text{Nb}_4$ ($x = 0, 0.2, 0.5, 0.8, 1.0$) glassy alloys.

The glass transition and crystallization behaviors of amorphous CoFeBSiMo and CoFeBSiNb were measured by using differential scanning calorimetry (DSC). The T_g and the T_x represent the onset of the glass transition temperature and the onset temperature of the first crystallization exothermal event, respectively, and the value of the supercooled liquid region ΔT_x is defined as T_x

Thermal and Elastic Properties of Cu-Hf-Al Bulk Metallic Glasses

Haein CHOI-YIM*

Department of Physics, Sookmyung Women's University, Seoul 140-742, Korea

(Received 22 December 2011)

Stable bulk glass forming alloys can be developed over a wide range of compositions in the Cu-Hf-Al ternary system. The compositional and the cooling rate dependences of the thermal and the elastic properties of Cu-Hf-Al alloys were systematically investigated. Evidence of the amorphous nature of the cast rods was provided by X-ray spectra. The glass transition temperature (T_g) and crystallization temperature (T_x) were obtained for the alloys by using differential scanning calorimetry (DSC) at a heating rate of 20 K/min. In the results, the differences between the glass temperature and the crystallization temperature ($\Delta T_x = T_x - T_g$) were measured to have values ranging up to 72 – 74 K. The elastic constants for several selected alloys were measured using ultrasonic methods. The values of the elastic shear modulus, bulk modulus, and Poisson's ratio are also given.

PACS numbers: 81.05.Kf, 61.43.Er, 62.20.Fe

Keywords: Bulk metallic glasses, Elastic constants, Glass-forming ability

DOI: 10.3938/jkps.60.485

I. INTRODUCTION

Bulk metallic glasses (BMGs) are attractive structural materials because of their unique mechanical properties. BMGs have high yield strength, elastic strain limit (about 2%), hardness, fracture toughness, wear and corrosion resistance [1–5]. Many BMGs have been developed in multicomponent systems such as Zr-, Ti-, Fe-, Ni-, Pd-, Mg-, Cu, and Co-based alloys [1–9]. BMGs based on certain late transition metals (*e.g.*, Fe, Co, Ni, and Cu) have many potential advantages over those based on early transition metals. These potential advantages include even a higher strength and elastic modulus and lower material cost, which are highly preferable for broad applications of BMGs as engineering materials.

Cu-based bulk metallic glasses have been developed because of the outstanding thermophysical and mechanical properties. The Cu-based BMGs offer high strength over 2 GPa and high glass-forming ability with 4 – 10 mm in diameter by copper mold casting [10–19]. In the Cu-based systems, the $\text{Cu}_{47}\text{Ti}_{34}\text{Zr}_{11}\text{Ni}_8$ (Vit 101) alloy can be cast into a thickness of 4 mm by copper mold casting. The Cu-(Zr, Hf)-Ti, Cu-Zr-Ti-Co, and Cu-Ti-Zr-Ni alloy systems show high glass-forming ability, stable thermal property, and high strength [10–13]. Moreover, Cu-based alloys have good ductility and lower cost [14–19].

The ductility of BMGs correlates to the Poisson's ratio (ν) and the ratio of the shear modulus (G) to the bulk

modulus (B). The low shear modulus allows shear collapse before the extensional instability of crack formation occurs. Therefore, a small G/B ratio causes the tip of a shear band to extend; then, the formation of high-density shear bands causes large ductility [20]. Lewandowski *et al.* found a correlation between the ductility and the ratio of G/B . When the G/B ratio ranges from 0.41 to 0.43 (or the Poisson's ratio (ν) < 0.31 – 0.32) for metallic glasses, the metal is brittle [21]. For instance, a Pt-based BMG with excellent compressive ductility shows a low G/B ratio of 0.165, a high Poisson's ratio of 0.42, and a high fracture toughness of 80 $\text{MPa}\cdot\text{m}^{-1/2}$ [20]. In addition, a recent molecular dynamics (MD) simulation indicates that the shear modulus of a metallic glass depends on its cooling rate [22]. This is supported from the existence of a fundamental difference in the Poisson's ratio between slowly-cooled bulk-type and rapidly-cooled ribbon-type metallic glasses.

The total potential energy barrier (W) for a shear transformation zone (STZ) is shown as [23,24]

$$W = \left(\frac{8}{\pi^2} \right) \gamma_c^2 G \xi \Omega, \quad (1)$$

where Ω is an effective zone volume, ξ is a correction factor arising from the matrix confinement, and γ_c is the critical shear strain limit shown to be a universal scale for metallic glasses. The details of ξ depend on the elastic constants and the Poisson's ratio (ν) for the material. Therefore, the cooperative shear model shows that the activation barrier height for shear flow is proportional to shear modulus (G), which indicates that the elastic

*E-mail: haein@sookmyung.ac.kr; Fax: +82-2-2077-7320

Table 1. Thermal and elastic properties of Cu-Hf-Al glassy alloys.

Composition	Thickness (mm)	T_g (K)	T_x (K)	ΔT_x (K)	G (GPa)	E (GPa)	B (GPa)	G/B	ν
Cu ₄₀ Hf ₄₂ Al ₀	3	—	—	—	42.78	115.46	128.40	0.335	0.349
	4	789.49	862.81	73.32	42.00	113.89	132.40	0.319	0.356
Cu ₄₈ Hf ₄₁ Al ₁₁	3	—	—	—	46.75	125.46	135.15	0.353	0.342
	4	786.29	858.44	72.15	48.62	129.59	131.75	0.377	0.333
Cu ₄₈ Hf ₄₃ Al ₉	3	—	—	—	43.10	116.55	131.48	0.328	0.352
	4	791.00	865.62	74.62	48.00	129.78	146.17	0.328	0.352

constants such as G , B , G/B , and the Poisson's ratio (ν), can be used to understand the mechanical behavior of metallic glasses. A lower G implies a higher Poisson's ratio and means that the shear flow barrier for an unstressed shear cooperative zone is relatively small.

The formation, thermal stability, and compositional and cooling rate dependences of the elastic properties of Cu-Hf-Al metallic glassy alloys were investigated in this research. In addition, the elastic properties, such as the shear modulus (G), the bulk modulus (B), and the Poisson's ratio (ν), were measured and are discussed.

II. EXPERIMENTS AND DISCUSSION

Ingots of Cu-Hf-Al alloys were prepared by arc-melting elemental metals with purities of 99.5% or higher under a Ti-gettered argon atmosphere. Each ingot was remelted at least three times to ensure chemical homogeneity. These alloys were re-melted under high vacuum in a quartz tube by using an induction-heating coil and were injected into a copper mold by using high purity argon at a pressure around 2.02×10^5 Pa. Transverse cross sections of all were characterized by using X-ray diffraction (XRD) with a D8 advance and Cu-K α radiation. The glass transition and the crystallization behaviors of the samples were studied with a differential scanning calorimeter (DSC) at a heating rate of 20 K/min. The Shear modulus, the bulk modulus, and the Poisson's ratio were obtained using an ultrasonic device by measuring the sound velocities of the longitudinal and the transverse waves in fully amorphous samples.

Figure 1 shows the X-ray diffraction patterns of cast rods with a diameter of 4 mm for selected Cu-Hf-Al alloys. Within the detectable limit of XRD, the diffraction patterns show a typical diffuse first nearest neighbor peak associated with amorphous metals. However, all samples show incipient crystallization.

The thermal properties of three glassy alloys were measured by using a differential scanning calorimeter (DSC). The results, such as the glass transition temperature T_g , the crystallization temperature T_x , and the supercooled liquid region $\Delta T_x (=T_g - T_x)$, are listed in Table 1, and the DSC curves of three glassy alloys are shown in Fig. 2. All the glassy alloys exhibit an endothermic reaction of

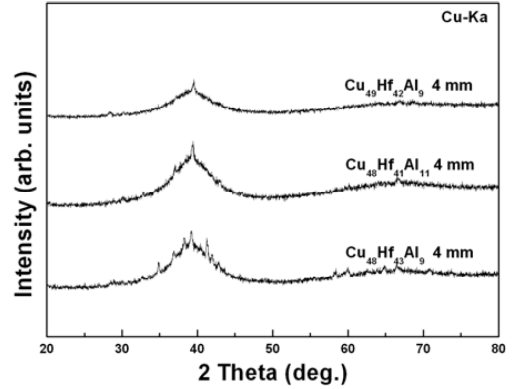


Fig. 1. X-ray diffraction patterns of Cu-Hf-Al samples.

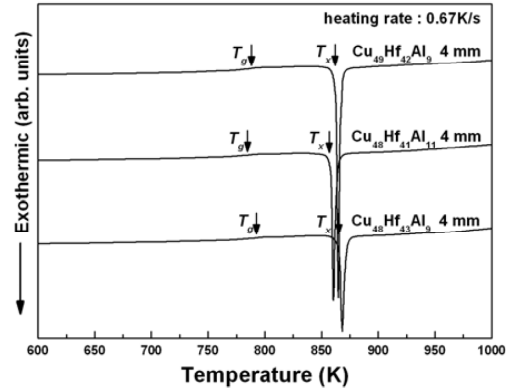


Fig. 2. DSC curves of glassy Cu-Hf-Al alloys.

the glass transition, followed by the onset of crystallization events and exothermic events due to crystallization. In the results, the differences between the glass temperature and the crystallization temperature ($\Delta T_x = T_x - T_g$) are measured to have values ranging up to 72 – 74 K. T_g and T_x range between 786.29 and 791 K and 858.44 and 865.42 K, respectively. The Cu₄₈Hf₄₃Al₉ alloy shows the largest ΔT_x value, which indicates that this composition is the most thermally stable.

The Young's modulus and the Poisson's ratio were ob-

Synthesis and Characterization of Co-Fe-based Amorphous Ribbons with Small Nb Additions

Hye-in JO and Haein CHOI-YIM*

Department of Physics, Sookmyung Women's University, Seoul 140-742, Korea

(Received 29 May 2012, in final form 12 June 2012)

The ferromagnetic metallic glass alloys in the Co-Fe-B-Si-Nb system were investigated. Such glassy alloys exhibit high glass-forming ability (GFA), as well as good mechanical and magnetic properties. Based on those results, $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) were chosen for this investigation with increasing Nb contents. Amorphous ribbons with $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) were obtained by single-roller melt spinning. As-spun ribbons were identified as fully amorphous by using X-ray diffraction (XRD). The thermal stability parameters, such as the supercooled liquid region ($\Delta T_x = T_g - T_x$), were measured by using differential scanning calorimetry (DSC). The mechanical properties of the Co-Fe-B-Si-Nb system were measured by using a nano-indentation test. Co-Fe-B-Si-Nb glassy ribbons also exhibited good soft magnetic properties, a low coercive force (H_c) and a high saturation magnetization (M_s). We report here that a small amount of Nb addition (microalloying) can enhance the thermal stability of the alloy without deteriorating its soft magnetic properties.

PACS numbers. 81.05.Kf, 81.40.Rs, 81.70.Pg

Keywords: Metallic glasses, Amorphous, Ferromagnetic, Soft magnetic, Magnetic properties

DOI: 10.3938/jkps.61.460

I. INTRODUCTION

With scientific and technological progress, novel materials having unique physical, chemical and mechanical properties are required. Many researchers have been devoting themselves to finding new materials for the needs of the times. After the first metallic glass was reported in 1960, metallic glass opened a new chapter of material science [1]. Among the various metallic glasses, Fe- and Co-based alloys are commercially the most important because they are plentiful relative to others, and they can be applied in a variety of industrial applications: magnetic heads, magnetic memories, electro-magnetic cores and sensors [2–4]. Especially, the (Fe, Co)-Si-B amorphous alloys have been used in many application fields as soft magnetic materials [5,6].

Extensive efforts have been made to improve the glass-forming ability (GFA) and the magnetic properties of ferro-group metal-based systems such as the Fe-(Al, Ga)-(P, C, B, Si, Ge), Fe-(Co, Ni)-(Zr, Hf, Nb)-B, and Co-Fe-(Zr, Nb, Ta)-B systems [7–9]. Recently, Inoue et al. reported the magnetic properties and glass-forming ability of (Fe-Co)-B-Si-Nb metallic glass alloys. Those bulk glassy alloys exhibited a super-high fracture strength of over 4000 MPa and good soft magnetic properties [10]. In addition, we reported preliminary results for Co-Fe-B-Si-

Nb and Co-Fe-B-Si-Mo semi-hard magnetic amorphous alloys [11]. With a different view of the composition, we make an attempt to modify the magnetic properties or the structure phase to a Nb-rich dendritic-phase-reinforced $(\text{Cu}_{0.6}\text{Zr}_{0.25}\text{Ti}_{0.15})_{93}\text{Nb}_7$ [12]. Thus, the composition Co-Fe-B-Si-Nb is selected as a base alloy. Nb contents are increased to enhance the thermal, mechanical or magnetic properties.

In this paper, the thermal, mechanical and magnetic properties of a series of $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) melt-spun amorphous alloys was investigated in detail. We aimed at obtaining metallic glasses with high mechanical and good soft magnetic properties and clarifying the effect of a small amount of Nb addition.

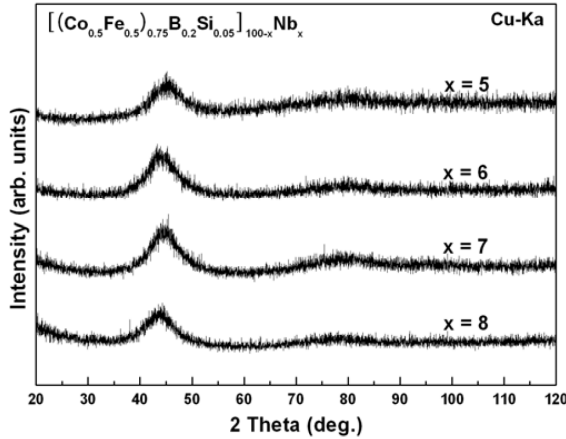
II. EXPERIMENTS AND DISCUSSION

Multicomponent pre-alloys with the composition of $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) were prepared by using an arc-melting furnace under a Ti-gettered argon atmosphere. A highly purified and ultrasonically-cleansed mixture of Co, Fe, B, Si, and Nb was remelted five times with high arc-power for chemical homogeneity. Subsequently, ingots of $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) were placed in an induction coil of a single-roller melt-spinning

*E-mail: haein@sookmyung.ac.kr; Fax: +82-2-2077-7320

Table 1. Thermal, mechanical and magnetic properties of $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) glassy alloys.

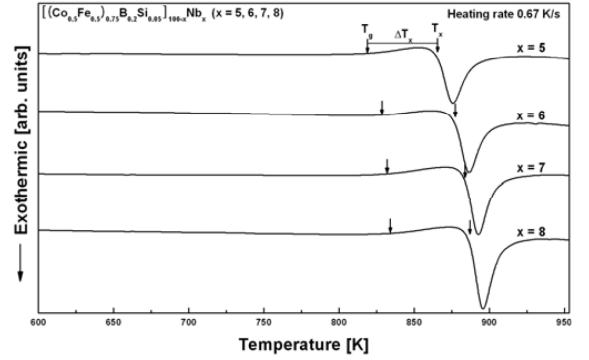
Alloys	Thermal Properties			Mechanical Properties		Magnetic Properties		
	T_g (K)	T_x (K)	ΔT_x (K)	H (GPa)	E_r (GPa)	H_c (Oe)	M_s (emu-cm ³)	M_r (emu-cm ³)
$[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{95}\text{Nb}_5$	819.0	865.1	46.1	15.5	158.2	0.70	756.590	7.444
$[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{94}\text{Nb}_6$	828.5	876.7	48.2	17.1	220.7	0.85	594.014	4.631
$[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{93}\text{Nb}_7$	831.9	883.4	51.5	16.3	218.0	0.87	530.154	8.796
$[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{92}\text{Nb}_8$	833.7	886.5	52.8	17.3	201.3	0.88	513.943	5.914

Fig. 1. XRD patterns of glassy $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) alloys.

apparatus. As-spun glassy ribbons were about 3 mm wide and 30- to 50- μm thick.

The structures of these ribbons were identified by using X-ray diffraction (XRD) with Cu-K α radiation. The X-ray diffraction (XRD) measurements were performed at room temperature for 2θ from 20° to 120° at a rate of $0.02^\circ/\text{s}$. The thermal properties associated with the glass transition temperature (T_g), the crystallization temperature (T_x) and the supercooled liquid region ($\Delta T_x = T_g - T_x$) were examined by using differential scanning calorimetry (DSC) under flowing purified argon gas at a heating rate of 0.67 K/s. The mechanical properties were characterized using a nano-indenter system with a triangular pyramid tip of Berkovich diamond. The magnetic properties, such as the saturated magnetization and the coercivity, were measured at room temperature by using a vibrating sample magnetometer (VSM). Based on the absence of crystal anisotropy in Co-Fe based ferromagnetic amorphous alloys, the samples were examined by using in-plane magnetic field in the range from -2000 Oe to 2000 Oe.

The structural, thermal, mechanical and magnetic properties of $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) were characterized. As shown in Fig. 1, the X-ray diffraction patterns of $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) confirmed an amorphous phase because of the lack of sharp diffraction peaks with high intensity corresponding to a crystalline phase. These typical broad and halo curves show that the as-spun ribbons are in a fully amorphous state.

Fig. 2. DSC curves of glassy $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) alloys.

$[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) confirmed an amorphous phase because of the lack of sharp diffraction peaks with high intensity corresponding to a crystalline phase. These typical broad and halo curves show that the as-spun ribbons are in a fully amorphous state.

The glass transition and crystallization behaviors of amorphous CoFeBSiNb were measured by using DSC. Figure 2 shows DSC scans of the amorphous $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) at a heating rate 0.67 K/s. The glass transition temperature (T_g) represents the onset temperature of the first glass transition endothermic heat event, and the crystallization temperature (T_x) represents the onset temperature of the first crystallization exothermal event. Also, the supercooled liquid region ΔT_x which is defined as $T_x - T_g$ indicates the transformation from a metastable undercooled liquid state to a crystalline compounds. In Table 1, thermal properties such as T_g , T_x , and ΔT_x are listed. The results of ΔT_x for $[(\text{Co}_{0.5}\text{Fe}_{0.5})_{0.75}\text{B}_{0.2}\text{Si}_{0.05}]_{100-x}\text{Nb}_x$ ($x = 5, 6, 7, 8$ at.%) are in the range of 46.1 – 52.8 K. According to the results, the values of T_g , T_x and ΔT_x increase gradually with small increases in the Nb content. Alloys containing 8 at.% Nb are more thermally stable than others because they have a wide supercooled liquid region (ΔT_x).

The nano-indentation technique is now widely used in evaluating the mechanical properties of thin and ultrathin films [13]. The nanohardness and the reduced

Initial Oxidation Structure of Chlorinated Si(001)

Hanchul KIM*

Department of Physics, Sookmyung Women's University, Seoul 140-742, Korea

(Received 5 January 2012, in final form 16 January 2012)

The Cl-saturated Si(001) surface is investigated by employing density functional theory calculations. From the electronic density of states, the low- and the high-bias filled-state images and the empty-state images are found to originate from Cl(3p) states directed in-plane, in parallel with the Si dimer row, and perpendicular to the surface (z), respectively. The bright triplet, which is a characteristic scanning tunneling microscopy (STM) feature of an unchlorinated bare Si dimer, is attributed to an occupied dangling bond state of the bare Si dimer that is hybridized with the neighboring Cl(3p_z) orbital. The previously observed split-dimer feature is confirmed to be the dimer-bond oxidized structure. Finally, we suggest that the initial back-bond oxidation does not occur on the chlorinated Si(001) surface because it is energetically not favored compared with dimer-bond oxidation and because its STM simulation does not match experimental images.

PACS numbers: 68.43.Bc, 68.43.Fg, 73.20.At, 73.20.Hb

Keywords: Density functional theory, Silicon, Oxidation, Chlorination

DOI: 10.3938/jkps.60.398

I. INTRODUCTION

Oxidation of the Si (001) wafer is one of the most important processes in modern electronic device technology. Forming a several-atomic-layer thin oxide film with an atomically flat interface is desirable in order to develop a high-speed, large-capacity device with reduced size. For that purpose, one has to understand the oxidation mechanism microscopically and to control the oxidation process atomically.

Silicon oxidation has been extensively studied using either oxygen (dry oxidation) [1–4] or water (wet oxidation) [5–7]. The very initial stage of Si oxidation is theoretically understood to occur through O atom incorporation into the dimer bond (DB) or into the back-bond (BB, the Si-Si bond formed between the surface and the subsurface atoms) by breaking Si-Si bonds. However, experimental confirmation of such initial oxidation process has been absent due to some technical difficulties like surface roughening upon oxidation [8–10].

In a series of work, Weaver and his collaborators were able to observe the DB-oxidized structure by using scanning tunneling microscopy (STM) on a residual-water exposed Si(001) surface that had been halogenated with Cl, Br, and I [11–14]. Point defects on Si(001), such as type-*C* defects and some dark missing dimers, were recently shown to be products of the adsorption of residual water molecules in ultrahigh vacuum chambers [15–19]. Chlorination seems to be efficient in isolating individ-

ual initial oxidation processes and in preventing surface roughening.

Weaver and collaborators observed three isolated STM features: bright triplet dimers, missing dimers, and split dimers (SD) [14]. The bright triplet feature is commonly observed on the halogenated Si(001) surface and has been proposed to be an unreacted bare or clean Si dimer (CD) [11–13], which was subsequently verified by using density functional theory (DFT) calculations [20]. The missing dimer is identified as the Si dimer vacancy (DV). Among the three STM features, SD was the only feature depending on the density of the type-*C* defect on the Si substrate and was proposed to be the DB oxidation structure (DB-o). To our knowledge, there has been neither a theoretical confirmation of that proposition nor an investigation of the BB oxidation structure (BB-o) on the chlorinated Si(001) surface.

In this paper, we report the pseudopotential DFT calculations on the initial oxidation of the chlorinated Si(001) surface. For the STM observed features (CD, DV, and SD) and the unexplored BB-oxidized (BB-o) structure, we obtain the relaxed geometry and the total energies. Simulated STM images are utilized to verify experimental observation, and the electronic density of states (DOS) and orbitals are used to understand the bias-voltage dependence of the STM images.

*E-mail: hanchul@sookmyung.ac.kr

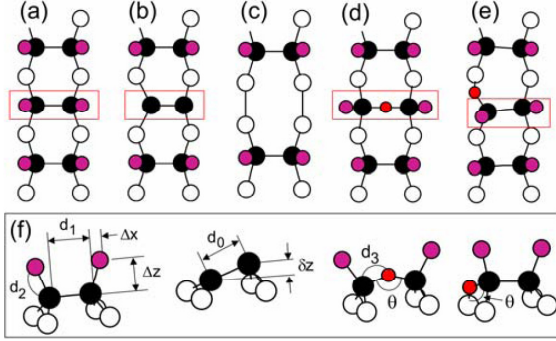


Fig. 1. (Color online) Structures of chlorinated Si(001): (a) the Cl-saturated Si(001)- 2×1 surface, (b) an unreacted clean Si dimer (CD), (c) a dimer vacancy (DV), (d) an oxidized DB (DB-o), and (e) an oxidized BB (BB-o). Some structural parameters are defined in (f). The pink, red, black, and white circles are Cl, O, first-layer Si, and second-layer Si atoms, respectively. The red (solid line) boxes indicate local defective sites.

II. COMPUTATIONAL METHODS

We have performed DFT calculations using the Vienna *Ab initio* Simulation Package [21] within the generalized gradient approximation [22]. A kinetic energy cutoff of 400 eV and a theoretical lattice constant of 5.46 Å are used. The Brillouin zone integrations are done using a Monkhorst-Pack $\mathbf{k}$ -point mesh [23] equivalent to an 16×16 mesh within the 1×1 surface Brillouin zone. The ions are represented by ultrasoft pseudopotentials [24, 25]. The Si(001) substrate is modeled by using the repeated slabs of six Si layers intervened with a 18 Å-thick vacuum. The top surface is $c(4 \times 2)$ reconstructed with a 4×8 surface supercell, and the bottom surface is passivated by using fictitious H atoms. All the atoms, except for those in the two bottom Si layers and passivating H layers, are relaxed so that the atomic forces are ≤ 0.002 eV/Å. The constant-current STM images are simulated within the Tersoff-Hamann approximation by using the fully relaxed geometries [26, 27].

III. ATOMIC STRUCTURES AND ENERGETICS

Upon chlorination, the buckled Si dimers of clean Si(001) surface become symmetric while retaining the Si-Si dimer bond. Because the electrons in the dangling bonds of Si dimers are consumed to form Si-Cl bonds, the Si dimer bond is elongated to 2.43 Å from the 2.35 Å of the clean dimer. The Si-Cl bond is *slanted* towards the troughs between the dimer rows and is about 2.07 Å, which is very similar to the bond length (2.06 Å) in the tetrahedral SiCl₄ molecule [28]. The resulting

Table 1. Structural parameters of chlorinated Si(001) defined in Fig. 1(f). For CD, DV, DB-o, and BB-o, the parameters are for the defective sites or for the nearest-neighbor sites. The numbers in the parentheses are structural parameters for the Si-Cl moiety apart from the BB-incorporated O atom.

	clean surface	Cl-saturated surface	CD	DV	DB-o	BB-o
d_0 (Å)	2.35	—	2.31	—	—	—
d_1 (Å)	—	2.43	2.43	2.45	3.13	2.43
d_2 (Å)	—	2.07	2.07	2.07	2.05	2.05
d_3 (Å)	—	—	—	—	1.64	1.65
δz (Å)	0.75	—	0.71	—	—	0.27
Δx (Å)	—	0.69	0.69	0.65	1.12	0.49
Δz (Å)	—	1.95	1.95	1.97	1.71	1.99
θ (°)	—	—	—	—	144.5	136.1

chlorinated surface shows 2×1 periodicity, as shown in Fig. 1(a). These results are in good agreement with previous DFT calculations [20, 29].

When a pair of Cl atoms are desorbed to create a CD, chlorinated Si dimers reveal virtually the same structural parameters as the perfectly chlorinated surface, as one can tell from Table 1. The clean Si dimer has a slightly shorter bond length compared with the clean Si(001), which seems to be caused by the structural hindrance of neighboring chlorinated dimers. This is consistent with the smaller buckling (δz) of the clean dimer (0.71 Å compared with 0.75 Å on the clean surface). If we remove the bare dimer from CD, the resultant structure is the DV structure with second-layer rebonds, as shown in Fig. 1(c). The nearest-neighbor Si dimers are slightly elongated from 2.43 to 2.45 Å and relax towards the DV by ~ 0.05 Å. The corresponding Si-Cl bonds become slightly more *upright* (smaller Δx and larger Δz in Table 1) compared with those distant from DV.

Regarding the oxidized structures, the Si-O bonds are 1.64 $\sim$ 1.65 Å in the DB-o and the BB-o structures. In DB-o, the Si-Cl bonds of O-reacted sites become more slanted (larger Δx and smaller Δz) and are strengthened, judging from the reduced bond length of 2.05 Å. The BB oxidation inevitably reduces the local symmetry, and the nearest-neighbor Si-Cl bonds are skewed, as seen in Fig. 1(e), and the Si-Cl closer to O is more upright while the farther Si-Cl is more slanted. The Si-O-Si bond angles (θ) of DB-o and BB-o deviate from the corresponding values for dry oxidation configurations by about 2 and -7% , respectively.

Now let us examine the energetics of the studied structures. The Cl saturation of the Si(001) surface is exothermic, and the energy is lowered by 5.06 eV/dimer rela-

Van der Waals Interaction between P_4 Molecules: Density Functional Theory Calculations with Dispersion Correction

Ji-Young NOH and Hanchul KIM*

Department of Physics, Sookmyung Women's University, Seoul 140-742, Korea

(Received 6 January 2012)

We investigated the structural and the binding properties of tetrahedral P_4 molecular pairs by employing density functional theory (DFT) calculations using both the local density approximation (LDA) and the generalized gradient approximation (GGA). To consider the van der Waals interaction, we employed the GGA with dispersion correction, specifically Grimme's approach (DFT-D2). This DFT-D2 in conjunction with the GGA (GGA-D2) well fixes the failure of the GGA in reproducing a reasonable binding energy curve. The binding energy and the equilibrium intermolecular distance are found to be in close correlation with the number of adjacent atomic pairs between two molecules. Application of the LDA and the GGA-D2 to a molecular crystal β - P_4 and comparison with experiments confirms that the GGA-D2 should be used for a reliable description of P_4 -based molecular systems.

PACS numbers: 71.15.Nc, 31.15.Ew, 61.50.Lt, 61.66.Bi

Keywords: Density functional theory, Van der Waals interaction, Phosphorus molecules

DOI: 10.3938/jkps.60.410

I. INTRODUCTION

Density functional theory (DFT) [1, 2] calculations have been greatly successful in describing the ground-state properties of materials, and now DFT has become a widely utilized tool in ever increasing fields. The generalized gradient approximation (GGA) for the electronic exchange-correlation is quite good in describing a strongly interacting system, but it has failed in predicting, particularly, the structural and the thermodynamic properties of van der Waals (vdW) materials because it is not capable of describing the long-range dynamical correlation between induced dipole moments. Thus, many efforts have been made to solve these problems by using approaches such as the vdW-density-functional (vdW-DF) scheme and the DFT with dispersion correction (DFT-D) [3–9]. Among them, a recent version of the DFT-D by Grimme [9], the so-called DFT-D2 method, has been widely used. This approach not only provides appropriate predictions but also requires minimal computational demands [10].

Van der Waals forces are very weak compared to chemical or crystalline bonds, but they play an important role in weakly-bound systems. The vdW interaction is crucial for electrically inactive atomic or molecular systems such as noble gas elements and non-polar molecules [11]. An appropriate description of the vdW interaction is nec-

essary for systems with π -electrons, such as hydrocarbon molecules with unsaturated C-C bonds, carbon allotropes (C_{60} , carbon nanotubes, graphite, *etc.*), and organic molecules in biological systems [5,6]. For instance, Chakarova-Käck *et al.* studied the vdW interactions between benzene and naphthalene molecules on graphite by using the vdW-DF [12]. Hasegawa and Nishidate developed a semiempirical method to calculate the interlayer binding energy of graphite and pointed out that the results of the local density approximation (LDA) underestimated the binding energy of graphite because of the missing vdW interaction [13]. The equilibrium distance of the in-plane benzene dimer and the interlayer distance of graphite were calculated using both the DFT-D2 and a non-empirical dispersion correction method, and the two schemes were found to yield virtually the same results [14,15]. Very recently, the vdW-DF method was applied to bulk solid systems, where vdW forces had been thought to be not important, and found that the original vdW-DF tended to overestimate the lattice constants [16].

It is intriguing to notice that, similarly to carbon, phosphorus has three kinds of allotropes (white, red, and black phosphorus) that have different dimensionalities. White phosphorus is known as a reactive molecular crystal whose building unit is the P_4 molecule. It has three structural modifications with different arrangements of P_4 molecules (α -, β -, and γ - P_4) [17–19]. Red phosphorus is made up of polymeric chains of P, and black phosphorus results from a stacking of puckered P layers [20].

*E-mail: hanchul@sookmyung.ac.kr

Judging from the structural characters of P allotropes, the vdW interaction seems to play a role in the structural and the thermodynamic properties. To our knowledge, there has not been a comprehensive investigation of the role of the vdW interaction in P allotropes except for a couple of very recent studies on black P [21,22].

In this research, we investigated the interaction between P₄ molecules by using the DFT-D2 method for the vdW correction. First, we calculated the cohesive properties of the isolated P₂ and P₄ molecules and looked at the density of states (DOS). Then, the binding energies and equilibrium distances of a pair of P₄ molecules were calculated by using three different methods: LDA, GGA, and GGA with dispersion correction. Finally, we optimized the structural parameters of a P₄ molecular crystal, β -P₄, to compare the performances of the LDA and the GGA with dispersion correction.

II. COMPUTATIONAL DETAILS

We performed the conventional DFT calculations and the dispersion-corrected DFT calculation by using the Vienna *Ab initio* Simulation Package (VASP) [23–26]. P ions were represented by using the projector-augmented-wave potential [27,28]. The exchange-correlation was described following Ceperley and Alder [29] and Perdew, Burke, and Ernzerhof (PBE) [30] for the LDA and the GGA, respectively. The DFT-D2 method implemented in VASP was also used to include the vdW correction. A kinetic energy cutoff up to 450 eV and the Γ -point for k-points sampling were used. An isolated molecule and two molecules were placed in large supercells to guarantee the convergence in total energy within 1 meV/P: a 24 Å × 24 Å × 24 Å cubic supercell was used for a single molecule, and a supercell of 24 Å × 24 Å × 40 Å was used to study the interaction between two molecules.

III. RESULTS

The phosphorus atoms of a P₄ molecule form a tetrahedral structure. Among the phosphorus clusters, the dimer (P₂) and the tetramer (P₄) are known to be stable and are experimentally observed, and the tetrahedral P₄ is stable up to 800 °C [31]. Before investigating the vdW interaction between P tetramers, we calculated the optimized bond lengths and the cohesive energies for both P₂ and P₄ molecules to compare their stabilities.

Table 1 shows the P-P bond lengths and cohesive energies of P₂ and P₄ obtained using the LDA, PBE, and PBE-D2. The calculated bond lengths for the three functionals are in good agreement with the experimental data, and the errors are within 1%. However, we note that the LDA shows a slightly shorter bond length compared with experiments and that both the PBE and the

Table 1. Bond length (Å) and cohesive energy (eV/atom) of P₂ and P₄ molecules.

	Bond length	Cohesive energy
P ₂	LDA	1.887
	PBE	1.901
	PBE-D2	1.901
	Exp. [32]	1.893
P ₄	LDA	2.184
	PBE	2.202
	PBE-D2	2.202
	Exp. [33,34]	2.210

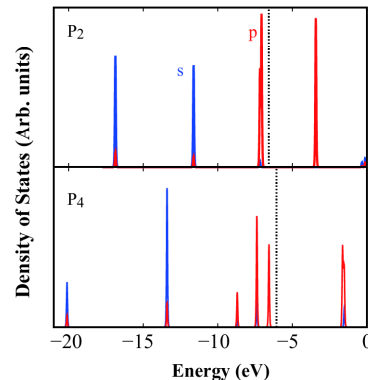


Fig. 1. (Color online) Electronic densities of states of P₂ and P₄ molecules. The blue and the red lines are for the 3s and the 3p orbitals, respectively. The energy is referred to the vacuum level, and the vertical dashed line indicates the Fermi level.

PBE-D2 yield slightly elongated bond lengths, confirming the typical tendency of overbinding in the LDA and reduced overbinding in the GGA. The cohesive energy is obtained by using the formula

$$E_{\text{coh}} = \frac{E_{\text{cluster}}}{N_{\text{cluster}}} - E_{\text{atom}},$$

where E_{cluster} and E_{atom} are the total energies of the cluster and of the isolated spin-polarized atom, respectively, and N_{cluster} is the number of P atoms in the cluster or molecule. The tetrahedral P₄ molecule is more stable than P₂ by 0.86 and 0.73 eV/atom for the LDA and the PBE (and PBE-D2), respectively, in accordance with a previous study [31]. Thus, forming a P₄ molecule is preferred to forming two P₂ molecules. The virtually same results in bond length and cohesive energy with and without dispersion correction indicate that the vdW interaction is not significant at the short P-P distance (~ 2 Å) where the chemical bond is formed.

Figure 1 shows the angular-momentum-decomposed partial DOSs for both P₂ and P₄. From the PDOS of P₂, one can tell that the 3s and the 3p states do not hybridize

Precursor State of Oxygen Molecules on the Si(001) Surface during the Initial Room-temperature Adsorption

Eunkyung HWANG, Yun Hee CHANG, Yong-Sung KIM and Ja-Yong KOO*
Korea Research Institute of Standards and Science, Daejeon 305-600, Korea

Hanchul KIM†
Department of Physics, Sookmyung Women's University, Seoul 140-742, Korea

(Received 2 May 2012, in final form 6 August 2012)

The initial adsorption of oxygen molecules on Si(001) is investigated at room temperature. The scanning tunneling microscopy images reveal a unique bright O₂-induced feature. The very initial sticking coefficient of O₂ below 0.04 Langmuir is measured to be ~ 0.16 . Upon thermal annealing at 250 – 600 °C, the bright O₂-induced feature is destroyed, and the Si(001) surface is covered with dark depressions that seem to be oxidized structures with –Si–O–Si– bonds. This suggests that the observed bright O₂-induced feature is an intermediate precursor state that may be either a silanone species or a molecular adsorption structure.

PACS numbers: 68.43.Hn, 68.43.Mn

Keywords: Silicon surface, Oxygen molecule, Precursor, Initial adsorption, Sticking Coefficient, Room temperature

DOI: 10.3938/jkps.61.1046

I. INTRODUCTION

A structure with thin insulating layers of SiO₂ on the Si(001) wafer is an essential component of current electronic devices, and the initial oxidation of the Si(001) surface has been studied intensively for decades. The process of dry oxidation uses O₂ molecules to oxidize the silicon surface at high temperatures. The dry oxidation of Si(001) requires the dissociation of O₂ molecules and a subsequent interaction between O and Si atoms. Hence, the initial step of dry oxidation would be either molecular adsorption followed by O₂ dissociation or barrierless dissociative adsorption.

For O₂ molecules with thermal energy corresponding to room temperature (RT), the initial sticking coefficient on the Si(001) surface has been measured by several groups without consensus, and the measured values are scattered over a wide range of 10^{-4} – 10^{-1} [1–5]. The initial sticking coefficient is known to depend on both the kinetic energy of the incident molecules and the substrate temperature, suggesting the existence of either a precursor or an intermediate state that traps incident O₂ molecules [2–4, 6–9]. Similar conclusions have been drawn from studies using infrared absorption [5], high-resolution electron-energy-loss spectroscopy [10], and

ultraviolet photoemission spectroscopy combined with photon stimulated desorption [11]. This has been supported by theoretical calculations that found metastable intermediate adsorption structures for O₂ on Si(001) [12–16]. On the other hand, a scanning reflection electron microscopy experiment claimed a barrierless oxidation of the first subsurface layer at RT under a high O₂ partial pressure [17]. Subsequent theoretical calculations proposed the existence of narrow channels for barrierless oxidation [18–20].

Scanning tunneling microscopy (STM) experiments reported that both bright and dark features were created by dosing with O₂ molecules and the bright protrusions were interpreted as ejected Si dimers [21–25]. In addition, type-*C* defects [26,27] were proposed as preferred reaction sites of O₂ molecules, and the pristine Si(001)-(2×1) surface was shown to be resistant against oxidation [16]. The inertness of a flat terrace was supported by a recent STM experiment that demonstrated selective reaction of O₂ molecules with a step edge at low temperatures near 80 K [28,29].

The Si(001)-(2×1) surface can have many extrinsic point defects, such as type-*C* defects and contamination-induced dimer vacancies [30], as well as the inevitable step edges [31]. Since point and line defects can be reactive with O₂ molecules and the O₂ partial pressure may affect the oxidation mechanism, the behavior of O₂ molecules on a flat, defect-free terrace with a very low O₂ coverage should be reinvestigated in order to under-

*E-mail: koojayon@kriss.re.kr

†E-mail: hanchul@sookmyung.ac.kr

stand the intrinsic oxidation process of the Si(001)-(2×1) surface.

In this work, we investigated the initial adsorption of O₂ molecules on a clean Si(001)-(2×1) surface quantitatively by using STM. We obtained high-resolution RT STM images and observed a single well-localized bright feature resulting from the initial O₂-Si(001) interaction. Then, we performed repetitive STM measurements at RT by varying the O₂ dosage to estimate the initial sticking coefficient. Finally, we annealed the O₂-dosed Si(001) surface and found that the initial STM features transformed to dark depressions. These observations are indicative of the existence of an intermediate precursor state in the dry oxidation of Si(001), and we propose a plausible molecular adsorption structure, in addition to the previously proposed silanone (Si=O) structures, for the precursor state.

II. EXPERIMENTAL

By using STM, we performed experiments in two connected UHV chambers with a base pressure $<1 \times 10^{-10}$ Torr. All the sample preparations were performed in a preparation chamber blocked off from the STM chamber by a gate valve because even a small partial pressure of O₂ molecules can damage the tungsten tip of the scanning tunneling microscope severely. The samples were phosphorus-doped Si(001) wafers with a resistivity of $1 - 10 \Omega\cdot\text{cm}$. Atomically flat Si(001)-(2×1) surfaces with a point-defect density below 0.5% were obtained by repeated flashing at temperatures up to 1200 °C. To prevent the O₂ molecules from encountering point defects such as type-*C* defects, we exposed the Si(001)-(2×1) surface to a tiny amount of O₂. The O₂ molecules were introduced through a leak valve and backfilled the preparation chamber for a predetermined time period keeping the pressure at 5×10^{-10} Torr. After the dose of O₂ molecules and other necessary processes such as thermal annealing, the sample was transferred to the measurement chamber and the surface was measured by using STM at RT.

III. RESULTS AND DISCUSSION

Figures 1(a) and 1(b) show the empty-state STM images of the Si(001)-(2×1) surface after dosing at 0.03 L and 0.2 L of O₂ molecules, respectively, where 1 L corresponds to 10^{-6} Torr-sec. A unique bright protrusion is commonly observed in both images. It is much brighter than the type-*C* defect in the empty state and is hardly visible at sample bias voltages smaller than 2 V. The density of the bright protrusions increases with increasing O₂ dosage whereas that of point defects such as type-*C* defects and dimer vacancies are scarcely affected by

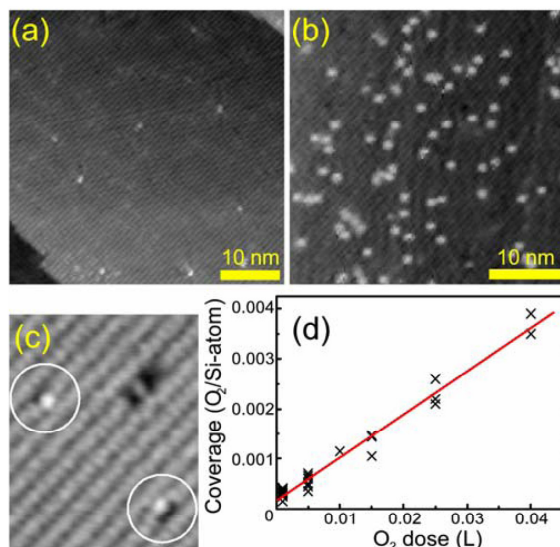


Fig. 1. (Color online) (a) and (b) Empty-state STM images of Si(001)-(2×1) surface dosed to 0.03 L and 0.2 L of O₂ molecules, respectively, at the sample bias voltage $V_s = 2$ V and the tunneling current $I_T = 0.2$ nA. (c) Filled-state image ($V_s = -2$ V) showing two bright O₂-induced features oriented in opposite directions. (d) The density of O₂-induced feature increases linearly with the O₂ dosage up to 0.04 L.

O₂ deposition. This implies that O₂ molecules do not encounter those point defects and guarantees that our investigation genuinely addresses the intrinsic properties of the O₂-Si(001)-(2×1) reaction.

The filled-state STM image in Fig. 1(c) shows better resolution than the empty-state images, and the O₂-induced feature can be characterized by using the following four observations: (i) It appears as a bright protrusion in both filled and empty states, (ii) two dimer sites are affected, (iii) two successive Si sites on the same side of a dimer row appear bright in contrast to the dark depression on the opposite side, and (iv) the bright side shows asymmetry, with one Si site being brighter than the other.

The plot in Fig. 1(d) shows the linear relation between the number of O₂-induced bright features and the amount of O₂ dosed on the surface. The sticking probability measured from Fig. 1(d) is 0.086 ML/L, where 1 ML stands for one molecule per surface Si atom. Estimating 1 L of O₂ dosage as 0.53 ML of O₂ [1], the sticking coefficient of O₂ molecules with thermal energy corresponding to RT on the Si(001)-(2×1) surface is determined to be ~ 0.16 .

Our observation is apparently in contrast with those of previous STM studies [21–24] that show dark depressions as well as bright bumps of ~ 1.4 Å in height [similar to the heights of monatomic steps on Si(001)]. Such an apparent discrepancy seems to originate from the differ-

Optical and Electronic Properties of Hydrogenated Silicon Nanoclusters and Nitrogen Passivated Silicon Nanoclusters: A Density Functional Theory Study

Seung Mi Lee^{1,*}, Kyung Joong Kim¹, Dae Won Moon¹, and Hanchul Kim^{2,*}

¹Korea Research Institute of Standards and Science, Daejeon 305-340, Korea

²Department of Physics, Sookmyung Women's University, Seoul 140-742, Korea

Silicon nanoclusters have become significant research interest due to their potential application to optoelectronic devices in visible range. We investigate the electronic and optical properties of hydrogenated and nitrogen-passivated silicon nanoclusters using density functional theory calculations. The energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of nanoclusters have varying sizes. They are systematically studied using the conventional local density approximation, the generalized gradient approximation, and the time-dependent density functional theory calculations with hybrid functional. The HOMO-LUMO gap is found to decrease monotonically as the size of nanocluster increases. Introducing one and two nitrogen passivants to a $\text{Si}_{29}\text{H}_{36}$ nanocluster, we find that the HOMO-LUMO gap decreases as the number of nitrogen passivants increases. It suggests that multi-nitrogen passivants may enable light emission in visible range from smaller clusters.

Keywords: Silicon Nanocluster, Electronic Structure, Surface Passivation Effect.

1. INTRODUCTION

Silicon (Si) has been used for electronic and memory devices due to its good electronic properties and well-established fabrication processes. The application to optical devices using Si has difficulties since it has indirect band gap. Semiconductor materials have new material properties when their size is reduced to nanoscale due to the quantum confinement effect. Different from the bulk structure, nanoscale Si and porous Si were reported to have optical emission near infrared regime. The band gap of Si nanoclusters increases as the cluster size decreases, and the frequency range covers all visible range,^{1,2} which opens the possibility of Si-based optical devices.

The high-temperature annealing of oxide silicon (SiO_x , $x < 2$) results in Si nanoclusters embedded within the SiO_2 matrix. Controlling the relative amount of oxygen, x , the peak position of optical emission was changed from near infrared to red.³ It was qualitatively explained by the quantum confinement effect of embedded Si nanoclusters and the local distortion of Si sp^3 bonding character by oxygen adatom.⁴ Although there are several theoretical reports,^{2,4,5} quantitative explanations are still unavailable. In the same manner, high-temperature annealing of SiN_x

($x < 1.33$) results in embedded Si nanoclusters in the Si_3N_4 matrix. As the ratio of Si to N increases, the emission peak position is shifted to lower energy side covering all visible range of red to violet,⁶ suggesting the possibility of Si-based optoelectronic device.

With the intention of comprehending the Si-based optical devices, the physical properties in atomistic level are essential. Density functional theory (DFT) calculations have been successfully applied for explanation and prediction of electronic properties, surface reconstruction, atomistic diffusion, adsorption, and desorption of semiconductor materials. Since the underestimation of energy gap is a well-known shortcoming of conventional exchange-correlation functionals, we have employed hybrid functional to get the exact value of energy gap for important geometries. In this report, we show the stable geometries and band gaps of Si nanoclusters with or without nitrogen adatom using DFT calculations. The size dependence and the effect of nitrogen adatom on electronic and optical structure of Si nanoclusters are discussed.

2. COMPUTATIONAL DETAILS

We performed quantum mechanical calculations within DFT framework using atomic orbital basis set calculations

*Authors to whom correspondence should be addressed.

of DMOL3 code⁷ and Gaussian03.⁸ The exchange-correlation functionals of local density approximation (LDA)⁹ and generalized gradient approximation (GGA)¹⁰ were used as implemented in DMOL3 code. We first optimized each possible geometry using LDA and GGA functionals and defined the most stable structure. Then, the most stable geometry was reoptimized using hybrid functional to get the exact value of the energy gap. We use time-dependent DFT with the B3LYP hybrid functional as implemented in Gaussian03 package. For basis sets, a double-numeric polarized basis set was used for DMOL3 calculations while various basis sets were tested for Gaussian03 code as described later. All dangling bonds of silicon nanoclusters were saturated by hydrogen atoms and the atomic structures were relaxed until the force on each atom was less than 0.1 eV/Å, the maximum displacement was 0.005 Å, and the total energy change is less than 5×10^{-5} eV.

3. RESULTS AND DISCUSSION

The experimentally observed size of silicon nanoclusters was about 5 nm, while the Bohr radius of silicon for quantum confinement effect was 4.9 nm⁶. To increase the calculation efficiency, we investigated the T_d symmetric silicon nanoclusters with a size of up to 3 nm. For nitrogen-adsorbed systems, we investigated a silicon nanocluster of 1 nm.

First, we investigated the smallest Si nanocluster, the SiH_4 molecule. After optimization, the bond length of Si–H was 1.492 Å (LDA) and 1.487 Å (GGA), respectively. The dependency of GGA functional of PW91, PBE, and BLYP was found to be negligible. As the next step, we systematically calculated the highest occupied molecular orbitals (HOMO) and the lowest unoccupied molecular orbitals (LUMO) gap, using hybrid functional to get the exact value. Table I presents the excitation energy and HOMO-LUMO gap for several basis sets. From systematic investigations, we decided to use 6–311 g++(2d,2p)

Table I. Calculated HOMO-LUMO gap and first excitation energy of a SiH_4 molecule by B3LYP hybrid functional and various basis set.

Name of basis set	Number of basis function	HOMO-LUMO gap (eV)	First excitation energy (eV) [oscillator strength]
SVP	38	10.9025	10.4665 [0.1175]
6-31g(d,p)	39	10.9523	10.8383 [0.7845]
6-31g++(d,p)	47	9.5163	8.6464 [0.1304]
6-31 g(2d,2p)	57	10.8894	10.5475 [0.0468]
6-31 g+(2d,2p)	61	9.6374	8.7311 [0.1358]
6-31 g++(2d,2p)	65	9.5207	8.5672 [0.1259]
6-311 g++(2d,2p)	75	9.5550	8.5562 [0.1169]
6-31 g++(3df,3p)	114	9.5460	8.5214 [0.1182]
AUG-cc-pVTZ	142	9.4026	8.4824 [0.1052]
AUG-cc-pVQZ	268	9.3890	8.5722 [0.1051]
AUG-cc-pV5Z	451	9.3514	8.5587 [0.1004]

basis set for further calculations, because it showed accuracy compared to the experimental value¹¹ and reasonable computational efficiency.

The HOMO-LUMO gap of optimized structures of hydrogenated Si nanoclusters is illustrated in Figure 1 as a function of the cluster size. The geometries of T_d symmetric nanoclusters of SiH_4 , Si_5H_{12} , $\text{Si}_{17}\text{H}_{36}$, $\text{Si}_{41}\text{H}_{60}$, $\text{Si}_{83}\text{H}_{128}$, $\text{Si}_{147}\text{H}_{148}$, and $\text{Si}_{239}\text{H}_{220}$ were studied. Points represent the calculated value while the dotted and dashed lines show the extrapolated values. All three different functionals of LDA, GGA, and hybrid functionals show the same trend. The HOMO-LUMO gap increases as the cluster size decreases. The LDA and GGA functionals give smaller value compared to hybrid functional, as expected.

To reveal the nitrogen passivant effect, we take T_d symmetric $\text{Si}_{29}\text{H}_{36}$ cluster with diameter of 1.06 nm. Considering substitutional sites with surface hydrogen atoms, a single nitrogen atom can substitute one, two, and three hydrogen atom (s). Geometrically possible substitutional sites are depicted in Figure 2(a). Two different cases are cited for one nitrogen atom substituting a hydrogen atom according to the number of Si neighbors of hydrogen; a single silicon atom compared in Figure 2(a) and two Si atoms compared in Figure 2(a). For the case of an N atom replacing two neighboring hydrogen atoms to form bonds with two Si atoms, two possible configurations, *D1* and *D2*, are cited in Figure 2(a). For the case of three neighboring hydrogen atoms, being replaced by an N atom, two different possible geometries, *T1* and *T2*, are indicated in Figure 2(a).

We performed geometry optimization for the six initial geometries within DFT calculations. Since the atom numbers of each species are different for each system, one cannot simply compare the total energy to identify relative stability. We then calculated the binding energy E_{bind} as $[E_{\text{total}} - n_{\text{Si}}\mu_{\text{Si}} - n_{\text{H}}\mu_{\text{H}} - n_{\text{N}}\mu_{\text{N}}]/n_{\text{total}}$, where E_{total}

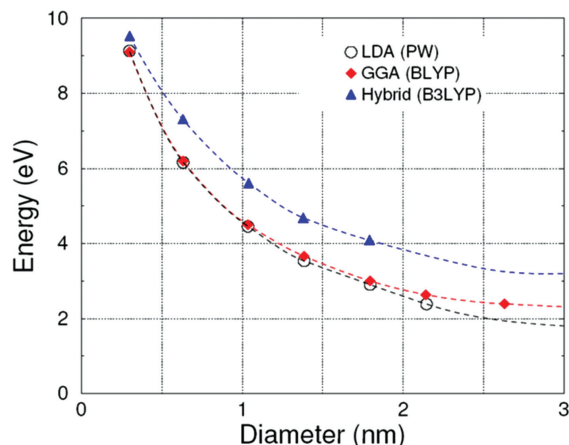


Fig. 1. HOMO-LUMO gap of silicon nanoclusters calculated with LDA-PW (open circle, black), GGA-BLYP (filled diamond, red), and hybrid-B3LYP (filled triangle, blue) functionals, respectively.

Theoretical investigation on the magnetic phase stability of Fe-doped Bi tellurides

Miyoung Kim^{1,a)} and Jung-Hwan Song²

¹*Division of Energy System Research, Ajou University, Suwon 443-749, Korea*

²*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60203, USA*

(Presented 1 November 2011; received 23 September 2011; accepted 19 October 2011; published online 14 February 2012)

The structural properties and magnetic phase stability of Fe-doped Bi₂Te₃ are studied in first principles by means of the highly precise full-potential linearized augmented plane-wave method within the generalized gradient approximation. Two different types of Fe doping, which are substitutional and interstitial, are examined, assuming about 8% of the Fe doping level, where the atomic structures are fully optimized by atomic force calculations. Our results show that: (i) ferromagnetic phases are energetically more stable over paramagnetic phases for both substitutional and interstitial Fe dopings, (ii) the substitutional Fe doping indicates a high magnetic phase with a largely enhanced Fe magnetic moment of 3.40 μ_B while the interstitial Fe doping shows the magnetic moment similar to the Fe bulk value, and (iii) the structural optimization reduces the Fe magnetic moments for both doping types while the overall effects on the total moment exhibit different results for different doping types. © 2012 American Institute of Physics. [doi:10.1063/1.3671777]

The Bi₂Te₃ alloys are well-known conventional materials for thermoelectric (TE) applications with the peak figure of merit (ZT) ~ 1.0 near room temperature.¹ To take advantage of the energy conversion directly from a variety of thermal sources via the TE mechanism, there have been enormous efforts to develop highly efficient TE materials with a high ZT exceeding that of the Bi₂Te₃ alloys over the last decade. Noting that the efficiency of the TE materials is determined by the figure of merit, $ZT = S^2\sigma/(\kappa_e + \kappa_L)$, where S is the Seebeck coefficient, σ is the electrical conductivity, and κ_e and κ_L are the electronic and lattice thermal conductivities, the majority of experimental studies have been carried out using the phonon blocking/electron transmitting mechanism. While an impressive ZT value of ~ 2.4 was reported in p -type Bi₂Te₃/Sb₂Te₃ superlattices almost 10 years ago,² the low dimensional thin films or nanostructure TE systems with relatively high ZT values are difficult to be put into practical use due to the thermal instability. Therefore, a search for an efficient and practically applicable TE material still remains a challenge where Bi₂Te₃ is playing an important role as one of the key ingredients.

Recently, special interest in Bi telluride and selenide has also been revived due to their unique electronic structures exhibiting a Dirac-like conducting band as the bulk topological insulators (TI).^{3–6} One particular interest in manipulating the Dirac-like conducting surface states is on the effect of magnetic impurities and ferromagnetism on the TI surface states.^{7–9} Experimentally, Mn doped Bi tellurides, Bi_{2–x}Mn_xTe₃, are found to be ferromagnetic for the doping level, $0.04 < x < 0.09$,^{8,10} and a weak ferromagnetism was reported for Bi_{2–x}Fe_xTe₃ (Ref. 11) while a theoretical report is

still lacking. Noting that the magnetic phase dependence, not only on the composition but also on the atomic structure or structural relaxation, has been reported for various transition metal alloys, a study on the magnetic stability for different doping types and structure relaxation is also desirable for the transition metal doped Bi tellurides.

In this research, we investigate the electronic structures and magnetic properties of the Fe-doped Bi tellurides with an 8% doping level, aiming to understand the magnetic phase stability and its dependence on the doping type and the atomic structural optimization due to relaxation. For this study, we considered two different types of Fe doping, *substitutional* and *interstitial*, based on a hexagonal Bi₂Te₃ structure, as shown in Fig. 1. For *substitutional* doping (Fig. 1(a)), two Bi atoms in a 60-atoms unit cell are substitutionally replaced by Fe atoms, while for the *interstitial* doping (Fig. 1(b)), two doped Fe atoms are located in the interstitial region between the Te layers. Assuming the fixed lattice constants with experimental values ($a = 4.386$ Å and $c = 30.497$ Å),¹² the internal atomic positions were fully relaxed via atomic force calculation to obtain the structural optimization. For each of these doping types, we examined two magnetic phases; the ferromagnetic (FM) and the paramagnetic (PM) phase. The electronic structures were calculated using the all-electron full-potential linearized augmented plane-wave (FLAPW) method,¹³ employing the exchange-correlation potential provided by the generalized gradient approximation.¹⁴ An energy cutoff of 13 Ry for the plane-wave basis and the special k-point samplings of 25 (*substitutional*) and 13 (*interstitial*) k-points in the irreducible Brillouin zone for k-space integration were employed.

In Table I, we showed our calculated result of the total energy differences between the FM and PM phases, $\Delta E = E_{\text{FM}} - E_{\text{PM}}$, in eV/(unit-cell). A total energy

^{a)}Author to whom correspondence should be addressed. Electronic mail: mykim.nu@gmail.com.

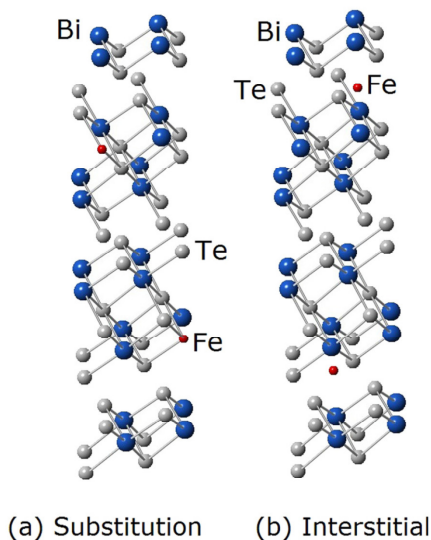


FIG. 1. (Color online) The atomic structure of two different types of Fe doping: (a) substitutional, and (b) interstitial. The 8% Fe doped Bi_2Te_3 alloy is simulated by replacing two Bi atoms of a 60-atom hexagonal unit cell substitutionally by Fe atoms for (a), while the doped Fe atoms are located in the interstitial region for (b).

comparison revealed that the FM is energetically favored over the PM phases for both dopings. The results also show that ΔE of the *substitutional* doping is three times larger than that of the *interstitial* doping, implying that the FM phase stability is relatively stronger when the Fe impurity is doped by replacing Bi atoms rather than by locating in the interstitial region. Therefore, while an 8% Fe doped $\text{Bi}_{2-x}\text{Fe}_x\text{Te}_3$ alloy prefers the FM phase, the stability can be diminished if the interstitial Fe impurities increase, which might be caused by an increased doping level or high temperature.

In Table I the calculated results of the Fe atomic spin moment and the total spin moment, which includes all atomic as well as interstitial moments in one unit-cell, are also shown. As can be seen, the substitutionally doped Fe exhibits a largely enhanced moment of $3.40 \mu_B$ while the interstitially doped Fe has a relatively small moment of $2.38 \mu_B$, which is close to its bulk value. The atom-, orbital-, and spin-projected density of states (DOS) of the valence states are calculated and presented in Fig. 2. The Fe 3d DOS

TABLE I. The calculated results of total energy difference between the FM and PM phases, $\Delta E = E_{\text{FM}} - E_{\text{PM}}$ [in eV/(unit-cell)] are shown. Note that negative ΔE means that the FM phase is energetically more stable than the PM phase. Also, the Fe atomic spin moments and total spin moments, including all atomic and interstitial moments in one unit-cell [in μ_B] for the fully optimized (after OPT) FM phase, are shown. Spin moments before optimization (before OPT) are also given for comparison.

	ΔE (eV/unit-cell)	Fe spin moments (μ_B)		Total spin moments (μ_B)	
		Before OPT	After OPT	Before OPT	After OPT
Substitutional	-4.87	3.54	3.40	9.45	9.95
Interstitial	-1.63	2.41	2.38	4.59	4.39

for substitutional doping certainly shows the enhanced DOS peaks for both spins and a large exchange splitting between the spin-up and spin-down DOS peaks compared to the interstitial doping, which results in a larger Fe magnetic moment. Although not shown in the table, small induced magnetic moments are also observed for the Bi and Te atoms, which are 0 to 0.04 (Bi) and 0 to 0.09 (Te) μ_B for substitutional doping and 0 to -0.03 (Bi) and 0.01 to -0.01 (Te) μ_B for interstitial doping. These induced moments are due to the (small) spin polarization of *p*-states which can be seen in Fig. 2 where the spin-projected DOS for the topmost valence *s*- and *p*-states of Bi and Te atoms are shown. Among the many atoms in the one unit-cell, we show the DOS of the Bi and Te atoms having the largest value of induced magnetic moments, which are the closest to the Fe atoms. While very small, the difference between the spin-up and spin-down *p*-DOS can be found. It is interesting to note that the Bi and Te moments show antiferromagnetic (AFM) coupling with the Fe moment for the interstitial doping, although the values are very small. The DOS plot indicates that the AFM coupling of Bi is due to a small development of the DOS peak occupied for the spin-down *p*-states, which does not exist for the spin-up DOS. As an overall result of small Fe moment and AFM coupling of Bi and Te, the total magnetic moment of the *interstitial* doping becomes only one half that of the *substitutional* doping.

In order to see the optimization effect on the magnetic property, the Fe and total spin moments before the optimization are given in Table I for comparison. We find that the Fe moment is reduced due to the optimization for both dopings, an effect which is more remarkable for the substitutional Fe. However, the total moment shows different optimization

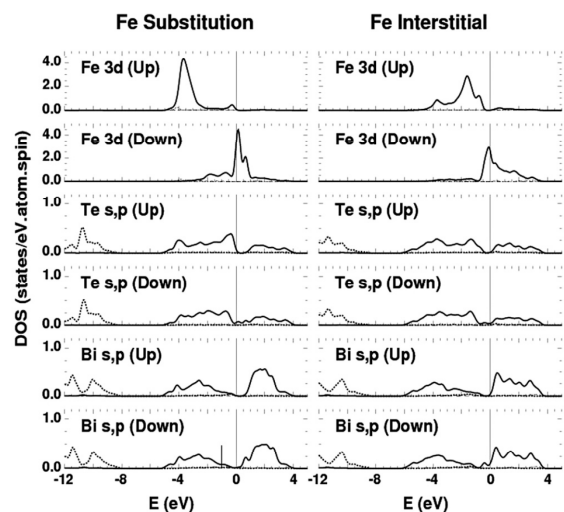


FIG. 2. The calculated atom-, orbital-, and spin-projected density of states (DOS) of the Fe-doped Bi_2Te_3 alloy for substitutional (left panels) and interstitial (right panels) doping. The DOS of the topmost valence states are shown, including the 3d states for Fe, the 5s (dotted lines) and 5p states (solid lines) for Te, and the 6s (dotted lines) and 6p (solid lines) states for Bi atoms. Among the many Bi and Te atoms in the unit-cell, we show the DOS of the representative atoms with the maximum value of the induced magnetic moments which are the closest to the Fe atoms.

Surface and Correlation Effect on the Half-metallicity of the Co_2FeSi Heusler Alloy

Miyoung KIM*

Division of Energy System Research, Ajou University, Suwon 443-749, Korea

Jae Il LEE

Department of Physics, Inha University, Incheon 402-751, Korea

(Received 30 January 2012, in final form 17 February 2012)

We report the *ab-initio* calculational results for the half-metallicity of the Co_2FeSi full Heusler alloy in (001) surfaces. Employing the +U corrections, in addition to the local density approximation (LDA) and generalized gradient approximation (GGA), we investigate the effects of surface and correlation interactions on the half-metallicity and the magnetic properties. The (001) surfaces are calculated to be metallic by both the LDA and the GGA due to the surface states the develop at the Fermi energy (E_F). Upon +U correction, the surface states are pushed down below E_F for the Co-terminated surface, which recovers the half-metallicity with a reduced band gap of 0.40 eV. In contrast, the FeSi-terminated surface is predicted to be metallic because the surface states remain at the Fermi level in spite of an enhanced t_{2g} - e_g band splitting caused by the correlation effect.

PACS numbers: 73.20.At, 71.20.Be

Keywords: Electronic structure, Half-metallicity, Transition-metal thin films, Surface states

DOI: 10.3938/jkps.60.1068

I. INTRODUCTION

Half metals are characterized by metallic electronic states for one spin channel while the other spin channel exhibits a finite band gap at the Fermi Energy (E_F), resulting in 100% spin polarization at E_F . This half-metallicity (HM) is one of the key ingredients for realization of spintronics device because it is known to maximize the spin-injection rate of spin-polarized carriers into semiconductors [1]. Among various classes of half metals, including magnetic oxides (CrO_2 and Fe_3O_4), colossal magnetoresistance materials in perovskites ($\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ and $\text{Sr}_2\text{FeMoO}_6$), zincblendes (CrAs and MnSb), and dilute magnetic semiconductors ($\text{Ga}_{1-x}\text{Mn}_x\text{As}$) [2], full Heusler alloys of the Co_2YZ type have drawn special attention because they also exhibit attractive properties, such as high Curie temperature (T_C), high magnetic moments, and good structural matching with conventional semiconductors in terms of crystal structure and lattice constant. While Co_2MnZ compounds have been extensively explored by theory and experiment [3], studies on Co_2FeZ type alloys are relatively lacking [4] in spite of their high T_C up to 1100 K and large magnetic moments of $6\mu_B$, which are the highest values for all known Heusler alloys [5].

Implementing the half metals into the devices raises another practical question: whether the HM is preserved at the half-metal/semiconductor interfaces. Actually, experiments report that the spin polarization and injection rates of half metals are far from theoretical predictions: the spin polarizations of epitaxially-grown Co_2MnGe and Co_2MnSi on GaAs substrates reveal to be 50 – 60% [6] and the spin injection rate from Co_2MnGe into the Al-GaAs/GaAs hetero-structure is reported to be only 27% at 2 K, which is even lower than the value of 40% reached with an Fe injector [7]. Therefore, it is important to clarify the interface dependence of the HM. In this context, the half-metallic surfaces are also an attractive research subject because they not only suffer abrupt change in electronic structures similarly with interfaces but also are more easily accessible experimentally than interfaces.

II. METHODOLOGY

In this paper, we present *ab-initio* electronic structure calculations for Co_2FeSi full Heusler alloy thin films with (001) surfaces within the density functional theory (DFT). The calculations were performed using the all-electron full-potential augmented plane wave (FLAPW) [8] method within the local density approximation (LDA), the general gradient approximation (GGA) [9]

*E-mail: mykim.nu@gmail.com

Table 1. Calculated equilibrium lattice constants and atomic spin magnetic moments for Co_2FeSi full Heusler alloys *in the bulk*. Spin minority band gaps are also given for the LDA+U and the GGA+U approaches.

	Lattice constant (Å)	Spin magnetic moments (μ_B)			Spin minority band gap (eV)
		Co	Fe	Si	
LDA	5.507	1.21	2.55	-0.01	-
LDA+U	5.565	1.52	3.15	-0.05	1.35
GGA	5.635	1.38	2.80	-0.01	-
GGA+U	5.705	1.57	3.20	-0.06	0.85

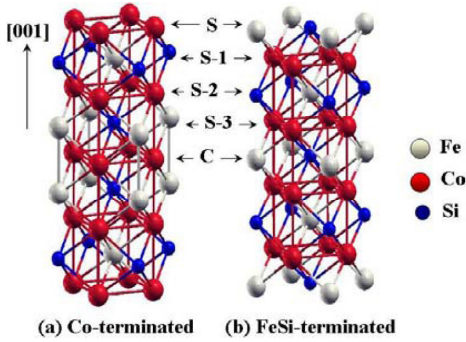


Fig. 1. (Color online) Atomic structures are shown for the freestanding Co_2FeSi (001) thin films in the $L2_1$ structure with nine atomic layers for the (a) Co-terminated and the (b) FeSi-terminated surfaces. The atomic layers are denoted from the center (C) to the surface (S).

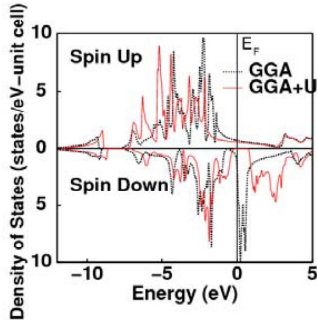


Fig. 2. (Color online) Calculated results of spin-resolved total DOS for Co_2FeSi *in the bulk* for spin majority (upper panel: spin up) and spin minority (lower panel: spin down) channels. The dotted and the solid lines indicate the DOS obtained using the GGA and the GGA+U, respectively.

and the +U corrections, that is, LDA+U and GGA+U. In the +U approach, the on-site Coulomb interaction characterized by a Hubbard U is implemented by using the second variation procedure self-consistently in solving the Kohn-Sham equation [10], adopting Coulomb (U) and exchange (J) energies of 5 and 1 eV, respectively, for 3d electrons of Fe and Co atoms. An energy cut off of

220 eV for the interstitial plane wave basis and $20 \times 20 \times 20$ (20×20) special k-point meshes are used for the 3D (2D) Brillouin Zone integral. The (001) surface is simulated by a free-standing thin film with nine atomic layers, as shown in Fig. 1. We examine two different surface terminations: a Co-terminated surface and an FeSi-terminated surface (*cf.* Figs. 1(a) and (b)). Vertical atomic positions are fully relaxed using the total energy and atomic force calculations assuming that the in-plane lattice constant is fixed to the equilibrium lattice of bulk alloy in the GGA. Note that the GGA gives the best agreement with the experimental lattice constant among the four different exchange-correlation potentials that are employed here, as explained below.

III. RESULTS AND DISCUSSION

For comparison, the electronic structure calculation for the Co_2FeSi *bulk alloy* is carried out, and the lattice constants calculated via total energy minimization and the atomic spin magnetic moments are listed in Table 1. The lattice constant is found to strongly depend on the exchange-correlation potentials. As usual for other transition metals, the GGA predicts a bigger equilibrium lattice constant than the LDA. The correlation effect through +U is found to increase the lattice constant by 1% for both the GGA and the LDA. The lattice constant is in best agreement with the experimental value for the GGA - with a difference of less than 0.1% - while it is underestimated by LDA and LDA+U and overestimated by the GGA+U. Calculated results also show that the magnetic moments are significantly affected by on-site Coulomb interaction. The atomic moments of transition metals are enhanced by 0.31 (Co) and 0.60 (Fe) μ_B for the LDA and by 0.19 (Co) and 0.40 (Fe) μ_B for the GGA. The induced moment for Si, which is negligible in the LDA and the GGA calculations, also increases up to $-0.06 \mu_B$ upon +U correction. As a result, total magnetic moment is calculated to be 6 μ_B for both the LDA+U and the GGA+U.

The integer value of the total magnetic moment indicates the half-metallicity. To confirm this, we show our calculated spin-resolved total density of states (DOS) plot for the GGA and the GGA+U in Fig. 2. As can be

Temperature and Carrier-concentration Dependences of the Thermoelectric Properties of Bismuth Selenide Dioxide Compounds

Tran Van QUANG

Department of Physics, Ajou University, Suwon 443-749, Korea

Hanjo LIM

Division of energy System Research, Ajou University, Suwon 443-749, Korea

Miyoung KIM*

Department of Nano Physics, Sookmyung Women's University, Seoul 140-742, Korea

(Received 8 December 2011, in final form 5 June 2012)

We performed an *ab-initio* electronic structure calculation for the bismuth selenide compound with oxygen substitution by using the all-electron full-potential linearized augmented plane-wave (FLAPW) method. The spin-orbit coupling in the second variational scheme and the screened-exchange local density approximation (sX-LDA) were adapted, and a large enhancement of the indirect band gap from 0.13 to 0.78 eV over the conventional LDA result was found. Exploiting the FLAPW wave-functions, we employed Boltzmann's equation in a constant relaxation-time approach to estimate the thermoelectric (TE) transport properties, including the Seebeck coefficient, the electrical conductivity and the power factor. The temperature and the carrier-concentration dependences of the TE properties were investigated in detail to find the effect of oxygen substitution on and to search for the optimal doping and temperature range for better thermoelectric performance.

PACS numbers: 71.20.-b

Keywords: First-principles calculation, Thermoelectric material, FLAPW, Bismuth selenide dioxide

DOI: 10.3938/jkps.61.1728

I. INTRODUCTION

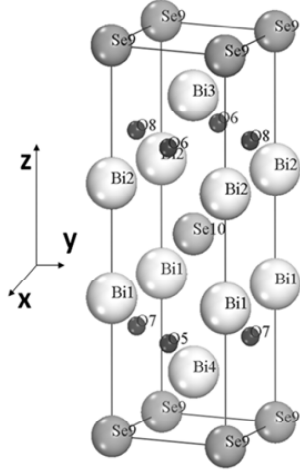
Bismuth selenide is one of the most investigated materials as a component for an n-type leg in thermoelectric (TE) applications. In spite of the huge efforts to improve the TE properties by introducing various nanostructures, dopings, or defects, the figure of merit, which is defined by $ZT = \sigma S^2 T / (\kappa_e + \kappa_L)$ where S is the Seebeck coefficient, σ is the electrical conductivity, and κ_e , and κ_L are the electronic and the lattice thermal conductivities, respectively, has been scarcely increased from its original value [1]. As can be seen from the definition, the efficiency of a TE material can be improved by decreasing the thermal conductivity while keeping the electrical conductivity unchanged (or even increased) or vice versa. In this conjecture, a recent experimental by Ruleova *et al.*, which reported that the thermal conductivity of Bi_2Se_3 was significantly reduced by substituting oxygen atoms to generate the $\text{Bi}_2\text{O}_2\text{Se}$ compound, is worthy of notice [2]. While the TE property was not much enhanced due

to a decrease in the electrical conductivity, they found that the oxygen-substituted bismuth-selenide compound still provided a good opportunity to obtain a high TE coefficient by searching for a possible mechanism to improve the electrical conductivity.

Therefore, in this study, we carried out *ab-initio* calculations to investigate the electronic structures of the $\text{Bi}_2\text{O}_2\text{Se}$ compound and its TE properties. The electronic structure was calculated via the highly-precise all-electron full-potential linearized augmented plane wave (FLAPW) method [3] in the local density approximation (LDA) where the spin-orbit coupling (SOC) was included in the second variational way [4,5]. In addition, the screened-exchange LDA (sX-LDA) [6] was used to obtain a more accurate description of the band gap and the excited states. The sX-LDA method is known to improve the underestimated LDA band gaps for most semiconductors [7,8] as well as the bulk Bi_2Te_3 , where the sX-LDA band gap (0.154 eV) was found to be in good agreement with the experimental band gap (0.162 eV) [9].

To calculate the TE coefficients, we estimate the Seebeck coefficient and the electrical conductivity by uti-

*E-mail: mykim.nu@gmail.com

Fig. 1. Atomic structure of the $\text{Bi}_2\text{O}_2\text{Se}$ compound.

lizing the distribution function based on Boltzmann's equation in the constant relation-time approximation. The group velocity included in the transport coefficients is calculated from the full intraband optical matrix element defined within the FLAPW method. Consequently, the electrical conductivity and the Seebeck coefficient are calculated as [10,11]

$$L_{\alpha\beta}^{(\nu)} = e^2 \tau \int \frac{d\varepsilon d\vec{k}}{8\pi^3} \left(-\frac{\partial f_0}{\partial \varepsilon} \right) (\varepsilon - \mu)^\nu \delta(\varepsilon - \varepsilon_{\vec{k}}) \nu_{\vec{k}}^\alpha \nu_{\vec{k}}^\beta, \quad (1)$$

$$\sigma_{\alpha\beta} = L_{\alpha\beta}^{(0)}, \quad S_{\alpha\beta} = -\frac{1}{eT} (L_{\alpha\beta}^{(0)})^{-1} L_{\alpha\beta}^{(1)}, \quad (2)$$

where τ , f_0 , and $\nu_{\vec{k}}$ denote the relaxation time, Fermi distribution function, and group velocity, respectively; e denotes the elementary charge; k_B Boltzmann's constant; μ the chemical potential; T the temperature; and $\delta(\varepsilon - \varepsilon_{\vec{k}})$ the Dirac delta function. Then, the power factor is computed by σS^2 . While most previous reports on the bismuth selenides dealt with n-type doping, we calculate the TE transport coefficients in both n- and p-type dopings.

II. ELECTRONIC STRUCTURES

The crystal structure of $\text{Bi}_2\text{O}_2\text{Se}$ is experimentally observed to be tetragonal with space group $I4/mmm$ which has 10 types of atoms in the unit cell with 8 atomic layers along the z-direction as shown in Fig. 1. Two different atomic coordinates have been experimentally reported for the $\text{Bi}_2\text{O}_2\text{Se}$ compound; one by Schmidt *et al.* [12] and the other by H. Boller [13]. Upon LDA calculations adopting these two sets of experimental coordinates, we find that the former results in a semiconductor while the

Table 1. Calculated band gap E_g (eV), valence energy split-off Δ_1 (eV), and conduction energy split-off Δ_2 (eV) within LDA, LDA-SOC, and sX-LDA and sX-LDA-SOC.

	LDA	LDA-SOC	sX-LDA	sX-LDA-SOC
E_g (eV)	0.34	0.13	1.22	0.78
Δ_1 (eV)	0.33	0.33	0.39	0.33
Δ_2 (eV)	0.60	0.19	0.86	0.39

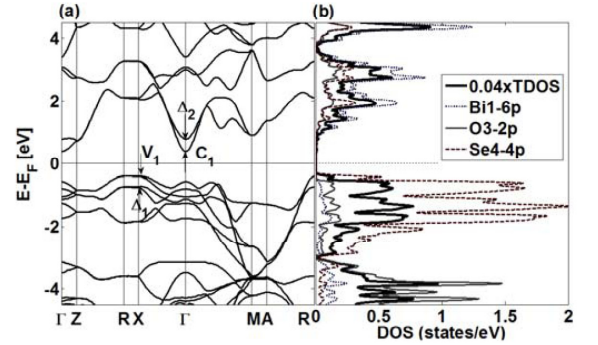


Fig. 2. Calculated band structure of $\text{Bi}_2\text{O}_2\text{Se}$ is shown in the left panel. The highest valence and the lowest conduction bands are indicated by V_1 and C_1 , respectively. Δ_1 is valence energy split-off and Δ_2 is the conduction energy split-off. The densities of states are shown in the right panel: total density of states (TDOS) rescaled by factor of 0.04 (bold solid line) and the projected densities of states of Bi1-6p (dotted line), O3-2p (dash-dotted line), and Se4-4p (dashed line) within sX-LDA with SOC. Atomic indices are as denoted in Fig. 1.

later results in a metal, and the semiconducting phase is energetically more stable over the metallic phase. In the semiconducting phase, through further total energy calculations, we find that the optimized lattice constants are $a_{\text{opt}} = 3.85 \text{ \AA}$ and $c_{\text{opt}} = 11.94 \text{ \AA}$. These results are in good with experimental values reported by Schmidt *et al.*, $a_{\text{exp}} = 3.88 \text{ \AA}$ and $c_{\text{exp}} = 12.16 \text{ \AA}$ [12].

The calculated band structure and the atom- and the orbital-projected densities of states (DOSs) within the sX-LDA with SOC are presented in Fig. 2. As can be seen, the sX-LDA band gap is found to be 0.78 eV. The valence band split-off energy indicated by Δ_1 in the figure is 0.33 eV, and the conduction-band split-off energy, Δ_2 , is 0.39 eV. For comparison, the calculated values of the band gap and the split-off energies of the valence, as well as the conduction, band for the LDA with and without SOC are given in Table 1. Within the LDA calculation without SOC, the band gap is found to be 0.34 eV. When SOC is included in the LDA (LDA-SOC), the band gap is reduced to 0.13 eV. Now, adapting the sX-LDA, the band gap is enhanced significantly to 1.22 eV without SOC (sX-LDA), which is finally reduced to 0.78 eV with SOC (sX-LDA-SOC). As was found for bulk Bi_2Se_3 , the band gap is reduced due to the SOC effect while the

화학과학

Atomic-level investigations on the amyloid- β dimerization process and its driving forces in water†

Song-Ho Chong and Sihyun Ham*

Received 21st October 2011, Accepted 5th December 2011

DOI: 10.1039/c2cp23326f

We report the spontaneous dimerization process of the full-length A β 42 proteins in water by using unguided, fully atomistic, explicit-water molecular dynamics simulations. Based on the thermodynamic analysis, we demonstrate that A β 42 dimerization in water occurs via a two-step nucleation–accommodation mechanism driven by water-induced force and by protein internal force, respectively.

Aggregation of misfolded amyloid-beta (A β) proteins and subsequent amyloid-fibril formation have been considered as a primary cause of Alzheimer's disease (AD).¹ An increasing number of studies indicate that soluble A β oligomers formed prior to amyloid fibrils are the main pathogenic agents in AD,^{2,3} leading to more focus on early stages of aggregation processes.⁴ The A β dimer is particularly important since it is the smallest oligomeric species that potentially impairs synapse structure and function in the AD brain as revealed by a recent clinical study.⁵ Yet, experimental investigations on the dimerization process of A β proteins in a physiological aqueous environment have been hindered by their strong tendency to aggregate in water.⁶ Computational approaches on the spontaneous dimerization of the full-length A β proteins also remain challenging due to heavy computational load, and previous studies have been carried out for short fragments of A β proteins⁷ or for the A β dimerization in a continuum water model.^{8,9} However, the importance of explicit treatment of water has been suggested to elucidate the role of water in protein self-assembly.^{10–12}

Here, we report the computational results from unguided, fully atomistic, explicit-water molecular dynamics (MD) simulations on the dimer formation of the full-length, 42-residue form of A β proteins (A β 42) as well as from the solvation free energy calculation for the simulated dimer conformations based on the integral-equation theory of liquids.^{13,14} From the simulation pathway analysis of the whole dimerization process starting from two isolated A β 42 monomers to the formation of a compact A β 42 dimer structure, we show that A β 42 dimerization occurs

via two steps, the initial nucleation stage followed by the structural accommodation stage. To elucidate the driving force in each stage, the solvation free energy along with the protein internal energy were computed. We demonstrate that those two stages are, respectively, driven by the water-induced force and the protein internal force. The atomic-level description of the A β 42 dimerization pathway and the thermodynamic understanding of the driving force in each stage of the dimerization process may be critical for the identification of therapeutic targets of AD.

We conducted A β 42 dimerization simulations at 300 K and 1 bar under neutral pH where each A β 42 monomer is in the -3 charged state due to its amino acid composition. The initial monomer structure (Fig. 1) was taken from our previous study, where the conformational transitions of the A β 42 monomer that occur in water prior to aggregation were investigated by the extensive all-atom, explicit-water MD simulations.¹⁵ Two monomers were initially placed at 45 Å apart from each other with a random orientation, which were solvated with 24 708 explicit water molecules. Six Na⁺ counter ions were added to neutralize the system. No artificial attraction force was employed between the two monomers. (See ESI† for more details on the simulation procedure.) The dimerization process was monitored *via* the center-of-mass distance between the two monomers as well as the number of inter-monomer heavy atom contacts (Fig. 1). Three independent dimerization simulations were performed with different random initial relative orientations and velocities. All the three trajectories exhibited similar A β 42 dimerization processes, except that the time scale differs for the dimerization events. In the following, we present the results based on the single representative trajectory.

Two distinctive stages were observed in the A β 42 dimerization pathway. In the first stage, the two monomers undergo transient formation of a contact structure followed by detachment, and this attachment–detachment process is repeated until a critical number of inter-monomer heavy atom contacts are developed. The relative orientation of colliding A β 42 monomers seems to play a crucial role in determining the lifetime of transient contacts as can be inferred from representative contact structures displayed in Fig. 1. We refer to this stage as the nucleation stage since it resembles the nucleation step in the protein folding where the formation of necessary contacts between nucleation sites is a critical event.¹⁶ Once the two approaching A β 42 monomers develop sufficient heavy atom contacts around potential nucleation sites, they do not fall apart and the dimerization proceeds to the next stage.

Department of Chemistry, Sookmyung Women's University, Hyochangwon-gil 52, Yongsan-gu, Seoul 140-742, Korea.
E-mail: sihyun@sookmyung.ac.kr; Fax: +82 2 2077 7321;
Tel: +82 2 2710 9410

† Electronic supplementary information (ESI) available: Details on the simulation method and on the solvation thermodynamic analysis using the integral-equation theory of liquids. See DOI: 10.1039/c2cp23326f

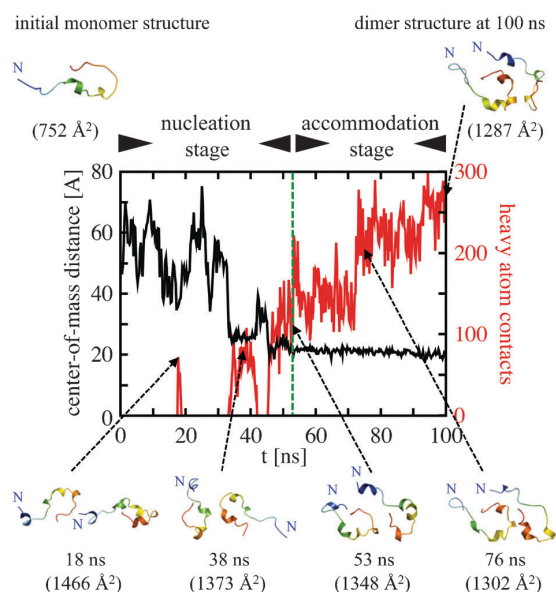


Fig. 1 Center-of-mass distance between two monomers (black curve, left axis) and the number of inter-monomer heavy atom contacts (red curve, right axis) as a function of time. Initial Aβ42 monomer structure taken from ref. 15 and representative Aβ42 contact structures along the dimerization simulation trajectory are also displayed. Each protein structure is color-coded according to the sequence, ranging from blue to red at N- and C-termini, respectively. Numbers in parentheses refer to the collision cross-section. The simulation trajectory can be divided into the nucleation and structural accommodation stages (separated by the vertical dashed line). In the nucleation stage, transient formations of contact structures are observed between 17.6 and 18.7 ns and between 33.0 to 41.9 ns. In the subsequent structural accommodation stage, a significant number of heavy atom contacts are developed after 53 ns.

In the subsequent stage, the number of inter-monomer heavy atom contacts further increases, and the two interacting monomers undertake conformational adjustment which results in a compact dimer structure. This second stage is the structural accommodation stage since the most distinctive feature here is the conformational adjustment of each of the two monomers to form favorable interactions between them. To validate the structural features of our Aβ42 dimerization simulation, we calculated the collision cross-sections along the dimerization trajectory (Fig. 1), whose experimental values are available from ion-mobility mass spectrometry.⁴ The computed collision cross-sections of both the initial Aβ42 monomer structure (752 Å²) and the final Aβ42 dimer structure (1287 Å²) are in good agreement with the experimental values⁴ (702 and 1256 Å², respectively). This assesses the relevance of our simulation results to experiment.

To elucidate the driving force in each stage of the dimerization process, we performed the calculation of the protein internal energy E_u (sum of intra- and inter-monomer energies) and the solvation free energy $\Delta\mu$ for Aβ42 dimer conformations taken from the simulation (Fig. 2). (See ESI† on the solvation free energy calculation based on the integral-equation theory of liquids.) The knowledge on the protein solvation free energy in water is an essential ingredient to understand the water-induced force between proteins such as the hydrophobic interaction.¹⁷

When E_u and $\Delta\mu$ are combined, one obtains the solvent-averaged protein potential energy $f = E_u + \Delta\mu$.¹⁸ Averaging f over dimer conformations yields the Gibbs free energy up to the protein configurational entropy.¹⁸ (The contribution from the pressure–volume term was found to be negligible, in which case Gibbs and Helmholtz free energies can be used interchangeably.) Since the protein configurational entropy would decrease upon dimerization, the driving force for the dimerization must originate from f , i.e., either from E_u or $\Delta\mu$.

Fig. 2 demonstrates that the force that drives the initial nucleation stage is the *decrease* in the solvation free energy of the two unfolded monomers upon their approach, which was acquired through their conformational transitions in water prior to dimerization.¹⁹ Such a water-induced force overcomes the large unfavorable increase in the protein internal energy originating from the electrostatic repulsion between the negatively charged Aβ42 monomers, and drives the two monomers to approach each other. On the other hand, the force that drives the second structural accommodation stage of the Aβ42 dimerization is the *decrease* in the protein internal energy (Fig. 2). In this stage, favorable protein–protein interactions, mainly due to van der Waals contacts, lead to the formation of compact dimer structures. Consequently, the collision cross-section is further decreased during the structural accommodation stage (Fig. 1). Interestingly, the solvation free energy of the Aβ42 dimer is increased during this structural accommodation stage (Fig. 2). The increased solvation free energy of the Aβ42 dimer may act as the driving force to recruit other Aβ42 proteins in the subsequent nucleation stage of further oligomerization processes.

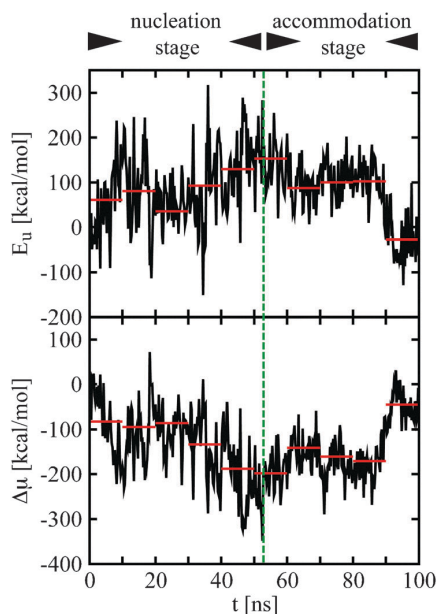


Fig. 2 Protein internal energy E_u (upper panel) and solvation free energy $\Delta\mu$ (lower panel) as a function of time. The initial values are set to zero. Horizontal red bars represent averages over each 10 ns time interval. Vertical dashed line refers to 53 ns separating the nucleation and structural accommodation stages.

Structural and Thermodynamic Characteristics That Seed Aggregation of Amyloid- β Protein in Water

Song-Ho Chong, Mirae Park, and Sihyun Ham*

Department of Chemistry, Sookmyung Women's University, Hyochangwon-gil 52, Yongsan-gu, Seoul, 140-742, Korea

ABSTRACT: Amyloid- β ($A\beta$) proteins undergo conformational transitions leading to aggregation-prone structures, which can initiate self-assembly to form soluble oligomers and eventually insoluble amyloid fibrils when transferred from the transmembrane phase to the physiological aqueous phase. Yet, how $A\beta$ proteins acquire an aggregation-prone nature during the conformational transitions in water remains elusive. Here, we investigate key structural and thermodynamic features of a 42-residue $A\beta$ ($A\beta_{42}$) protein that seed aggregation based on the fully atomistic, explicit-water molecular dynamics simulations as well as on the integral-equation theory of liquids for solvation thermodynamic analysis. We performed a structure-based analysis on the solvation free energy, a major determinant of the protein hydrophobicity/solubility that influences the aggregation propensity of $A\beta_{42}$ protein in water. In addition, the Gibbs free energy and its constituents including protein internal energy, protein configurational entropy, solvation enthalpy, and solvation entropy were computed to elucidate thermodynamic driving forces for the conformational transitions of $A\beta_{42}$ protein in water. On the basis of the atomic-decomposition analysis of these thermodynamic functions, we demonstrate how N-terminal (residues 1–11) and C-terminal (39–42) regions as well as the central region (16–18) contribute significantly to decreasing the solubility of $A\beta_{42}$ protein upon its conformational transitions in water. These results are consistent with the recent experimental and computational implications and further provide the molecular origin for why the C terminus may serve as an “internal seed” for aggregation and the N-terminal segment may act as a “catalyst” in inducing the $A\beta_{42}$ self-assembly. This work takes a step forward toward the identification of structural and thermodynamic features of the $A\beta_{42}$ monomer that seed the aggregation process in water.

■ INTRODUCTION

A number of human diseases, such as Alzheimer's disease, Parkinson's disease, and type II diabetes, are associated with the deposition of insoluble protein aggregates.¹ Protein aggregation occurs in multistep processes which involve initial conformational changes to form aggregation-prone structures capable of initiating self-assembly, sometimes called protein misfolding.² Misfolded proteins can self-assemble into a variety of neurotoxic aggregate species, ranging from soluble oligomers to amyloid fibrils.³ Understanding the driving factors and mechanistic information of the initial misfolding and subsequent aggregation has been one of the most actively investigated issues in relation to the development of therapeutics and medical applications for the cure of protein aggregation diseases.^{4,5} Recently, an increasing number of evidence has accumulated, indicating that soluble oligomers rather than mature amyloid fibrils are the primary toxic agent,^{6,7} leading to more focus on early stages of oligomerization involving monomers and small oligomers.⁸

Alzheimer's disease is linked to the aggregation of amyloid- β ($A\beta$) proteins.^{6,9} Deducing the aggregation mechanism of $A\beta$ proteins requires studies on the structures of monomeric and oligomeric species as well as on the kinetics and thermodynamics of their interconversions. However, experimental investigation on the atomistic structures of the monomer and each oligomer species has been hindered by their strong propensity to aggregate in water.¹⁰ Molecular dynamics (MD) simulations have therefore been extensively used to characterize atomic-level structural features of $A\beta$ monomers, the starting point for aggregation. To enable efficient sampling of $A\beta$

conformations, numerous groups have studied $A\beta$ fragments in implicit or explicit solvent and full-length $A\beta$ proteins in implicit solvent.^{11–39} To fully take into account the detailed protein–water interactions, explicit-solvent MD simulations on full-length $A\beta$ proteins have also been performed.^{40–50}

Despite these numerous computational efforts, key structural features of $A\beta$ monomers that seed aggregation in water remain elusive. In this regard, the knowledge of the protein solvation free energy is an essential ingredient since the change in the hydrophobicity/solubility of a protein is closely correlated with the propensity to form protein aggregates^{51,52} and the solvation free energy is a major determinant of the hydrophobicity/solubility.^{53,54} The solvation free energy of a protein in water depends on the protein conformation, chemical composition of the protein surface, and the protein–solvent interaction. For example, it has been observed that, when transferred from the organic phase to the aqueous phase, $A\beta$ protein acquires a strong propensity to aggregate in water.⁴¹ This occurs since the protein conformational transitions take place in response to the change in the surrounding physicochemical environment, and the protein surface acquires a more hydrophobic nature, i.e., becomes more prone to aggregate, during those conformational transitions. Therefore, in elucidating the key structural features that seed aggregation, it is crucial to follow how the change in the hydrophobicity/solubility of $A\beta$ protein occurs during the course of the conformational transitions in water. On the other hand, it is also important to note that it is not the solvation free

Received: October 26, 2011

Published: January 7, 2012



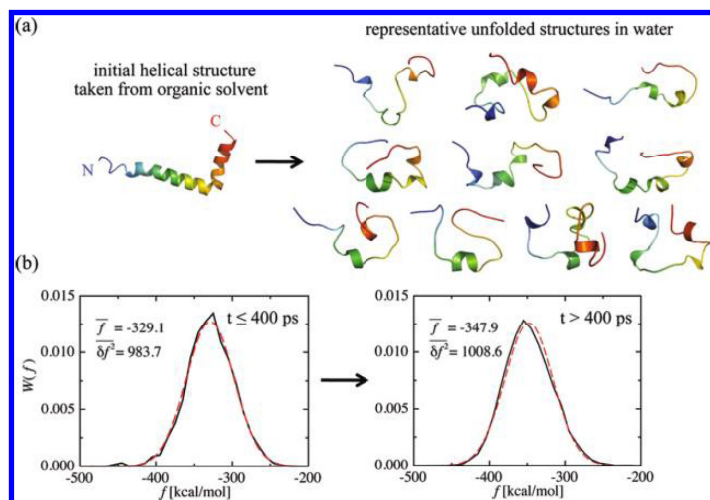


Figure 1. (a) NMR structure of the A β 42 monomer determined in an 80:20 hexafluoroisopropanol/water mixture at neutral pH (PDB ID: 1IYT)⁵⁷ (left) and representative aqueous-phase A β 42 monomer structures determined in ref 46 from the clustering analysis of each of 10 independent simulation trajectories (right). Each structure is color-coded according to sequence, ranging from blue to red at the N and C termini, respectively. (b) Probability distribution function $W(f)$ (solid curves) of the solvent-averaged protein potential energy f of A β 42 in water for the short-time regime ($t \leq 400$ ps) related to the initial helical configurations representing organic phase structure (left panel) and the one for the long-time regime ($t > 400$ ps) for the aqueous-phase configurations (right panel). Dashed curves denote the fit by the Gaussian function. Numerical values of $\bar{f}$ and δf^2 in units of kcal/mol and (kcal/mol)² are reported in each panel.

energy, but the Gibbs free energy, that determines whether such conformational transitions occur spontaneously. Therefore, a detailed thermodynamic analysis in terms of the Gibbs free energy and its constituents—protein internal energy, protein configurational entropy, and solvation free energy along with solvation enthalpy and solvation entropy components—is necessary to probe driving forces for protein misfolding and aggregation.

In this paper, we report a detailed thermodynamic investigation on the full-length, 42-residue form of A β protein (A β 42) in water based on explicit-water MD simulations as well as on the solvation thermodynamic analysis using the integral-equation theory of liquids.^{55,56} The aim of this work is to address key structural and thermodynamic features of the A β 42 monomer that seed the subsequent oligomerization process in water. We used the helical structure determined by NMR in a membrane-mimicking organic solvent⁵⁷ as the initial structure and carried out fully atomistic MD simulations of the A β 42 monomer with explicit water. This closely resembles the *in vivo* situation that the A β 42 protein is produced through the endoproteolysis of the transmembrane amyloid precursor protein and then is released to an extracellular aqueous environment where the A β 42 aggregation takes place.³ The relevance of the aqueous-phase structures of the A β 42 monomer generated from our MD simulations is validated through the comparison of the collision cross-section measured by ion-mobility mass spectrometry.⁵⁸ Protein internal energy was directly computed during the simulations, while the integral-equation theory for molecular systems^{55,56} was used to yield solvation thermodynamic quantities. These results were combined to obtain protein configurational entropy,⁵⁹ allowing one to determine the Gibbs free energy change and all of its constituents upon the A β 42 conformational transitions in water. The atomic decomposition analysis of the solvation free energy⁶⁰ was further conducted to elucidate how A β 42 protein

acquires an aggregation-prone nature during the conformational transitions and to reveal the regions of the A β 42 protein that significantly affect its aggregation propensity in water.

RESULTS

Aqueous-Phase A β 42 Monomer Structures from Simulations Are in Accord with Experimental Data. We performed MD simulations to study the conformational transitions of A β 42 monomer in water starting from the helical structure determined in a membrane-mimicking organic solvent [see Figure 1a].⁴⁶ In total, 10 independent simulations of 50 ns were performed at a temperature $T = 300$ K and a pressure $P = 1$ bar under neutral pH (see Methods). Representative aqueous-phase A β 42 monomer structures are displayed in Figure 1a, which were determined from the clustering analysis of each of the 10 independent simulated trajectories. The initial event of conformational transitions for A β 42 simulation in water is found to be the unfolding of the C-terminal region,⁴⁶ which is in accord with the experimental observation that the C-terminal unfolding was detected when A β 42 was exposed to more water from the organic solvent.⁴¹ Our simulated structure also agrees with the NMR dynamic study in that the C terminus of A β 42 monomer in water is rather structured and buried.⁶¹ The J-coupling constants from our simulations are also found to be in agreement with the experimental NMR data.^{42,46}

For further corroboration, we compared the collision cross-section of A β 42 monomer structures from our simulations with that determined from the ion-mobility mass spectrometry.^{15,58} In order to compare with the solvent-free experiments on samples electrosprayed from the solution phase, each protein structure from our simulations was instantaneously dehydrated through energy minimization in a vacuum. The collision cross-section of the dehydrated structure was then calculated using the trajectory method implemented in the MOBICAL software.⁶² We obtained 738 ± 25 Å² from the 10



Metal ion shuttling mechanism through thiacalix[4]crown: a computational study

Jooyeon Hong, Sunhee Cho, Sihyun Ham^{*}

Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-Ku, Seoul 140-742, Republic of Korea

ARTICLE INFO

Article history:

Received 18 January 2012

Revised 6 February 2012

Accepted 7 February 2012

Available online 15 February 2012

Keywords:

Thiacalix[4]crown

Density functional theory

Shuttling

Crown ether

Molecular recognition

ABSTRACT

The first computational study on the alkali metal ion shuttling mechanism through thiacalix[4]biscrown-5 has been performed using density functional theory (MPWB1K/6-31G(d)//B3LYP/6-31G(d)). The activation free energy for K⁺ ion shuttling through the calix tube of thiacalix[4]biscrown-5 is computed to be 12.7 kcal/mol which is in good agreement with the experimental value. Throughout the shuttling process, K⁺ ion maintains the maximum electrostatic interactions with ether oxygens of crown rings, which contributes to lower the activation barrier of the shuttling process.

© 2012 Elsevier Ltd. All rights reserved.

For decades, a lot of attention has been accumulated on calix[4]-arene to its potential accessibility for molecular recognition, ion transport phenomena, enzyme model building and the construction of sensors for a wide variety of applications.^{1,2} Of particular interest in this regard is thiacalix[4]arene (**1**) since the bridging sulfur atoms in the macrocyclic cavity can provide greater dimensionality on the cavity size, functional modification and extra binding sites for guest complexation compared to calix[4]arene.^{3–10} Among the thiacalix[4]arene derivatives, thiacalix[4]biscrown with crown-ether linkages was synthesized and its metal ion extractability is found to be lower than that of conventional calix[4]biscrown.¹¹ Consequently, one of the most interesting features of thiacalix[4]biscrown-5 (**2**) with crown-5 rings is the metal ion shuttling through the aromatic cavity due to its relatively weak metal binding affinity^{12,13} (Scheme 1). Temperature-dependent (TD) ¹H NMR spectroscopic study revealed coalescence temperatures T_c for potassium ion exchange in **2**.¹² Recently, the structural and energetic preferences of **2** with and without alkali metal ions (Na⁺, K⁺, Rb⁺ and Cs⁺) have been theoretically investigated by using molecular dynamic simulations and density functional theory (DFT) methods.¹⁴ However, the detailed mechanism of the metal ion shuttling process through the calix tube of **2** has not been determined. This phenomenon is particularly important because of its biological implication to understand the metal ion transport mechanism through the ion channel protein in cell membrane. Conceivably, the detailed theoretical study on the metal ion shuttling process

in thiacalix[4]biscrown tube may provide valuable molecular insights for understanding the ion transport mechanism in membrane proteins.

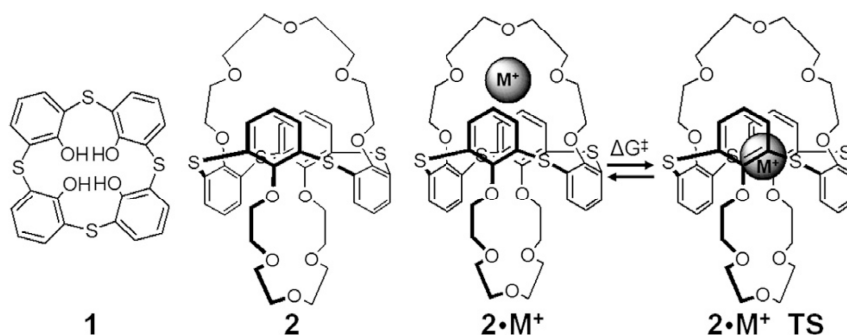
Here, we report the first theoretical investigation on the metal ion shuttling mechanism through the calix tube of **2** for various alkali metal ions (Na⁺, K⁺, Rb⁺ and Cs⁺) by using DFT method at the MPWB1K/6-31G(d)//B3LYP/6-31G(d) level. We elucidate structural and thermodynamic features for the metal ion shuttling process in **2** via a transition state (TS) characterization. The computed activation free energy for K⁺ shuttling is compared with the experimental one.

The global minimum structure of **1** is known to be the cone conformation by experimental³ as well as computational⁹ studies. The cone of **1** is the most stable due to the cyclic array of the intramolecular hydrogen bonds between hydroxyl groups. The conformational preference of **2** was reported to be 1,3-alternate by both experimental¹¹ and computational¹⁴ studies. The computational study predicts that the ether oxygens in the crown-5 ring rather than sulfur atoms in the thiacalixarene ring would be involved in cation recognition based on the atomic charge analysis.¹⁴

For each alkali metal ion complex of **2** (2·M⁺), two energy minima complexes were located with 2·M⁺_A, the lowest energy complex, and 2·M⁺_B, the second lowest at the B3LYP/6-31G(d) level.¹⁵ In Figure 1, the potassium ion is depicted for M⁺ since it shows the highest binding affinity toward **2**. For all 2·M⁺_A complexes, five ether oxygens in the crown-5 ring on one side of **2** are involved in the electrostatic interactions to recognize the metal ion. On the other hand, four ether oxygens exhibit attractive interactions to M⁺ in the 2·M⁺_B complexes, except the case of 2·Na⁺_B where

^{*} Corresponding author. Tel.: +82 2 710 9410; fax: +82 2 2077 7321.

E-mail addresses: sihyun@sookmyung.ac.kr, sihyun@sm.ac.kr (S. Ham).



Scheme 1. Schematic representation of thiactalix[4]arene **1**, thiactalix[4]biscrown-5 **2**, metal ion complex of **2** (**2·M⁺**), and transition state for metal ion shuttling (**2·M⁺·TS**).

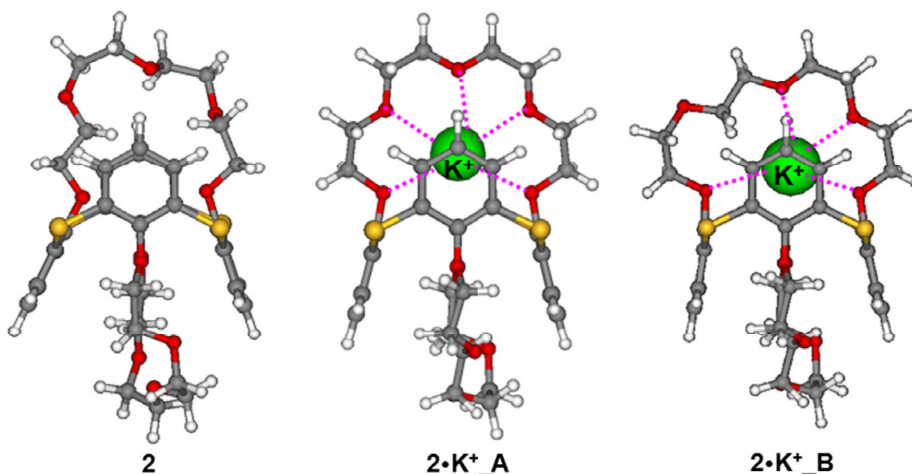


Figure 1. Optimized geometries for **2**, **2·K⁺_A**, and **2·K⁺_B** computed at the B3LYP/6-31G(d) level. The pink dotted lines represent the electrostatic interactions between K⁺ and ether oxygens. Oxygen atoms are shown in red and sulfur atoms are in yellow.

only three ether oxygens display attraction force to Na⁺ ion. The average bond distances between metal ion and ether oxygens are listed in Table 1. Overall, **2·M⁺_A** exhibits greater complexation energy than **2·M⁺_B** as shown in Table 1. The most distinctive structural change from a free state **2** to cation-binding complex (**2·M⁺**) is the reorientation of the phenole-ether ring to maximize the electrostatic interaction with a cation by ether oxygens. The cation recognition of **2** is mainly attributed by the electrostatic interaction between ether oxygens and a metal ion together with the minor contribution of the cation- π interaction between two facing phenyl rings around the metal ion.

Since **2** exhibits the highest selectivity for K⁺ ion among alkali metals by experiments¹² as well as by the computations herein, the geometrical features of **2·K⁺** complexes are discussed in detail. For the **2·K⁺_A**, the average distance between five crown ether oxygens and K⁺ ion is 2.88 Å, whereas that between four ether oxygens and K⁺ ion is 2.81 Å for the **2·K⁺_B**. The distances between two phenolic oxygens nearby K⁺ ion in the upward cavity are 5.08 and 5.31 Å for **2·K⁺_A** and **2·K⁺_B**, respectively. The dipole moment for the **2·K⁺_A** is computed to be 21.9 D which is larger than 6.1 D for **2·K⁺_B** at the MPWB1K/6-31G(d)//B3LYP/6-31G(d) level (Supplementary data). Despite the large dipole moment, **2·K⁺_A** is located to be more stable than **2·K⁺_B** due to the maximized electrostatic interactions between ether oxygens and K⁺ ion in the crown-5 ring. Although the binding affinity between **2** and K⁺ ion is stronger than that between **2** and other alkali metal cations, it

Table 1

The complexation energies^a (kcal/mol) and the average lengths (Å) between M⁺ ion and ether oxygens for **2·M⁺_A** and **2·M⁺_B** complexes are listed at the B3LYP/6-31G(d) and MPWB1K/6-31G(d)//B3LYP/6-31G(d) level of theory

M ⁺		Na ⁺	K ⁺	Rb ⁺	Cs ⁺
2·M⁺_A	B3LYP	—	−40.5	−29.6	−12.9
	MPWB1K	—	−58.4	−48.0	−32.0
	av. $r(\text{O} \cdots \text{M}^+)$	2.73	2.88	2.95	3.05
2·M⁺_B	B3LYP	—	−24.0	−11.2	+6.6
	MPWB1K	—	−42.6	−31.4	−14.0
	av. $r(\text{O} \cdots \text{M}^+)$	2.40	2.81	2.98	3.09

^a Complexation energy = $E_{\text{complex}} - (E_{\text{host}} + E_{\text{alkali metal ion}})$.

is weaker than −72.0 kcal/mol for 18-crown-6 complex with K⁺ ion (MP2/6-31+G(d) level).¹⁶ Relatively weak complexation energy for K⁺ ion by the host **2** may contribute to K⁺ ion shuttling process in the calix tube of **2** as observed experimentally.

To elucidate the metal ion shuttling mechanism through **2·M⁺** complex, the transition state complexes (**2·M⁺·TS**) for shuttling process through the calix tube of **2** with four alkali metal ions were located. The activation barriers for the shuttling process are summarized and compared with experimental data as shown in Table 2. According to TD ¹H NMR experiment, the activation free energy (ΔG[‡]) of K⁺ ion shuttling through the calix cavity of **2** is 14.7 kcal/mol.¹² The ΔG[‡] for K⁺ ion shuttling from **2·K⁺_B** is computed to be



Component analysis of the protein hydration entropy

Song-Ho Chong, Sihyun Ham*

Department of Chemistry, Sookmyung Women's University, Hyochangwon-gil 52, Yongsan-gu, Seoul 140-742, Republic of Korea

ARTICLE INFO

Article history:

Received 24 December 2011

In final form 10 March 2012

Available online 10 April 2012

ABSTRACT

We report the development of an atomic decomposition method of the protein solvation entropy in water, which allows us to understand global change in the solvation entropy in terms of local changes in protein conformation as well as in hydration structure. This method can be implemented via a combined approach based on molecular dynamics simulation and integral-equation theory of liquids. An illustrative application is made to 42-residue amyloid-beta protein in water. We demonstrate how this method enables one to elucidate the molecular origin for the hydration entropy change upon conformational transitions of protein.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Water plays a crucial role in many aspects of life phenomena [1]. Solvation forces are among the major determinants of structure, stability, dynamics, and function of biomolecules such as proteins [2–4]. The knowledge on solvation thermodynamic properties is therefore an essential ingredient to understand driving forces for biological processes in water such as protein folding, protein–protein association, and protein–ligand binding. Recently, a critical role of the solvation entropy has been pointed out in the determination of compact protein structures where backbones and side chains are tightly packed [5,6]. This stems from the notion that the formation of intra-protein hydrogen bonds and van der Waals contacts, which leads to a favorable decrease in the protein internal energy, necessarily accompanies dehydration of surrounding water molecules, and the gain in the protein internal energy is largely canceled out by an unfavorable increase in the solvation enthalpy that reflects the dehydration penalty. Interestingly, the dehydration penalty can be relatively larger [7,8]. It has also been demonstrated that the hydration entropy quantifies the tightness of interface packing between subunits constituting F_1 -ATPase, a rotary-motor protein complex [9].

However, the thermodynamic characterization by a single number, e.g., a value of the solvation entropy assigned to each protein conformation, does not provide detailed microscopic insights. To understand a change in macroscopic thermodynamic data in terms of microscopic interactions, the decomposition of thermodynamic properties into contributions from groups of atoms (e.g., functional

groups, residues) and individual components of potential energy terms (e.g., van der Waals, electrostatic terms) is indispensable. Such a decomposition, often referred to as component analysis, has been widely employed in an attempt to unravel the role of particular functional groups or interaction terms in the protein stability, protein–protein association, and protein–ligand binding [10–16]. However, the development of the decomposition method and its application so far have been limited mainly to the free energy or to the solvation free energy. (See Refs. [17–19] on the decomposition of the solvation entropy based on a different perspective, namely, into translational and rotational components and their further partitioning into different orders of correlation contributions and into terms which are dependent on the volume, area, and integrated curvatures of the space enclosed by the solvent-accessible surface.)

Herein, we report an exact decomposition method of the protein hydration entropy into atomic and individual potential energy contributions. Our method can be implemented in a straightforward manner via a combined approach based on molecular dynamics (MD) simulation and integral-equation theory of liquids. An illustrative application is made to 42-residue amyloid-beta protein ($A\beta$ 42) in water, whose aggregation is associated with Alzheimer's disease [20,21]. In organic solvent, $A\beta$ 42 protein has α -helical structure and does not exhibit an aggregation [22]. When transferred to aqueous phase, on the other hand, $A\beta$ 42 protein attains a great propensity to aggregate [23]. Investigation on the conformational transitions of $A\beta$ 42 protein that occur in water prior to aggregation, called the protein misfolding [24], has been an active field of research to elucidate the mechanism of the aggregation and to develop therapeutics for Alzheimer's disease. We demonstrate how the decomposition method enables one to elucidate the molecular origin for the solvation entropy change upon $A\beta$ 42 misfolding in water.

* Corresponding author. Fax: +82 2 2077 7321.

E-mail address: sihyun@sookmyung.ac.kr (S. Ham).

2. Method

2.1. Atomic decomposition method of the solvation entropy

Our decomposition method of the solvation entropy is based on the Kirkwood charging formula [25] for the solvation free energy. We denote by $u_\gamma(\mathbf{r})$ the interaction potential acting on the water site γ , oxygen or hydrogen, of position $\mathbf{r}$ which is generated by atoms in protein. The equilibrium distribution function of the water site γ around the protein shall be denoted by $g_\gamma(\mathbf{r})$. The protein solvation free energy in water is then given by [25]

$$\Delta\mu = \rho \sum_\gamma \int_0^1 d\lambda \int d\mathbf{r} \frac{\partial u_\gamma(\mathbf{r}; \lambda)}{\partial \lambda} g_\gamma(\mathbf{r}; \lambda) \quad (1)$$

Here ρ denotes the average number density of water, and the coupling parameter λ gradually switches on the protein–water interaction potential from no interaction at $\lambda = 0$, $u_\gamma(\mathbf{r}; \lambda = 0) = 0$, to the full interaction at $\lambda = 1$, $u_\gamma(\mathbf{r}; \lambda = 1) = u_\gamma(\mathbf{r})$. $g_\gamma(\mathbf{r}; \lambda)$ denotes the equilibrium water distribution function when the interaction potential is $u_\gamma(\mathbf{r}; \lambda)$. Various integration paths for λ may be considered (see below on our choice of the path), but the integrated result for $\Delta\mu$ is identical.

While the solvation free energy does not depend on whether the solvation is considered under the isochoric (constant volume) or isobaric (constant pressure) condition, the solvation entropy depends on such a condition [26]. Denoting the solvation entropies under the isochoric and isobaric conditions as Δs_V and Δs_P , respectively, these are related by [26]

$$T\Delta s_P = T\Delta s_V + \frac{T\alpha_P}{\kappa_T} V_u \quad (2)$$

in terms of the isobaric thermal expansion coefficient α_P and the isothermal compressibility κ_T of water, and the partial molar volume V_u of the protein. In (2), the temperature T is multiplied so that the solvation entropy has units of energy. As we will see below, the second term on the right-hand side of (2) is practically negligible as far as the change in the solvation entropy upon protein conformational transition is concerned. We shall therefore develop a decomposition method for the isochoric solvation entropy. From here on, the isochoric solvation entropy shall be simply referred to as the solvation entropy, and is denoted as Δs without the subscript V for brevity.

Since Δs is related to the temperature derivative of $\Delta\mu$ at constant density, $\Delta s = -(\partial\Delta\mu/\partial T)_\rho$, one obtains from (1)

$$\Delta s = -\rho \sum_\gamma \int_0^1 d\lambda \int d\mathbf{r} \frac{\partial u_\gamma(\mathbf{r}; \lambda)}{\partial \lambda} \left(\frac{\partial g_\gamma(\mathbf{r}; \lambda)}{\partial T} \right)_\rho \quad (3)$$

We notice here that the most commonly used form of the protein–water interaction potential $u_\gamma(\mathbf{r})$ is represented by a sum of radially symmetric Lennard–Jones (LJ) and Coulomb electrostatic terms which are centered on the position $\mathbf{r}_\alpha$ of atom α in protein

$$u_\gamma(\mathbf{r}) = \sum_{\alpha=1}^N \left[u_{x\gamma}^{(\text{LJ})}(|\mathbf{r} - \mathbf{r}_\alpha|) + u_{x\gamma}^{(\text{elec})}(|\mathbf{r} - \mathbf{r}_\alpha|) \right] \quad (4)$$

Here N is the number of atoms in the protein, and $u_{x\gamma}^{(\text{LJ})}(r) = 4\epsilon_{x\gamma}[(\sigma_{x\gamma}/r)^{12} - (\sigma_{x\gamma}/r)^6]$ and $u_{x\gamma}^{(\text{elec})}(r) = q_\alpha q_\gamma / r$ with $\epsilon_{x\gamma}$, $\sigma_{x\gamma}$, q_α , and q_γ being the LJ parameters and atomic charges, respectively. When both the LJ and electrostatic potentials are present, it is necessary to treat them separately by introducing two coupling parameters λ_1 and λ_2 in the Kirkwood charging formula. The parameters λ_1 and λ_2 shall be chosen to scale the LJ parameter ($\lambda_1 \sigma_{x\gamma}$) and the atomic charge in protein ($\lambda_2 q_\alpha$), respectively. We consider the integration path in which the LJ interaction is turned on first involving the integration over λ_1 from 0 to 1 with $\lambda_2 = 0$, and then the electrostatic interaction is switched on by performing the integration over λ_2

from 0 to 1 with $\lambda_1 = 1$. As argued in Ref. [15], this is the most natural and practical choice when dealing with the solvation of macromolecule such as protein. Thereby, we obtain from (3) and (4) the following exact partitioning of the solvation entropy Δs into atomic contribution (Δs_α), which can further be decomposed into the non-electrostatic ($\Delta s_\alpha^{(\text{LJ})}$) and electrostatic ($\Delta s_\alpha^{(\text{elec})}$) terms:

$$\Delta s = \sum_{\alpha=1}^N \Delta s_\alpha \quad \text{with} \quad \Delta s_\alpha = \Delta s_\alpha^{(\text{LJ})} + \Delta s_\alpha^{(\text{elec})} \quad (5)$$

in which

$$\Delta s_\alpha^{(\text{LJ})} = -4\pi\rho \sum_\gamma \int_0^1 d\lambda_1 \int r^2 dr \frac{\partial u_{x\gamma}^{(\text{LJ})}(r; \lambda_1)}{\partial \lambda_1} \left(\frac{\partial g_{x\gamma}(r; \lambda_1, \lambda_2 = 0)}{\partial T} \right)_\rho \quad (6)$$

and

$$\Delta s_\alpha^{(\text{elec})} = -4\pi\rho \sum_\gamma \int_0^1 d\lambda_2 \int r^2 dr \frac{\partial u_{x\gamma}^{(\text{elec})}(r; \lambda_2)}{\partial \lambda_2} \left(\frac{\partial g_{x\gamma}(r; \lambda_1 = 1, \lambda_2)}{\partial T} \right)_\rho \quad (7)$$

Here $g_{x\gamma}(r)$ (the dependence on λ_1 and λ_2 is dropped for brevity) denotes the radial distribution function between atom α in protein and atom γ in water, which is obtained from $g_\gamma(\mathbf{r})$ by shifting the origin of the coordinate to the position $\mathbf{r}_\alpha$ of atom α followed by the orientational average, i.e.,

$$g_{x\gamma}(r) = \frac{1}{4\pi} \int d\hat{\mathbf{r}} g_\gamma(\mathbf{r}_\alpha + \mathbf{r}) \quad (8)$$

with $\hat{\mathbf{r}} = \mathbf{r}/r$ and $r = |\mathbf{r}|$.

2.2. Computational procedure

We implement our decomposition method by combining the MD simulation and the three-dimensional reference-interaction-site-model (3D-RISM) integral-equation theory [27,28], and apply it to investigate the solvation entropy change upon A β 42 misfolding in water. We previously performed explicit-water MD simulations for A β 42 protein starting from the NMR determined helical structure in organic solvent (PDB ID: 1IYT [22], Figure 1), and investigated the misfolding process of A β 42 protein in water [29]. For each protein conformation generated by the simulations, one can determine the interaction potential $u_\gamma(\mathbf{r}; \lambda_1, \lambda_2)$ around the protein along the integration path for λ_1 and λ_2 . Based on the knowledge of $u_\gamma(\mathbf{r}; \lambda_1, \lambda_2)$, one can compute the equilibrium water distribution function $g_\gamma(\mathbf{r}; \lambda_1, \lambda_2)$ around the protein using the 3D-RISM theory (see below). The solvation entropy and its decomposition can then be obtained from (5)–(7) after the numerical integration over λ_1 and λ_2 .

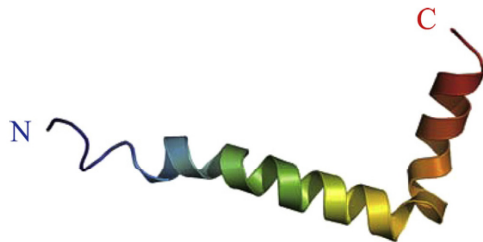


Figure 1. NMR structure of A β 42 monomer determined in hexafluoroisopropanol/water 80:20 mixture at neutral pH (PDB ID: 1IYT [22]). The structure is color-coded according to sequence, ranging from blue to red at the N- and C-termini, respectively. (For interpretation of references to color in this figure legend, the reader is referred to the web version of this article.)

Impact of chemical heterogeneity on protein self-assembly in water

Song-Ho Chong and Sihyun Ham¹

Department of Chemistry, Sookmyung Women's University, Hyochangwon-gil 52, Yongsan-gu, Seoul 140-742, Korea

Edited by* Peter J. Rossky, The University of Texas at Austin, Austin, TX, and approved March 15, 2012 (received for review December 20, 2011)

Hydrophobicity is thought to underlie self-assembly in biological systems. However, the protein surface comprises hydrophobic and hydrophilic patches, and understanding the impact of such a chemical heterogeneity on protein self-assembly in water is of fundamental interest. Here, we report structural and thermodynamic investigations on the dimer formation of full-length amyloid- β proteins in water associated with Alzheimer's disease. Spontaneous dimerization process—from the individual diffusive regime at large separations, through the approach stage in which two proteins come close to each other, to the structural adjustment stage toward compact dimer formation—was captured in full atomic detail via unguided, explicit-water molecular dynamics simulations. The integral-equation theory of liquids was then applied to simulated protein structures to analyze hydration thermodynamic properties and the water-mediated interaction between proteins. We demonstrate that hydrophilic residues play a key role in initiating the dimerization process. A long-range hydration force of enthalpic origin acting on the hydrophilic residues provides the major thermodynamic force that drives two proteins to approach from a large separation to a contact distance. After two proteins make atomic contacts, the nature of the water-mediated interaction switches from a long-range enthalpic attraction to a short-range entropic one. The latter acts both on the hydrophobic and hydrophilic residues. Along with the direct protein–protein interactions that lead to the formation of intermonomer hydrogen bonds and van der Waals contacts, the water-mediated attraction of entropic origin brings about structural adjustment of constituent monomer proteins toward the formation of a compact dimer structure.

solvation thermodynamics | dehydration | protein interface

Water-mediated interaction such as hydrophobic interaction plays a key role in protein folding and protein self-assembly (1–3), yet elucidating its physical and chemical basis still remains a challenge. Recent theoretical advances in understanding the hydrophobic effect have elucidated that the hydrophobic interaction between small apolar groups is markedly different from that between large hydrophobic solutes, with the cross-over length-scale being of the order of 10 Å (see, e.g., refs. 4 and 5 for review). However, most of previous studies employed solutes such as cylinders, plates, or hydrophobic chains, and a direct applicability of findings therefrom to proteins is not obvious. A major obstacle here is the fact that protein surfaces are both topologically and chemically heterogeneous (6, 7): Protein surfaces are irregular in shape and hydrophobic residues are laced with hydrophilic ones. The main focus of the present work is to contribute to the understanding of the impact of chemical heterogeneity on protein self-assembly in water via the study of solvation for a realistic system of protein and water.

Self-assembly of large hydrophobic solutes is thought to be driven by a drying transition of interfacial water (4, 5). It was demonstrated by simulation that, when two large hydrophobic plates are brought together to a certain critical distance, interfacial water exhibits a drying transition (8). Such a dewetting transition gives rise to a strong driving force for hydrophobic collapse as confirmed by theory and subsequent simulations (9–11).

Possible relevance of the dewetting-induced collapse to protein self-assembly has also been examined (12–14). (See ref. 15 for a critical view on the potential role of the dewetting-induced self-assembly, refs. 16 and 17 on the “lubrication” scenario in the absence of the dewetting transition, and ref. 18 on the possible role of water confined in hydrophobic pores.) The existence of the dewetting transition was found to be sensitive to the details of protein–water interactions (12–14). In particular, the surface chemical heterogeneity is expected to have a profound effect. Indeed, recent works on model hydrophilic plates have addressed a related issue and shown that the presence of hydrophilic group provokes wetting against dewetting behavior (19–21). This line of study was extended to a protein with flattened surface topology, where a high sensitivity of the dewetting to the local hydrophobicity/hydrophilicity of the planar protein interface was demonstrated (22).

Here, we expand these pioneering works on the nature of the protein self-assembly in water, and this is done in twofold. Firstly, we deal with a realistic protein whose surface is both topologically and chemically heterogeneous. Secondly, a whole dimerization process of two proteins is examined from the individual diffusive regime at large separations, through the approach stage in which two proteins come close to each other, to the structural adjustment stage toward compact dimer formation. Thereby, we address the role of hydrophobic and hydrophilic residues in each stage of the self-assembly process. To this end, we present structural and thermodynamic investigations on the dimer formation of full-length, 42-residue amyloid- β (A β 42) proteins in water. Aggregation of A β proteins is linked to Alzheimer's disease (23). A β dimer is of particular interest because it is the smallest neurotoxic A β oligomeric species as revealed by a recent clinical study (24). So far, computational studies have been carried out for the dimer formation of A β fragments (14, 25, 26) and for the A β dimerization in a continuum solvent (27, 28) due to the heavy computational load. Here, a spontaneous A β 42 dimerization process was captured in full atomic detail via an unbiased, explicit-water molecular dynamics (MD) simulations starting from two isolated A β 42 monomers. The simulated dimer structure was validated through the collision cross-section measured by the ion-mobility mass spectrometry (29). We then performed solvation thermodynamic and potential of mean force analyses by applying the integral-equation theory of liquids (30, 31) to simulated protein conformations. The component analysis of the hydration thermodynamic properties (32, 33) was further conducted to elucidate how hydrophobic and hydrophilic patches on the protein surface manifest themselves in the thermodynamic driving force for the A β 42 self-assembly in water.

Author contributions: S.-H.C. and S.H. designed research; S.-H.C. and S.H. performed research; S.-H.C. and S.H. analyzed data; and S.-H.C. and S.H. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

To whom correspondence should be addressed. E-mail: sihyun@sookmyung.ac.kr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1120646109/-DCSupplemental.

Results

Structural Characteristics of the Dimerization Process. We carried out A β 42 dimerization simulations at the temperature $T = 300$ K and the pressure $P = 1$ bar under neutral pH (see *Materials and Methods* and *SI Text*). The initial monomer structure was taken from our previous study (34), whose collision cross-section is computed to be $752 \text{ \AA}^2$ that is in fair agreement with the experimental value $702 \text{ \AA}^2$ (29) considering the size and complexity of A β 42 protein. Two monomers were initially placed at $45 \text{ \AA}$ apart from each other with a random orientation, and no artificial attraction force was employed between them. The dimerization process was monitored via the center-of-mass (COM) distance between two monomers, the number of intermonomer heavy atom contacts, and the collision cross-section (Fig. 1*A–C*). Three independent dimerization simulations were performed with different random initial relative orientations and velocities. All three trajectories exhibit a similar A β 42 dimerization process, except the time scales for the dimerization events vary (see *Figs. S1–S4*). In the following, we present the results based on the single representative trajectory.

The A β 42 dimerization occurs via two stages—the approach and structural adjustment stages—that stem out from diffusive dynamics (Fig. 1). Up to around 32 ns, two monomers exhibit individual diffusive motions. The collision cross-section stays around $1,500 \text{ \AA}^2$ in this time regime that is simply a double of that for monomer ($752 \text{ \AA}^2$). A transient contact is formed at around 18 ns, but it readily dissociates into two diffusing monomers. After 32 ns, a significant decrease in the COM distance is observed up to 47 ns, and we refer to this time regime as the approach stage. A large drop in the collision cross-section also occurs here that indicates that a considerable structural overlap between two monomers starts to be developed. The approach stage is not monotonous and involves relative reorientational dynamics of two monomers. The second contact that is formed at 33 ns dissociates at 42 ns because relative orientation of two monomers in this contact does not allow the development of sufficient intermonomer contacts to stabilize the dimer structure. Two monomers start to contact again at 45 ns with a different relative orientation. The number of intermonomer heavy atom contacts further increases after 47 ns, and two monomers do not dissociate till the end of our 100 ns simulation. The time regime between 47 and 100 ns is the structural adjustment stage because the most distinctive feature here is the conformational adjustment of two monomers to form favorable interactions between them. Such conformational adjustment leads to a more

compact dimer structure that is reflected in the further decrease in the collision cross-section during the structural adjustment stage. The computed collision cross-section ($1,287 \text{ \AA}^2$) of the final A β 42 dimer structure is in good agreement with the experimental value ($1,256 \text{ \AA}^2$) (29), and a similar level of agreement is achieved from other trajectories (*Figs. S1 and S3*).

Thermodynamic Features of the Dimerization Process. The two-stage nature of the A β 42 dimerization process has its thermodynamic counterpart. We computed the total protein internal energy E_u^{tot} (the sum of intramonomer and intermonomer energies) directly from the force field used for the simulations. Solvation free energy $\Delta\mu^{\text{tot}}$ and its enthalpy (Δh^{tot}) and entropy ($T\Delta s^{\text{tot}}$) components along the simulation trajectory were calculated using the integral-equation theory of liquids (30, 31) (see *SI Text*). The approach stage of the dimerization (32 to 47 ns) is characterized by the increase in the protein internal energy and by the decrease in the solvation free energy (Fig. 2*A and B*). The decrease in the solvation free energy is dominated by its enthalpy component (Fig. 2*C*). On the other hand, just the opposite changes are observed in the structural adjustment stage (47 to 100 ns). The decrease in the protein internal energy here reflects the energetic stabilization of the dimer structure. These thermodynamic features are reproduced in other independent trajectories (*Figs. S2 and S4*).

Enthalpy and entropy components of the Gibbs free energy reveal the thermodynamic driving force for the dimerization process. By combining the protein internal energy and the solvation free energy, we obtain a free energy $\mathcal{G} = E_u^{\text{tot}} + \Delta\mu^{\text{tot}}$ whose average over protein conformations yields the Gibbs free energy up to the protein configurational entropy (35). (The contribution from the pressure–volume term was found to be negligible, in which case Gibbs and Helmholtz free energies can be used interchangeably.) Because the protein configurational entropy would decrease upon dimerization, the driving force for the dimerization must come from $\mathcal{G}$. In fact, the free energy $\mathcal{G}$ decreases as the dimerization proceeds (Fig. 2*D and Table 1*). We find that the approach stage is mainly driven by its enthalpy component $\mathcal{H} = E_u^{\text{tot}} + \Delta h^{\text{tot}}$ (Fig. 2*E and Table 1*). Because the solvation enthalpy decreases while the protein internal energy increases in this stage, it is the solvation enthalpy that initiates the approach stage of the dimerization. On the other hand, the enthalpy $\mathcal{H}$ in the structural adjustment stage increases because the increase in the solvation enthalpy is found to prevail over the decrease in the protein internal energy. It is therefore a favorable increase in the

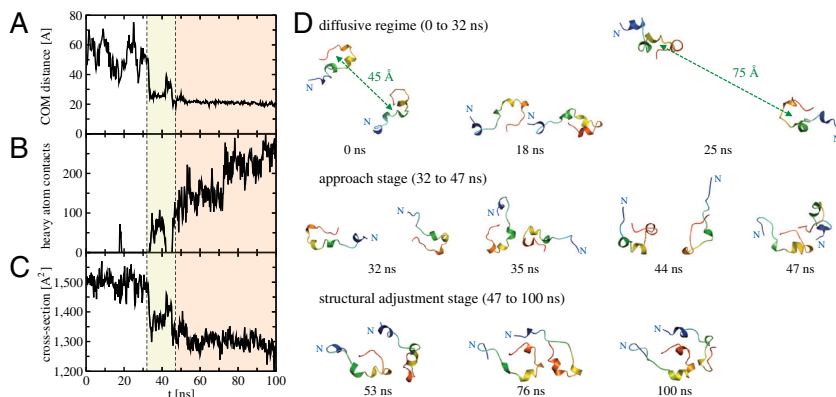


Fig. 1. Structural characteristics of the A β 42 dimerization process. (A) COM distance between two A β 42 monomers, (B) the number of intermonomer heavy atom contacts (the heavy atom contact is counted when the distance between two heavy atoms is less than $5.4 \text{ \AA}$), and (C) the collision cross-section as a function of time. Vertical dashed lines refer to 32 and 47 ns separating the diffusive regime (0 to 32 ns), the approach stage (32 to 47 ns, colored by light yellow), and the structural adjustment stage (47 to 100 ns, colored by light orange). (D) Representative A β 42 dimer conformations during the dimerization process. Each monomer structure is color-coded according to the sequence, ranging from blue to red at N and C termini, respectively.

Elucidating the Molecular Origin of Hydrolysis Energy of Pyrophosphate in Water

Jooyeon Hong,[†] Norio Yoshida,[‡] Song-Ho Chong,[†] Chewook Lee,[†] Sihyun Ham,^{*,†} and Fumio Hirata^{*,§}

[†]Department of Chemistry, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-Ku, Seoul 140-742, Korea

[‡]Department of Chemistry, Faculty of Sciences, Kyusyu University, Fukuoka 812-8581, Japan

[§]Department of Theoretical Molecular Science, Institute for Molecular Science, Okazaki 444-8585, Japan

S Supporting Information

ABSTRACT: The molecular origin of the energy produced by the ATP hydrolysis has been one of the long-standing fundamental issues. A classical view is that the negative hydrolysis free energy of ATP originates from intramolecular effects connected with the backbone P–O bond, so called “high-energy bond”. On the other hand, it has also been recognized that solvation effects are essential in determining the hydrolysis free energy. Here, using the 3D-RISM-SCF (three-dimensional reference interaction site model self-consistent field) theory that integrates the *ab initio* quantum chemistry method and the statistical mechanical theory of liquids, we investigate the molecular origin of hydrolysis free energy of pyrophosphate, an ATP analogue, in water. We demonstrate that our theory quantitatively reproduces the experimental results without the use of empirical parameters. We clarify the crucial role of water in converting the hydrolysis free energy in the gas phase determined solely by intramolecular effects, which ranges from endothermic, thermoneutral, to highly exothermic depending on the charged state of pyrophosphate, into moderately exothermic in the aqueous phase irrespective of the charged state as observed in experimental data. We elucidate that this is brought about by different natures of solute–water interactions depending on the charged state of solute species: the hydration free energy of low-charged state is mainly subjected to short-range hydrogen-bonds, while that of high-charged state is dominated by long-range electrostatic interactions. We thus provide unambiguous evidence on the critical role of water in determining the ATP hydrolysis free energy.

INTRODUCTION

ATP conversion into ADP, called ATP hydrolysis, is the source of energy in all biological functions.¹ It has originally been proposed and is still conceived widely that the negative hydrolysis free energy is associated with the backbone P–O bond in ATP, so-called “high-energy” P–O bond.^{2–13} On the other hand, the critical importance of solvation effects has also been recognized in determining the hydrolysis free energy.^{14–20} George and co-workers proposed that the hydrolysis free energies of phosphate compounds are determined mainly by the differences in solvation free energies of reactants and products, and not by intramolecular effects.¹⁴ Subsequent experimental investigation in aqueous media containing different concentrations of organic solvents has shown that the hydrolysis free energy depends strongly on the water activity.¹⁶ However, the joint role of special intramolecular properties connected with the P–O bond and hydration effects that vary significantly with the solute charged state remains to be revealed. For example, the hydrolysis free energies of various kinds of phosphate compounds in aqueous phase are experimentally found to be comparable,¹⁴ despite the large difference in sign and magnitude for the gas-phase values.⁷ Thus, the role of water, though its significance is recognized, should largely vary for different solute species, but such an effect has yet been comprehensively understood at the molecular level.

Most of the previous theoretical studies employed an implicit solvation model such as the polarizable continuum model to understand the role of water and solvation free energy of

phosphate compounds. Reliable calculations of solutes in solution must however take into account not only bulk electrostatics (i.e., long-range electrostatic polarization effects) but also shorter-range effects such as the free energy of cavity formation, dispersion, and solvent structural effects, the latter including both hydrogen bonding and exchange repulsion effects.²¹ Recently, a microsolvation-continuum approach was applied to compute solvation free energies of phosphate compounds since the simple polarizable continuum model was found to be unable to yield accurate solvation free energies.²² However, the treatment requires an empirical determination of the optimal number of water molecules to be explicitly handled in order to reproduce experimental values of the hydration free energies. Nevertheless, this work pointed out a necessity of molecular treatment of solvation effects in handling the phosphate compounds.

Here we investigate the molecular origin of hydrolysis free energy of a model compound of ATP, pyrophosphate, in water using the 3D-RISM-SCF (three-dimensional reference interaction site model self-consistent field) theory that integrates the *ab initio* quantum chemistry method and the statistical mechanical theory of liquids.^{23–25} The theory treats the electronic structure of solute and the statistical solvent distribution around it in a self-consistent manner, and properly accounts for chemical specificities of solute–water interactions such as hydrogen bonding similarly to molecular simulations

Received: February 4, 2012

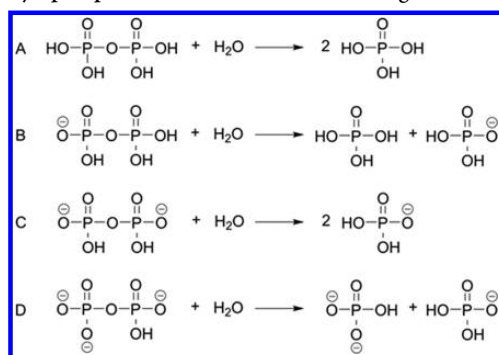
Published: June 3, 2012

with explicit solvent. We apply this method to four different charged states of pyrophosphate for which experimental data for hydrolysis free energies are available. By performing the calculations for the reaction free energies both in the gas phase and in the aqueous phase without the use of any empirical parameters, we quantitatively and unambiguously demonstrate the role of water in the hydrolysis of pyrophosphate. In particular, we clarify different natures of solute–water interactions depending on the charged state of pyrophosphate. We argue that such differences explain why the experimental results for the aqueous-phase hydrolysis free energies are moderately exothermic irrespective of the charged state of pyrophosphate.

RESULTS AND DISCUSSION

Hydrolysis Free Energies. For the four possible charged states of pyrophosphates (see Scheme 1), the computed

Scheme 1. Schematic Description of the Hydrolysis Reaction of Pyrophosphate for the Four Possible Charged States



reaction free energies of the hydrolysis reactions are summarized and compared with experimental data in Table 1.

Table 1. Reaction Free Energies in the Gas Phase and in the Aqueous Phase at 298.15 K and 1.0 atm Computed by DFT at the B3LYP/6-31+(d) Level and the 3D-RISM-SCF Theory with Units in kcal/mol

reaction	gas phase (DFT)	aqueous phase (3D-RISM-SCF)	exptl ^a
A	−1.7	−8.9	−9.5
B	21.3	−6.2	−7.5
C	−56.6	−8.1	−7.7
D	−119.5	−7.7	−7.1

^aExperimental values from ref 14.

Components of the gas-phase reaction free energies, i.e., reaction enthalpies and entropies, are listed in Table 2. The gas-phase geometries computed at the B3LYP/6-31+G(d) level are shown in Figure 1. As shown in Figure 1, the pyrophosphate molecules display maximized hydrogen bondings in a gas phase, and the average hydrogen bond distance is computed to be decreased as the total charge of pyrophosphate becomes more negative. In the gas phase, the hydrolysis reaction is close to thermoneutral for $\text{H}_4\text{P}_2\text{O}_7$ (A), endothermic for the monoanionic $\text{H}_3\text{P}_2\text{O}_7^-$ (B), and exothermic for the dianionic $\text{H}_2\text{P}_2\text{O}_7^{2-}$ (C) and the trianionic $\text{HP}_2\text{O}_7^{3-}$ (D). The results in the gas phase can be explained in terms of the localization of

Table 2. Gas-Phase Reaction Free Energies (in kcal/mol), Enthalpies (in kcal/mol), and Entropies (Multiplied by the Temperature, in kcal/mol) at 298.15 K and 1.0 atm Computed by DFT at the B3LYP/6-31+G(d) Level

reaction	ΔG_{gas}	ΔH_{gas}	$T\Delta S_{\text{gas}}$
A	−1.7	1.4	3.1
B	21.3	24.6	3.3
C	−56.6	−53.1	3.5
D	−119.5	−117.8	1.7

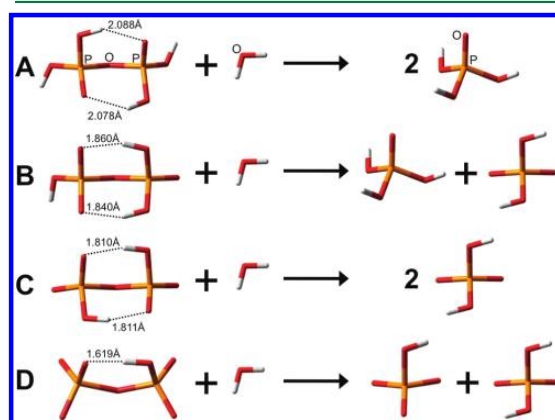


Figure 1. Optimized geometries in the gas phase of the reactants and products of the reactions A–D computed by DFT at the B3LYP/6-31+G(d) level. The dotted line and the number (in Å) denote hydrogen bonding and its distance, respectively.

excess electrons in a molecule and of the Coulomb repulsions among those electrons. In case A, the reaction free energy is vanishingly small because there is no excess electron either in the reactant or product states. In case B, the reaction becomes endothermic because an excess electron is localized in a phosphate molecule in the product. In cases C and D, the reaction becomes exothermic because excess electrons are localized more in the reactants causing electrostatic repulsion between negatively charged sites of pyrophosphates than in the products. Overall, the gas-phase hydrolysis free energies are significantly dependent on the reactant charged states. It is noted that the gas-phase reaction free energies (ΔG_{gas}) are mainly determined by the enthalpy term (ΔH_{gas}), and the relative contribution of the entropy term ($T\Delta S_{\text{gas}}$) is rather small at 298.15 K and 1.0 atm except for reaction A (Table 2). Nevertheless, the gas-phase hydrolysis free energy results are greatly deviated from the experimental data in which the hydrolysis reaction is moderately exothermic regardless of the reactant charged states.¹⁴

To obtain the hydrolysis free energy of pyrophosphates in the aqueous phase, we performed the 3D-RISM-SCF calculations^{23–25} for all molecular species in Scheme 1. The optimized geometries computed by the 3D-RISM-SCF theory are displayed in Figure 2. Compared to the gas-phase geometries, the computed structure for the neutral species of $\text{H}_4\text{P}_2\text{O}_7$ (A) in the aqueous phase shows the most distinctive difference among pyrophosphates by having no intramolecular hydrogen bonding. Since the hydrogen atoms in $\text{H}_4\text{P}_2\text{O}_7$ (A) are free from forming H-bonding with phosphate oxygens, they are available to make H-bonding to solvent water in the

Aqueous interaction site integral-equation theory that exactly takes into account intramolecular correlations

Song-Ho Chong and Sihyun Ham^{a)}

Department of Chemistry, Sookmyung Women's University, Cheongpa-ro 47-gil 100, Yongsan-Ku, Seoul 140-742, South Korea

(Received 7 June 2012; accepted 26 September 2012; published online 15 October 2012)

We report the development of a formally exact integral equation for the three-dimensional hydration structure around molecular solutes of arbitrary complexity. A distinctive feature of our theory—termed aqueous interaction site (AXIS) integral-equation theory—is that it fully takes into account the intramolecular structural correlations of solvent water, which has been missing in the previous integral-equation theories such as the three-dimensional reference interaction site model (3D-RISM) theory. With a simplifying approximation in which the intermolecular bridge function is neglected, an illustrative application of the AXIS theory is made on the equilibrium oxygen and hydrogen distributions of solvent water surrounding a solute water molecule at ambient and supercritical conditions. We demonstrate through a comparison with molecular dynamics simulation results that the inclusion of the exact intramolecular correlations improves upon the 3D-RISM theory in describing the water distribution around molecular solute, in particular near the surface region of the solute molecule, though there still remain quantitative differences from the simulation results. To further improve the quantitative accuracy of the theory, one needs to incorporate the intermolecular bridge function, and a possible formulation for the approximate bridge function is suggested based on the angular decomposition. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4758072>]

I. INTRODUCTION

Liquid integral-equation theory constitutes an important field of research in chemical physics.^{1,2} It allows us to obtain equilibrium solvent distribution around a solute, and thereby the solvation thermodynamics, without the sampling issue that arises in simulations with explicit solvent molecules.³ A particularly useful and successful theory for handling molecular solutions has been the reference interaction site model (RISM) theory pioneered by Chandler and Andersen⁴ and extended by Hirata, Pettitt, and Rossky.^{5–8} The three-dimensional version of the theory (3D-RISM) has subsequently been developed by Beglov and Roux,⁹ Cortis, Rossky, and Friesner,¹⁰ and Kovalenko and Hirata,^{11,12} which makes it possible to take into account the full distribution of the solvent around complex solutes of irregular shape in three-dimensional space. Such a three-dimensional generalization is vital for applications to chemically and biologically interesting systems since, for solute molecules with complex internal structures such as proteins, knowledge of the solvation structure at specific surface and interior regions is essential for understanding complicated and localized solute-solvent interactions. Recent applications of the 3D-RISM theory include molecular recognition processes in biomolecules,¹³ development of ligand mapping method on protein surfaces¹⁴ and screening of bioactive and biomimetic solutes¹⁵ toward computational drug design, salt-induced effects on the DNA conformational change,¹⁶ mutation effects on the aggregation thermodynamics and conformational stability of proteins,^{17,18}

uncovering the impact of chemical heterogeneity on protein self-assembly in water,¹⁹ and elucidating the molecular origin of the hydrolysis free energy of pyrophosphate, an ATP (adenosine triphosphate) analog, in water.²⁰

While the three-dimensional integral-equation theory is becoming such a tool of great utility, further development of the theory is necessary to achieve quantitative accuracy. In particular, it is well known that, when supplemented with the hypernetted-chain (HNC) closure, the 3D-RISM theory sometimes predicts unphysically high or numerically divergent first peak in the distribution function. It is this observation that motivated Kovalenko and Hirata to elaborate a hybrid of the HNC and Percus-Yevick closures, called the Kovalenko-Hirata (KH) closure, which avoids difficulties for convergence in solving the 3D-RISM equation.¹² More recently, Pettitt and co-workers incorporated an approximate bridge function to the HNC closure of the 3D-RISM equation.²¹ While this yielded a better result for a protein in terms of agreement with simulation, somewhat worse result was obtained for a solute water, i.e., the improvement was not systematic. Yet another kind of attempt for improving the 3D-RISM theory has been explored on the basis of the recognition that intramolecular bonding effects between constituent atoms of solvent are missing in the closure relation. To address the lack of such intramolecular correlations, Du, Beglov, and Roux²² introduced and optimized empirical water hydrogen and oxygen bridge functions for improving the accuracy of the 3D-RISM theory. Kovalenko and Hirata proposed a similar correction to the 3D-RISM theory, termed the repulsive bridge correction (KH-RBC), that incorporates the Boltzmann factor for a repulsive core part of the interaction potential through the

^{a)} Author to whom correspondence should be addressed. Electronic mail: sihyun@sookmyung.ac.kr.

intramolecular distribution function.²³ They found that KH-RBC drastically improves the agreement of the hydration structure and thermodynamics of hydrophobic solutes compared with simulation data. However, those empirical corrections of the closure relation do not follow from a formal and rigorous statistical mechanical treatment.

Here, we develop a three-dimensional integral-equation theory that exactly takes into account the intramolecular correlations of solvent water molecule. Since our theory builds upon the interaction site formalism, it has a distinctive structure compared to the 3D-RISM equation, and it is applicable exclusively to aqueous solutions (see below for the reason), we term it aqueous interaction site (AXIS) integral-equation theory. The derivation of the AXIS equation is based on the density functional theory (DFT) of non-uniform polyatomic fluid formulated by Chandler, McCoy, and Singer.^{24,25} One of the greatest advantages of adopting the DFT is that, by employing the methods of functional differentiation and integration, one can obtain in a straightforward manner formally exact results for distribution functions and thermodynamic functions.²⁶ This feature is particularly suitable for the interaction site formalism of molecular fluid with which it is non-trivial to obtain diagrammatically proper integral equation that practically improves upon the RISM result.^{27,28} Through an illustrative application of the AXIS theory, we investigate the effect of incorporating full intramolecular correlations on the equilibrium hydration structure of molecular solute.

The paper is organized as follows. In Sec. II, we derive the AXIS integral-equation theory based on the DFT of non-uniform polyatomic fluid. We will also derive with a common perspective the previous related theories—the 3D-RISM and rational-closure theories—to highlight the distinctive features of our theory. An illustrative application of the AXIS theory is presented in Sec. III, where results are also compared with those from the 3D-RISM theory and molecular dynamics simulation. Main findings of the present work and possible further extension of the theory are discussed in Sec. IV. Section V concludes the paper. Appendix A is devoted to the introduction and discussion of higher-order direct correlation functions. Appendix B presents technical details in evaluating the double integral that is necessary to solve the AXIS integral equation.

II. THEORY

A. Density functional theory

We start from summarizing the DFT formulated in Ref. 24 and its generalization incorporating higher-order direct correlation functions to an extent that is necessary to understand the main result of the present article. Details on the higher-order direct correlation functions are presented in Appendix A. We consider as a solvent a one-component molecular fluid and employ the standard notation of the interaction site formalism.²⁹ In particular, $\rho_\alpha(\mathbf{r})$ stands for the density at position $\mathbf{r}$ of solvent site α . The equilibrium ensemble average of the solvent density field in the presence of a molecular solute, which can be thought of as a source generating external potential energy fields $u_\alpha(\mathbf{r})$, is denoted by $\langle \rho_\alpha(\mathbf{r}) \rangle$. We

do not claim any specific functional form of $u_\alpha(\mathbf{r})$: it typically consists of the Lennard-Jones and Coulomb electrostatic potential functions of standard classical force fields, but continuous *ab initio* charge distributions obtained from solute electronic wave functions can also be taken into account. We work with the grand canonical ensemble characterized by the chemical potential μ , volume V , and temperature T . The grand potential $\log \Xi$ is a natural functional of the fields

$$\psi_\alpha(\mathbf{r}) = \beta\mu_\alpha - \beta u_\alpha(\mathbf{r}), \quad (1)$$

where $\beta^{-1} = k_B T$ with k_B denoting Boltzmann's constant, and μ_α is the chemical potential of the solvent site α . The condition of chemical equilibrium is represented as $\mu = \sum_\alpha \mu_\alpha$.

The principal result of the DFT formulated by Chandler, McCoy, and Singer²⁴ is that the equilibrium density fields $\langle \rho_\alpha(\mathbf{r}) \rangle$ can be determined from the variational principle

$$\frac{\delta \mathcal{A}}{\delta \langle \rho_\alpha(\mathbf{r}) \rangle} = \psi_\alpha(\mathbf{r}). \quad (2)$$

Here, $\mathcal{A}$ is the free energy functional, which is obtained by the Legendre transformation of $\log \Xi$, and is a natural functional of $\langle \rho_\alpha(\mathbf{r}) \rangle$. Its formal expression reads

$$\mathcal{A} = \sum_\alpha \int d\mathbf{r} \langle \rho_\alpha(\mathbf{r}) \rangle [\log \langle \rho_\alpha(\mathbf{r}) \rangle \lambda^3 - 1] + \mathcal{A}_{\text{bond}} + \Delta \mathcal{A}, \quad (3)$$

where the first term represents the standard entropy density functional for a gas of uncorrelated particles, the second term ($\mathcal{A}_{\text{bond}}$) describes the effects of the intramolecular bonding between constituent particles, and the third term ($\Delta \mathcal{A}$) accounts for the intermolecular interactions. Here, λ , which has units of length, is chosen so that the chemical potential $\mu^{(0)}$ of an ideal-gas system of uniform density ρ reads $\beta\mu^{(0)} = \log \rho \lambda^3$: λ can then be specified by inertia parameters of a solvent molecule, whose value however has no effect on the equilibrium density fields. To separately deal with intramolecular and intermolecular effects, we introduce the functionals $\mathcal{D}_\alpha$ and $\mathcal{C}_\alpha$ such that

$$\mathcal{D}_\alpha(\mathbf{r}) \equiv -\frac{\delta \mathcal{A}_{\text{bond}}}{\delta \langle \rho_\alpha(\mathbf{r}) \rangle}, \quad \mathcal{C}_\alpha(\mathbf{r}) \equiv -\frac{\delta \Delta \mathcal{A}}{\delta \langle \rho_\alpha(\mathbf{r}) \rangle}. \quad (4)$$

A formal solution to the variational principle (2) can then be expressed as

$$\langle \rho_\alpha(\mathbf{r}) \rangle = \lambda^{-3} \exp [\psi_\alpha(\mathbf{r}) + \mathcal{D}_\alpha(\mathbf{r}) + \mathcal{C}_\alpha(\mathbf{r})]. \quad (5)$$

With n denoting the number of interaction sites in a solvent molecule, a formally exact expression for the intramolecular functional $\mathcal{D}_\alpha(\mathbf{r})$ is given by²⁵

$$\begin{aligned} \mathcal{D}_\alpha(\mathbf{r}) = \log \int d\mathbf{r}_1 \cdots \int d\mathbf{r}_n s_{1\dots n}(\mathbf{r}_1, \dots, \mathbf{r}_n) \\ \times \left\{ \delta(\mathbf{r}_\alpha - \mathbf{r}) \prod_{\gamma(\neq \alpha)} \exp [\psi_\gamma^{(0)}(\mathbf{r}_\gamma)] \right\}. \end{aligned} \quad (6)$$

Here, $s_{1\dots n}(\mathbf{r}_1, \dots, \mathbf{r}_n)$ represents the n -site intramolecular distribution function,²⁹ i.e., the probability distribution for finding within the same molecule the n different sites (labeled

Human IgA-binding Peptides Selected from Random Peptide Libraries

AFFINITY MATURATION AND APPLICATION IN IGA PURIFICATION^{*(S)}

Received for publication, June 8, 2012, and in revised form, October 17, 2012. Published, JBC Papers in Press, October 17, 2012, DOI 10.1074/jbc.M112.389742

Takaaki Hatanaka[‡], Shinji Ohzono[§], Mirae Park[¶], Kotaro Sakamoto^{‡,1}, Shogo Tsukamoto[‡], Ryohei Sugita[‡], Hiroyuki Ishitobi[§], Toshiyuki Mori^{||,1}, Osamu Ito[§], Koichi Sorajo[§], Kazuhisa Sugimura[‡], Sihyun Ham^{¶,2}, and Yuji Ito^{‡,3}

From the [‡]Graduate School of Science and Engineering, Kagoshima University, Kagoshima 890-0065, Japan, the [§]General Research Laboratory, Otsuka Chemical Co., Ltd., 771-0193, Japan, the [¶]Department of Chemistry, Sookmyung Women's University, Hyochangwon-gil 52, Yongsan-gu, Seoul 140-742, Korea, and the ^{||}Molecular Targets Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, Maryland 21702

Background: Pharmaceutical application of human IgA requires a highly specific IgA purification system.

Results: A peptide affinity ligand for IgA was designed and optimized for affinity/specificity using randomized phage libraries and mutational studies.

Conclusion: The designed IgA-binding peptide has high affinity and specificity for human IgA.

Significance: This IgA-binding peptide can be used for specific purification of human IgA.

Phage display system is a powerful tool to design specific ligands for target molecules. Here, we used disulfide-constrained random peptide libraries constructed with the T7 phage display system to isolate peptides specific to human IgA. The binding clones (A1–A4) isolated by biopanning exhibited clear specificity to human IgA, but the synthetic peptide derived from the A2 clone exhibited a low specificity/affinity ($K_d = 1.3 \mu\text{M}$). Therefore, we tried to improve the peptide using a partial randomized phage display library and mutational studies on the synthetic peptides. The designed Opt-1 peptide exhibited a 39-fold higher affinity ($K_d = 33 \text{ nM}$) than the A2 peptide. An Opt-1 peptide-conjugated column was used to purify IgA from human plasma. However, the recovered IgA fraction was contaminated with other proteins, indicating nonspecific binding. To design a peptide with increased binding specificity, we examined the structural features of Opt-1 and the Opt-1-IgA complex using all-atom molecular dynamics simulations with explicit water. The simulation results revealed that the Opt-1 peptide displayed partial helicity in the N-terminal region and possessed a hydrophobic cluster that played a significant role in tight binding with IgA-Fc. However, these hydrophobic residues of Opt-1 may contribute to nonspecific binding with other proteins. To increase binding specificity, we introduced several mutations in the hydrophobic residues of Opt-1. The resultant Opt-3 peptide

exhibited high specificity and high binding affinity for IgA, leading to successful isolation of IgA without contamination.

IgA is most abundantly produced in the human body, particularly in the mucous membrane, and plays an important role in mucosal immunity, which is the first line of defense against bacterial and viral invasion (1). This protein is also the second most abundant immunoglobulin in human plasma after IgG (2). IgA is secreted in the mucus as a dimer (secretory form of IgA) connected by a joining chain via intermolecular disulfide bridges, whereas the major form of IgA in plasma is monomeric (3). IgA has two subclasses: IgA1 and IgA2. The latter lacks the 13-amino acid proline-rich sequence present in the hinge region of IgA1 and therefore is more resistant to bacterial proteases (4).

Immunological studies on IgA have focused on vaccine development for mucosal immunity to control bacterial and viral infections at the mucosal site (5, 6). The Fc receptor for IgA (CD89 or Fc α RI) has been identified and characterized (7, 8). This receptor is constitutively expressed on the surface of immune cells, including monocytes/macrophages, polymorphonuclear leukocytes, neutrophils, and eosinophils (9). The IgA receptor can mediate the various activities, including phagocytosis, oxidative burst, cytokine release, and antibody-dependent cellular cytotoxicity performed by these immune effector cells (10). Although IgG1 is popularly used as anti-cancer therapeutic antibody at present, IgA may also be a potential candidate for anti-tumor therapeutics given its antibody-dependent cellular cytotoxicity ability. Indeed, a recombinant IgA specific to tumor cell surface markers kills tumor cells in the presence of whole blood or neutrophils (11, 12). Furthermore, human IgA (hIgA)⁴ induces antibody-dependent cellular cytotoxicity more

* This work was supported by funds from the Japan Science and Technology Agency. This work was also supported by the Basic Science Research Program of the National Research Foundation of Korea funded by Ministry of Education, Science and Technology Grant 2011-0012096 and by Japan Society for the Promotion of Science Grant-in-Aid 238036 (to T. H.).

^(S) This article contains supplemental Figs. S1 and S2.

¹ Present address: Pharmaceutical Research Division, Takeda Pharmaceutical Company Limited, Fujisawa 251-8555, Japan.

² To whom correspondence may be addressed. Tel.: 82-02-710-9410; Fax: 82-02-2077-7321; E-mail: sihyun@sookmyung.ac.kr.

³ To whom correspondence may be addressed: Graduate School of Science and Engineering, Kagoshima University, Kagoshima 890-0065, Japan. Tel.: 81-99-285-8110; Fax: 81-99-285-8110; E-mail: k2174603@kadij.jp.

⁴ The abbreviations used are: h, human; MD, molecular dynamics; Fmoc, N-(9-fluorenyl)methoxycarbonyl; SPR, surface plasmon resonance; Fc α R, Fc α receptor.

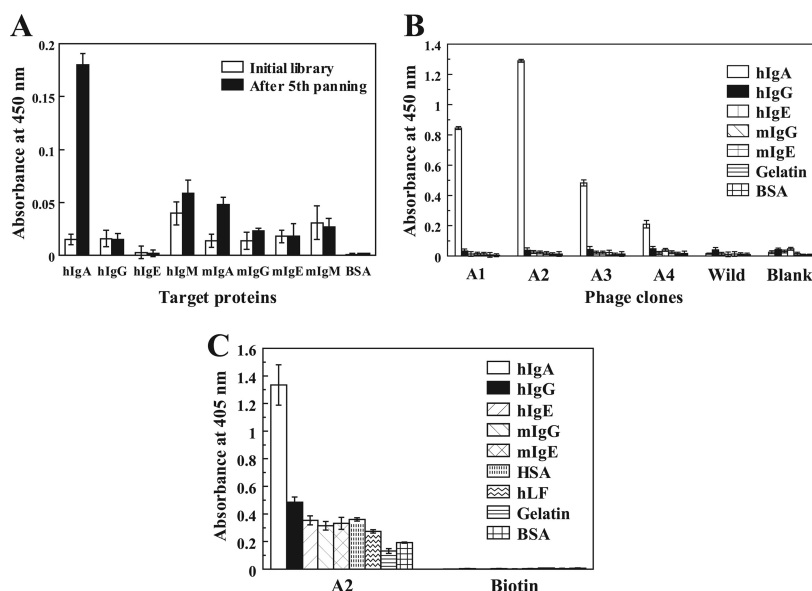


FIGURE 1. Isolation of hIgA-specific phage clones from T7 phage-displayed random peptide libraries. A, binding specificities of the phages from the initial library and the $X_3CX_{7-10}CX_3$ library enriched by five rounds of biopanning against hIgA were examined by ELISA. B, binding specificities of the four phage clones (A1–A4) isolated by biopanning were confirmed by ELISA. C, binding specificities of the A2 synthetic peptide were examined by ELISA. *hlg*, human immunoglobulin; *mIg*, mouse immunoglobulin; *HSA*, human serum albumin; *hLF*, human lactoferrin.

effectively against solid tumor cells or lymphoma than the relevant IgG (11, 13). These findings encourage the development of production methodologies, including cell culture (14) and purification systems, for the clinical use of IgA.

Several IgA purification methods have been reported. Jacalin, a lectin targeting IgA1-specific sugar (a galactosyl β -1,3-*N*-acetylgalactosamine) has been used as an affinity ligand for IgA purification (15). A conventional method using ion exchange and size exclusion chromatography has also been reported (16). However, these methods are not suitable for antibody purification on an industrial scale, like the protein A column for IgG purification. On the other hand, specific IgA-binding proteins have been identified in *Staphylococcus* group A (*Staphylococcus pyogenes*) and group B bacteria (17, 18). Typically, the M proteins from *S. pyogenes*, M22, Arp4, or Sir22, exhibit binding activities toward IgA1 and A2 or the monomeric and secretory forms of IgA. These proteins recognize the interdomain region between CH2 and CH3 of IgA-Fc and can inhibit the interaction between IgA and IgA receptor (19–22). A simple and highly efficient affinity purification method using IgA-binding proteins was previously described (16). However, the use of bacterial proteins for pharmaceutical antibody purification requires careful attention to prevent contamination with endotoxin or bacterial proteins, given their highly toxic and antigenic nature, as described for the protein A/G column used in IgG purification for pharmaceutical use (23).

To solve this problem, synthetic low molecular weight ligands, such as TG19318 (24) and protein A mimetic compounds (25), have been investigated. However, these compounds still suffer from insufficient specificity and affinity for use in IgA purification. Sandin *et al.* (26) reported the purification of hIgA using a synthetic peptide comprising 50 residues extracted from the IgA-binding domain (Sap) of the Sir22 M

protein. Using a dimerized form of this peptide (peptide M), Sandin *et al.* successfully detected and purified hIgA; however, the dimerized peptide of 100 residues is too large for industrial applications.

Here, we report a novel hIgA-binding peptide that was isolated from random peptide T7 phage libraries by biopanning against hIgA. The essential residues in the peptide were identified, and the peptide was optimized for affinity/specificity. Our peptide is only 16 residues long and exhibits high specificity/affinity for hIgA, which is suitable for hIgA purification.

EXPERIMENTAL PROCEDURES

Materials—Polyclonal hIgA1/IgA2, IgE, and IgG were purchased from Acris Antibodies GmbH (Herford, Germany), Athens Research & Technology (Athens, GA), and ICN/Cappel Biomedicals (Aurora, OH), respectively. Anti-HER2 IgG1 humanized antibody, Trastuzumab (Herceptin), and anti-human IL13-specific hIgA2 were from Chugai Pharmaceutical Corp. Ltd. (Tokyo, Japan) and Invivogen (San Diego, CA). The recombinant human Fc α R/CD89 was obtained from R&D Systems (Minneapolis, MN). Mouse IgG and IgE were from PharMingen (San Diego, CA).

Construction of the T7 Phage Display Library and Biopanning—The T7 phage libraries displaying typically $X_3CX_{7-10}CX_3$ random peptides, where *X* represents the randomized amino acid positions generated using mixed oligonucleotides on template DNA, were constructed using the T7 Select vector 10–3b from Merck (Tokyo, Japan), according to methods described previously (27).

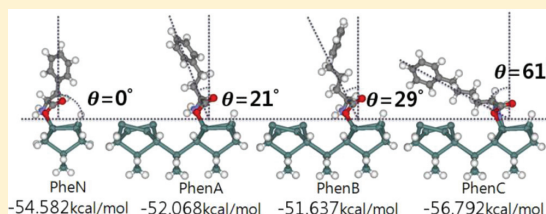
Microplate wells (Nunc Maxisorp) were coated with polyclonal hIgA (1 μ g/300 μ l/well) and blocked with 0.5% BSA in PBS. The T7 phage libraries (5×10^{10} pfu) of $X_3CX_{7-10}CX_3$ were incubated for 1 h in wells coated with hIgG and HSA to

Role of Chain Length in the Adsorption Structures and Geometric Configurations of Phenylalanine Derivatives on Ge(100) Surfaces

Youngchan Park, Heeseon Lim, Sena Yang, and Hangil Lee*

Department of Chemistry, Sookmyung Women's University, Seoul 140-742, Republic of Korea

ABSTRACT: The variation of the adsorption structures and geometric configurations among phenylalanine derivatives (2-amino-4-phenylpropanoic acid (PhenA), 2-amino-5-phenylpropanoic acid (PhenB), and 2-amino-6-phenylpropanoic acid (PhenC)) adsorbed on Ge(100) surfaces was investigated using density functional theory (DFT) calculations. The chain length of a linker bonded to the α -carbon of the phenylalanine derivatives influenced the preferred adsorption structure in self-assembled monolayers (SAMs). We confirmed that the "O–H dissociated-N dative bonded structure" is the most favorable structure among the five possible adsorption structures. Interestingly, geometric differences in the adsorption configurations indicated that the phenyl rings of PhenA and PhenB were tilted with respect to the Ge(100) surface, with an angle 30° from the surface normal. In contrast, the phenyl ring of PhenC was tilted with an angle of 60° with respect to the surface normal. We therefore confirmed that PhenA and PhenB are available to use in SAMs considering the geometrical configurations and the adsorption energies. The theoretical results were further supported by conducting DFT calculations of the reaction pathways, leading to the favored adsorption geometries.



1. INTRODUCTION

Over the past several decades, semiconductor-based materials and devices have become important components in modern technology. In this sense, many researchers pay attention to functionalization of organic molecules on semiconductors. Particularly important are the interactions between semiconductors and bioorganic molecules, such as amino acids. Amino acids are the fundamental components of proteins and play a vital role in the development of biochips, biosensors, and microelectronic devices.^{1–5} Biosensors and electrochemical devices are an important potential application of self-assembled monolayers (SAMs), and SAMs have attracted significant interest over the past several years. However, it is notable that only a few examples of organized SAMs bonded directly to bare semiconductor surfaces have been reported because examples of highly organized films almost exclusively involve oxides and coinage metal substrates.^{6–9}

In this paper, we used a Ge(100)-2 $\times$ 1 surface as a semiconductor substrate to observe the interactions between bioorganic molecules and semiconductor surfaces. The adsorption structures and geometric configurations of phenylalanine derivatives were calculated using density functional theory (DFT). Germanium substrates permit a greater product selectivity than is achieved using any other semiconductor surface due to the tendency of germanium to impose relatively large activation energy barriers on surface reactions.^{10–12} Large activation energy barriers are an important factor for preparing stable multifunctional systems, for example, amino acids, which can react via many possible routes. Precise interfacial control can be achieved using atomically clean surfaces, which can facilitate such applications. Because Ge(100) surfaces hold a weaker π -bond between pairs of surface dimer atoms, Ge dimers are buckled away from a

symmetric configuration at room temperature. The tilted dimers produce an unequal distribution of charge, resulting in zwitterion-like properties.^{13,14} The "up" Ge atom is an electron-rich nucleophile, and the "down" Ge atom is an electron-poor electrophile. These properties provide a framework for understanding the adsorption reactions of amino acids that present several functional groups at a semiconductor surface.^{10–13}

Here, we begin by introducing adsorption structures and geometric configurations of phenylalanine and three of its derivatives on Ge(100) surfaces. Phenylalanine is an essential amino acid and plays a critical role in our body because it provides the starting material for the synthesis of tyrosine, which is involved in several brain chemicals.^{14,15} As an amino acid, phenylalanine includes a carboxyl group ($-\text{COOH}$), an amino group ($-\text{NH}_2$), a hydrogen atom ($-\text{H}$), and a benzyl ring as a side chain. Previously, our group examined the electronic configurations and structures of several amino acids adsorbed onto Ge(100)-2 $\times$ 1 surfaces. We observed that amino acids form the "O–H dissociated-N dative bonded structure" as the most stable adsorption structure.^{16–21}

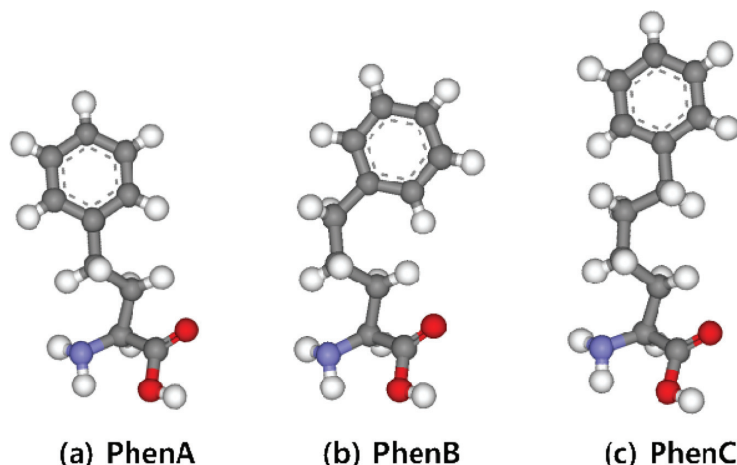
In this study, phenylalanine (2-amino-3-phenylpropanoic acid) will be referred to as PhenN, 2-amino-4-phenylpropanoic acid as PhenA (Scheme 1(a)), 2-amino-5-phenylpropanoic acid as PhenB (Scheme 1(b)), and 2-amino-6-phenylpropanoic acid as PhenC (Scheme 1(c)). The phenylalanine derivatives are distinct in the number of carbon atoms present. PhenN includes a benzyl ring attached to the α -carbon of the amino acid. PhenA includes one more carbon than is present in PhenN, and

Received: March 27, 2012

Revised: May 17, 2012

Published: May 23, 2012

Scheme 1. (a) 2-Amino-4-phenylpropanoic Acid (PhenA), (b) 2-Amino-5-phenylpropanoic Acid (PhenB), and (c) 2-Amino-6-phenylpropanoic Acid (PhenC)^a



^aEach gray, blue, red, and white-colored ball indicates carbon, nitrogen, oxygen, or hydrogen, respectively.

the benzyl ring of PhenA attaches to the β -carbon of the amino acid. The benzyl rings of PhenB and PhenC attach to the γ -carbon and δ -carbon of the amino acids, respectively. The adsorption energies and geometric configurations of the phenylalanine derivatives were examined using DFT calculations. Although these systems were not examined experimentally, the theoretical studies are meaningful because they provide important information about the adsorption structures and geometric configurations of the various amino acid derivatives when it applies to SAMs.

Here, we focus on the role of the side chain length on the molecular tilt angle with respect to the surface normal upon adsorption onto Ge(100) surfaces. In addition, we considered the possibility of self-assembled structure for phenylalanine derivatives. SAMs form spontaneously during immersion of an appropriate metal or semiconductor substrate into a solution or gas-phase medium. Widely known, single molecules are stable on inorganic surfaces with a tilt angle of 30° relative to the surface normal.²² π - π interactions with the phenyl ring can contribute to self-assembly, which suggests that phenylalanine derivatives might be capable of well-ordered self-assembly as well.^{23,24} The computational study described here furthers our understanding of surface chemistry and contributes to the obtainment of critical information for considerable development of applications. To our knowledge, no prior research has examined phenylalanine derivatives on Ge(100) surfaces.

II. COMPUTATIONAL DETAILS

DFT calculations were conducted to predict the energetics of the reaction pathways and geometrically optimized structures of the phenylalanine derivatives adsorbed onto Ge(100) surfaces. All DFT adsorption energy calculations were carried out using the JAGUAR 9.1 software package employing a hybrid density functional method that included Becke's three-parameter nonlocal exchange functional with the correlation functional of Lee–Yang–Parr (B3LYP).²⁵ These calculations considered four dimer (Ge_3H_{32}) cluster models. The geometries corresponding to the important local minima on the potential energy surface were determined at the B3LYP/LACVP** level of theory. The LACVP** basis set is a mixed basis set that uses the LACVP

basis set to describe Ge atoms and the 6-31G basis set to describe the remaining atoms. The LACVP basis set describes atoms beyond Ar in the periodic table using the Los Alamos effective core potentials developed by Hay and Wadt.²⁶ For each cluster, optimization was performed by fixing the bottom two layers of the Ge atoms in ideal Ge crystal positions while allowing the top layer of the Ge atoms (including the dimer atoms) and the atoms of the chemisorbed adsorbate to relax. The geometries of important local minima and transition states on each energy surface were calculated. Local minima and transition states were confirmed to be optimized using the same basis sets.²⁷

III. RESULTS AND DISCUSSION

DFT calculations were performed to determine the optimized configurations and reaction pathways associated with the adsorption of phenylalanine derivatives onto Ge(100) surfaces. On the basis of previous phenylalanine studies, we first examined the five possible adsorption structures for the three phenylalanine derivatives adsorbed on Ge(100) surfaces. The following five configurations were included in this examination for PhenA: (1) N dative bonded structure, as shown in Figure 1(a); (2) O dative bonded structure, as shown in Figure 1(b); (3) N–H dissociated structure, as shown in Figure 1(c); (4) O–H dissociated structure as shown in Figure 1(d); (5) O–H dissociated–N dative bonded structure, as shown in Figure 1(e). In addition, DFT calculations were performed to model each of the five possible adsorption structures of PhenB and PhenC.

Figures 1(a) and 1(b) show that the nitrogen atom of the amine group or the oxygen atom of the carboxyl group may bond to the Ge(100) surface through a dative bonding configuration. The –NH moiety of the amine group and the –OH moiety of the carboxyl group may also potentially be bonded to the Ge(100) surface via N–H dissociation (Figure 1(c)) and O–H dissociation (Figure 1(d)), respectively. Moreover, we considered the “O–H dissociated–N dative bonded structure” as well (Figure 1(e)). Toward a comprehensive understanding, we calculated the adsorption energies of all possible structures. The specific values of every adsorption structure are listed in Table 1, for phenylalanine derivatives when adsorbed onto Ge(100) surfaces.

Comparison and Contrast Analysis of Adsorption Geometries of Phenylalanine versus Tyrosine on Ge(100): Effect of Nucleophilic Group on the Surface

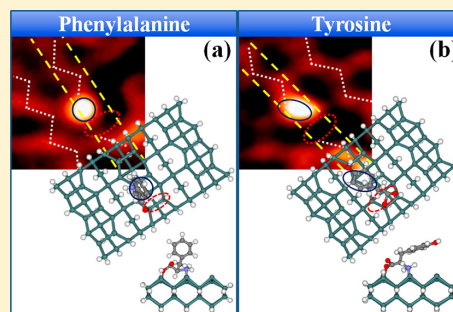
Sena Yang,[†] Heeseon Lim,[†] Eun Hee Park,[†] Yaewon Kim,[‡] Young Hwan Min,[†] Hee-Seung Lee,^{*,†} Sehun Kim,^{*,†} and Hangil Lee^{*,‡}

[†]Molecular-Level Interface Research Center, Department of Chemistry, KAIST, Daejeon 305-701, Republic of Korea

[‡]Department of Chemistry, Sookmyung Women's University, Seoul 140-742, Republic of Korea

Supporting Information

ABSTRACT: The discrepancy of geometric configuration between phenylalanine and tyrosine adsorbed on Ge(100) surfaces was investigated using scanning tunneling microscopy (STM) in conjunction with density functional theory (DFT) calculations and core-level photoemission spectroscopy (CLPES). The study focused on the role of nucleophilic group (hydroxyl group) on phenyl ring of tyrosine, and we elucidated the difference of the adsorption geometry between phenylalanine and tyrosine on Ge(100) surfaces. We first confirmed that the “O–H dissociated–N dative bonded structure” was the most favorable structure in both molecules at low coverage by results of CLPES and DFT calculations. Geometric differences for the adsorption configurations between phenylalanine and tyrosine were observed: the phenyl ring of phenylalanine was aligned axially with respect to the Ge(100) surface, whereas that of tyrosine was tilted, as determined by DFT calculations. In sequence, we found out the results of STM images to confirm DFT results. We determined the different geometric configurations are attributed to the nucleophilic hydroxyl group of tyrosine, which creates an uneven charge distribution.



1. INTRODUCTION

As the industrial applications of semiconductors are extended, research into the chemical modifications and molecular adsorption structures adsorbed on semiconductor surfaces has become important in a variety of industrial and academic fields.^{1–3} Organic materials may be functionalized to tune their chemical properties via substitution or addition reactions, and the adsorption of such organic materials may be readily exploited in applications such as biosensors, biochips, and electronic devices.^{4–8} Recently, several studies have examined the adsorption of a variety of bioorganic molecules on surfaces.^{9–12} Specially, research into the adsorption of amino acids, which include carboxyl, amino, sulfur, or a variety of functional groups, onto semiconductor surfaces is expected to provide fundamental knowledge of semiconductor functionalization for various biorelated applications.^{13–16} Among various semiconductors, germanium (Ge) is a potential candidate for the new channel material because of its superior electronic properties.¹⁷ Specifically, Ge(100)-2×1 reconstructed surfaces feature tilted dimers with a strong σ bond and a weak π bond at room temperature. The tilted dimers create an uneven charge distribution over the dimer, resulting in an electron-rich (nucleophilic) “up” atom and an electron-deficient (electrophilic) “down” atom, and thus the up and down atoms of the

dimers show zwitterionic character and function as a Lewis base and acid, respectively, which makes Ge(100)-2×1 reconstructed surfaces an excellent substrate for various adsorption reactions with multifunctionalized molecule.^{4,18}

We have examined the adsorption structures and behaviors of 10 natural amino acids on Ge(100)-2×1 surfaces.^{19–26} We have identified the general trend among the adsorption structures of amino acids that the “O–H dissociated–N dative bonded structure” is the most stable structure at low coverage. We also have investigated the effects of functional groups and steric hindrance on the adsorption geometries to explain the observed differences.

In this study, we focused on the differences in geometric configurations between phenylalanine and tyrosine in spite of their structural similarities when they adsorb on Ge(100) surfaces. Phenylalanine and tyrosine are composed primarily of a central carbon atom bonded to a carboxyl group (–COOH), an amino group (–NH₂), and either a phenyl or a phenol ring, respectively. Except for the presence of the hydroxyl group on the phenyl ring of tyrosine, the two amino acids have identical

Received: August 30, 2012

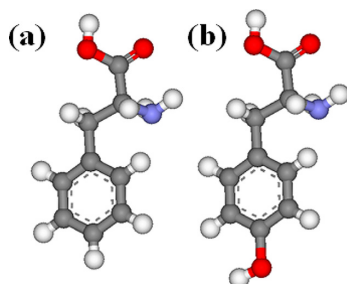
Revised: November 15, 2012

Published: November 21, 2012

structures. We assume that the hydroxyl group of tyrosine differentiates the two amino acids adsorption configurations.

The goal of this study is to clarify a discrepancy of geometric configurations between phenylalanine (Scheme 1a) and

Scheme 1. Models of the Chemical Structures of (a) Phenylalanine and (b) Tyrosine^a



^aThe gray, blue, red, and white balls indicate carbon (C), nitrogen (N), oxygen (O), and hydrogen (H), respectively.

tyrosine (Scheme 1b) adsorbed on the Ge(100) surface and to identify how the hydroxyl group (nucleophilic site) of tyrosine affects the geometric configurations in the adsorption process. Information about the adsorption structures and geometric configurations of phenylalanine and tyrosine on Ge(100) surfaces was obtained using core-level photoemission spectroscopy (CLPES) in conjunction with density functional theory (DFT) calculations and scanning tunneling microscopy (STM). We determined the effects of the $-OH$ group of tyrosine on the adsorption geometry in terms of the interactions between the hydroxyl group (nucleophile) and adjacent down dimer (electrophile) of the Ge(100) surface. To our knowledge, this is the first experimental and theoretical characterization of the adsorption structures and geometrical configurations of phenylalanine and tyrosine on Ge(100) surfaces using CLPES, DFT calculations, and STM techniques.

2. EXPERIMENTAL AND COMPUTATIONAL SECTION

Ge(100) surfaces (p -type, $R = 0.10\text{--}0.39\ \Omega$) were cleaned by several cycles of sputtering with 1 keV Ar^+ ions at 700 K for 20 min, followed by annealing at 900 K for 10 min. The cleanliness of the Ge(100)- 2×1 surfaces was checked using low energy electron diffraction (LEED) and STM. Phenylalanine ($C_9H_{11}NO_2$, 99% purity) and tyrosine ($C_9H_9NO_3$, 99% purity) were purchased from Aldrich and were further purified through several sublimation and pumping cycles to remove dissolved gases prior to exposure to the Ge(100) surface.

We performed CLPES at the 10D beamline of the Pohang Accelerator Laboratory. The C 1s, N 1s, and O 1s core-level spectra were obtained with a PHOIBOS 150 electron energy analyzer equipped with a two-dimensional charge-coupled device (2D CCD) detector (Specs GmbH) using photon energies of 340, 460, and 588 eV to enhance the surface sensitivity. The binding energies of the core-level spectra were calibrated with respect to the clean Au 4f core-level spectrum (84.0 eV) collected at the same set of photon energies. The base pressure in the chamber was maintained below 9.5×10^{-11} Torr. All spectra were recorded in the normal emission mode. The core-level spectra were carefully analyzed using a standard nonlinear least-squares fitting procedure with Voigt functions.²⁷

DFT calculations were used to analyze the energetics of the potential pathways for the tyrosine and phenylalanine reactions on the Ge(100) surface. All DFT calculations of the adsorption energies were performed using the JAGUAR 9.1 software package, which applies a hybrid density functional method and includes Becke's three-parameter nonlocal-exchange functional with the correlation functional of Lee–Yang–Parr (B3LYP).²⁸ The calculations were performed using a six-dimer (Ge_6H_{40}) cluster model. The geometries of important local minima on the potential energy surface were determined at the B3LYP/LACVP** level of theory. The LACVP** basis set is a mixed basis set that uses the LACVP basis set to describe the Ge atoms and the 6-31G basis set to describe the remaining atoms. The LACVP basis set is useful for describing atoms heavier than Ar in the periodic table and is based on the Los Alamos effective core potentials developed by Hay and Wadt.²⁹ Optimization of each cluster was performed by fixing the bottom two layers of the Ge atoms in the ideal Ge crystal positions, permitting the top layer of Ge atoms (including the dimer atoms) and the atoms of the chemisorbed adsorbate to relax. The geometries of important local minima and transition states on each energy diagram were calculated. The local minima and transition states in the optimized structures were verified using the same set of basis sets.³⁰

STM observations were performed in an ultrahigh vacuum (UHV) chamber equipped with an OMICRON VT-STM instrument at a base pressure below 1.0×10^{-10} Torr. All STM images were recorded with electrochemically etched W-tips at bias voltages of $V_s = -2.0$ V and a tunneling current of $I_t = 0.1$ nA to obtain the filled-state images.

3. RESULTS AND DISCUSSION

We first acquired CLPES data for phenylalanine and tyrosine adsorbed on Ge(100) surfaces. Figure 1 shows the C 1s, N 1s, and O 1s core-level spectra being included in both molecules. The top and bottom panels indicate the three core-level spectra of the phenylalanine and tyrosine adsorbed on the Ge(100) surfaces at 300 K. Figure 1a shows the C 1s core-level spectrum obtained after 0.15 monolayer (ML) phenylalanine deposition on the Ge(100) surface (at low coverage: below half

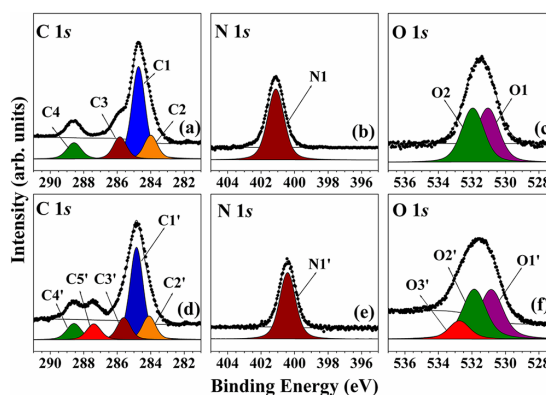


Figure 1. C 1s, N 1s, and O 1s core-level spectra (a)–(c) of phenylalanine and (d)–(f) of tyrosine obtained at a low coverage (0.15 ML) deposition on a Ge(100) surface. The dots indicate experimental values, and the solid lines represent the results of peak fitting.

Bioinspired Morphogenesis of Highly Intricate and Symmetric Silica Nanostructures

Jaeseok Yi,[†] Hee Su Jang,[‡] Jin Seok Lee,^{*,‡} and Won Il Park^{*,†}

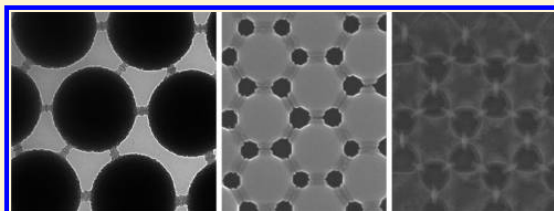
[†]Division of Materials Science Engineering, Hanyang University, Seoul 133-791, Korea

[‡]Department of Chemistry, Sookmyung Women's University, Seoul 140-742, Korea

S Supporting Information

ABSTRACT: Biosilification is of interest due to its capability to produce a highly intricate structure under environmentally friendly conditions. Despite the considerable effort that has been devoted toward biomimetic silification, the synthesis of highly complex silica structures, as found in the structures of diatom cell walls, is still in its infancy. Here, we report the bioinspired fabrication of well-organized and symmetric silica nanostructured networks, involving phase separation and silicic acid polymerization processes, in analogy to the morphogenesis of diatom cell walls. Our approach exploits self-assembled silica spheres as a self-source of the silicic acids as well as scaffolds that, interplayed with droplets of ammonium hexafluorosilicate, direct the site-specific silification. Moreover, we have achieved multiple morphological evolutions with subtle changes in the process, which demonstrates exquisite levels of control over silica morphogenesis.

KEYWORDS: Biomimetic synthesis, biosilification, biochemistry, nanopatterning, silicic acid polymerization, phase separation



Silica, existing in various allotropic forms of silicon dioxide, not only makes up ~12% of the earth's crust but can also be found in human bodies and plant resources as well as in unicellular marine algae. With its material abundance and diverse application areas, various types of silica have been produced synthetically.^{1,2} However, traditional methods are usually performed under very harsh conditions of high temperature and extreme pH and also lack control over the morphology and structure of the products.^{2–4}

In comparison, biological organisms can produce the complex and elaborated silica nanostructures with a wide diversity in form under mild conditions with lower energy consumption.^{5–9} Outstanding examples for such biosilica are the silicified cell walls of diatoms, single-celled marine algae.^{10–12} Motivated by a characteristic feature of biogenetic production of silica structures, considerable efforts have been directed toward mimicking biosilification. The identification of organic biomolecules such as polyamine and polypeptides as major biosilica constituents and their specific interaction with silicic acid provided insight into *in vitro* synthesis of silica with its various forms including spheres,^{5,6} and nanotubes.⁸ In addition, a model based on the phase separation events, which are associated with the amphiphilic properties of polycationic peptides, has been proposed to explain biosilica morphogenesis in diatoms.^{13–18} However, nanopatterning of silica with high complexity to produce exquisite structures comparable to those of diatom cell walls is still in its infancy.^{19,20} This is primarily because of our limited understanding of the mechanism underlying genetic control of this biosilification process and the difficulty in creating laboratory conditions identical to those

of the *in vivo* biosilification event; silica deposition in diatoms takes place inside specialized intracellular compartments, silica deposition vesicles (SDVs), in which selective release of silicic acid and its polymerization are biologically controlled.^{21–24}

Inspired by the morphogenesis in biosilica, we demonstrate the spontaneous transformation from self-assembled silica spheres to highly intricate and symmetric networked structures. This method is different from the previous biomimetic approaches that generally utilize a homogeneous mixture of silicic acid derivatives and biomolecular additives or their synthetic analogues.^{5–8,19,20,25} Instead of using these macromolecular constituents, we introduce very common additives of hydrofluoric acid (HF) and ammonium fluoride (NH₄F) to yield the relevant constituents via chemical reaction with the silica spheres. Moreover, precipitation of silica at specific sites is induced within the scaffolds of the silica spheres, which thereby enables the synthesis of tailored, well-organized silica networks. One simple example is sphere-bridge network (SB-NW) structures, in which silica spheres are interlinked with the neighboring spheres via nanobridges (Figure 1). The key steps to fabricate SB-NWs include (i) self-assembly of a hexagonally closed-packed monolayer of silica spheres on a PMMA layer, (ii) anchoring the silica spheres to the PMMA layer, and (iii) immersing the sample in a near neutral solution (pH 6.7–6.8) of a 1:30 (v/v) mixture of 49% HF and 40% NH₄F (see methods in the Supporting Information).

Received: April 26, 2012

Revised: June 11, 2012

Published: June 27, 2012



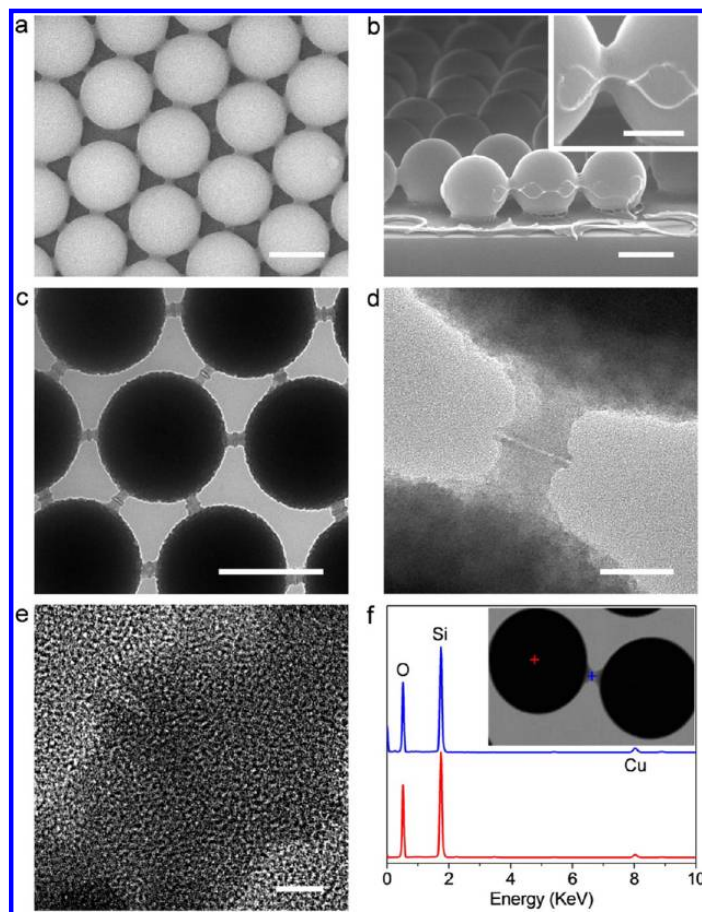


Figure 1. Sphere-bridge network (SB-NW) structures. (a) Top-view and (b) cross-sectional SEM images of SB-NW structures. Scale bars: 500 nm. Inset of (b): high-magnification image showing the bridge and the traces of severed bridges on the sphere surfaces. Scale bar: 200 nm. (c) TEM image of SB-NWs. For measurement, sample was prepared directly on the TEM grid. Scale bar: 500 nm. (d, e) Higher-magnification TEM images highlighting the amorphous bridge consisting of a pair of cylinders touching each other. Scale bars: 50 nm (d), 10 nm (e). (f) EDS spectra measured at the centers of the silica sphere (red) and bridge (blue); the measurement positions are marked by crosses in the inset of the dark-field STEM image.

Figure 1a,b shows the scanning electron microscopy (SEM) images of the representative SB-NW structures. Cylindrical bridges were formed at the nearest points between the silica spheres. The process of SB-NW formation is different from previous cases that relied on high-temperature sintering to produce overlapping spheres with wide necks and subsequent isotropic etching to form the adjoining bridges.²⁶ To make this point clear, we performed a time-resolved analysis on the morphology development (Figure S1). Initially, closed-packed silica spheres were isotropically etched, and almost simultaneously, its reverse reaction, i.e., nanobridge formation, occurred. Since silica growth prevails over the etching process, the diameter of the bridge increased steadily to ~ 150 – 200 nm for the first 10 s; however, with the further elapse of time, the bridges began to be etched from the center. After that, the silica bridges became elongated and narrower at 15–18 s and finally disappeared after 20 s (Figure S2). In addition, the reaction rate was further decreased by reducing the concentration of HF, with which we could monitor the early stage dynamics of SB-NW formation. A series of SEM images in Figure S3 illustrate

the isotropically etched silica spheres at the outset of the process and the subsequent development of silica bumps, demonstrating that the bridges grow out from the silica surface.

Further structural characterization of the SB-NWs was carried out using transmission electron microscopy (TEM) for the sample prepared directly on the lacey carbon-coated copper grid. Amorphous bridges with a mean diameter less than ~ 60 nm protrude out from both spheres (Figure 1c,d). Moreover, higher magnification image obtained by zooming in one of the nanobridges reveals that the spheres are completely interlinked by the bridges without producing any gap (Figure 1e). Energy dispersive X-ray spectrometry (EDS) spectra measured at the centers of the sphere and bridge confirmed the existence of Si and O elements (Figure 1f). Further, the atomic ratio of the Si:O in the bridge was estimated to be 1:1.61, which is quite similar to that in the sphere (1:1.66) and close to the stoichiometric composition of SiO_2 .

On the basis of our experimental observations, we propose a mechanism for the formation of the SB-NW structures that is analogous to biosilica morphogenesis. Studies on biosilification

Neurite Outgrowth



In Vitro Developmental Acceleration of Hippocampal Neurons on Nanostructures of Self-Assembled Silica Beads in Filopodium-Size Ranges**

Kyungtae Kang, Sung-Eun Choi, Hee Su Jang, Woo Kyung Cho, Yoonkey Nam,*
Insung S. Choi,* and Jin Seok Lee*

The neurite outgrowth and path-finding behaviors of neurons are governed by two protrusive, actin-based molecular structures, filopodia and lamellipodia, the diameter of which is generally in the range of 100–300 nm.^[1] The dynamic stability of filopodia and lamellipodia allows neuronal cells to recognize the surrounding environments at the nanometer scale and to subsequently modify their cytoskeletal structures in response to stimuli/cues.^[2] For example, permissive cues impede the retrograde flow of actin filaments and promote their assembly by generating tensions to the filopodia,^[3] resulting in the overall advance of neurites toward the cues.

Neurons encounter nanotopographical distributions of permissive and non-permissive cues in vivo, exemplified by protein structures in the extracellular matrix and tissue scaffolds or external morphologies of supporting cells. Although the surface roughness is known to affect the neuronal behaviors in vitro in various aspects, such as attachment and survival,^[4] axonal guidance,^[5] and neurite outgrowth,^[6] there have been few reports on developmental responses of neurons to nanotopographies.^[7,8] Because the polarized morphology and specialized compartments (e.g., axons and dendrites) of neurons are crucial and decisive

characteristics for neuronal activities, it is much more beneficial to the precise control of neuron/materials interfaces to manipulate neuronal development and neurite outgrowth than to do general behaviors, such as attachment and survival. The concrete understanding of nanotopographical effects also would contribute to the biomedical applications, which demand the controlled development or regeneration of nerve systems at a certain position towards a pre-determined direction, as well as to fundamental studies on nanotopographical role for neuronal development in vivo.^[9]

To understand the function of filopodia as an antenna for the environmental exploration,^[2] it is necessary to investigate the developmental responses of neurons to nanostructures, the feature size of which is comparable to that of filopodia (100–300 nm). Herein, we systematically varied the feature sizes of nanostructures by organizing spherical nanoparticles with different diameters on a glass substrate, and found that the neuritogenic acceleration of hippocampal neurons occurred on the nanostructures the period of which was larger than 200 nm.

For the generation of nanosurfaces with various feature sizes, silica beads with the diameters ranging from 100 to 700 nm were synthesized by hydrolysis of tetraethyl orthosilicate.^[10] The diameter was controlled to be 110, 190, 320, 480, or 670 nm by changing the re-growth cycles of seeds and the concentrations of reactants (See the Supporting Information for the experimental details; Figure S1). The silica beads were denoted as SB-110, SB-190, SB-320, SB-480, and SB-670, respectively. Among the methods for organizing nano-sized beads into two-dimensional arrays on a solid substrate,^[11] rubbing was reported to be a highly effective in the monolayer assembly with the precise control of the interparticle distance, and simple, fast, and reproducible on a large area (Figure 1a).^[12] Figure 1b shows the SEM images of the bead monolayers with the different diameters, formed by the rubbing method. This simple method yielded high-quality monolayers with uniform interparticle distance regardless of the bead diameters. To prevent the undesired detachment of silica beads from the surface during cell culture, we baked the substrate at 350 °C for 3 h (see the Supporting Information, Figure S2, for the experimental details). Figure 1c shows a representative SEM image of a hippocampal neuron cultured on a monolayer of SB-110, where the beads are smaller than the filopodia.

The surface of silica beads was made to be neuron-adhesive by 2 min-O₂-plasma treatment, followed by 1 h dipping in aqueous poly-D-lysine solution (0.1 mg mL⁻¹ in

[*] K. Kang,^[†] Dr. W. K. Cho, Prof. Dr. I. S. Choi
Molecular-Level Interface Research Center, Department of Chemistry, KAIST
Daejeon 305-701 (Korea)
E-mail: ischoi@kaist.ac.kr
Homepage: <http://cisgroup.kaist.ac.kr>
Prof. Dr. Y. Nam, Prof. Dr. I. S. Choi
Department of Bio and Brain Engineering, KAIST
Daejeon 305-701 (Korea)
E-mail: ynam@kaist.ac.kr
Homepage: <http://neuros.kaist.ac.kr>
S.-E. Choi,^[†] H. S. Jang, Prof. Dr. J. S. Lee
Department of Chemistry, Sookmyung Women's University
Seoul 140-742 (Korea)
E-mail: jinslee@sookmyung.ac.kr
Homepage: <http://sookmyung.ac.kr/~FETNS>

[†] These authors contributed equally to this work.

[**] This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, KRF-2008-313-d00614) and the Basic Science Research Programs through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2011-0001318, 2009-0077751, 2009-0080081, and 2010-0025065) and the Brain Research Center of the 21st Century Frontier Research Program.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201106271>.

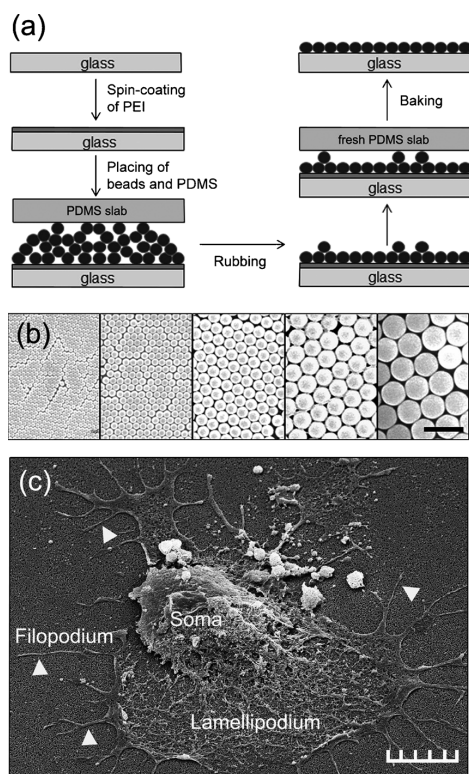


Figure 1. a) A schematic illustration of the formation of two-dimensional nanostructures of silica beads on a glass substrate. b) SEM images of the nanostructures assembled with SB-110, SB-190, SB-320, SB-480, and SB-670 (from left to right). The scale bar is 1 μm . c) SEM image of a neuron at 1 DIV cultured on SB-110. The scale bar is 5 μm . The arrowheads indicate filopodia.

deionized water). Hippocampal neurons were then plated at the density of 50 cells/ mm^2 . Figure 2a shows phase-contrast optical images of the neurons at 1 DIV (day in vitro) and 2 DIV. Based on neuronal morphologies, the substrates were categorized into two groups: Group-I (SB-110 and SB-190) and Group-II (SB-320, SB-480, and SB-670). We observed noticeable differences in neuronal development between the two groups even at 1 DIV. In Group-I, most of neurons had either lamellipodia or short minor neurites at 1 DIV, and the ratio between neurons with and without neurites was about 4:6. This ratio was similar to neurons grown on a polylysine-treated glass coverslip at 1 DIV.^[13] In a stark contrast, more than 80% of Group-II neurons had neurites at 1 DIV, and the length of the neurites was significantly longer than that of Group-I neurons. In terms of axonal development, more neurons in Group-II had a major neurite ("putative" axon) than those in Group-I, which implied that axonal polarization processes were more active on Group-II substrates than Group-I substrates. At 2 DIV, Group-II neurons had one long branching axon and a few minor neurites, which was a phenotype of cultured hippocampal neurons.^[13] The immu-

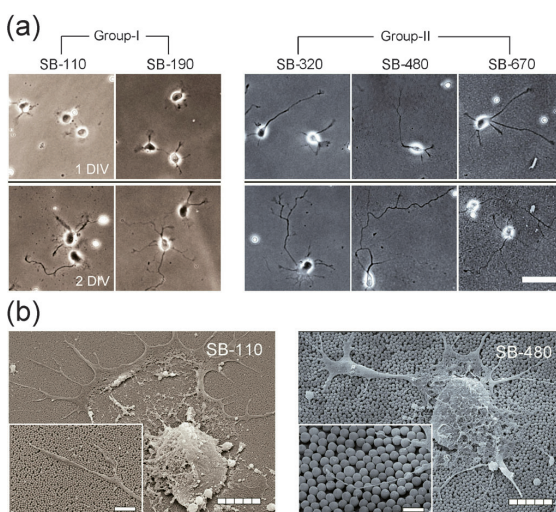


Figure 2. a) Phase-contrast optical micrographs of hippocampal neurons cultured on bead-packed substrates. The scale bar is 50 μm . b) SEM images of the hippocampal neurons on SB-110 (left) and SB-480 (right). The scale bar is 5 μm . Insets: Magnified images of filopodial tips. The scale bar is 1 μm . The original gray images were rendered with a monochromatic color for the comparison between the groups.

nostaining showed that the microtubules and filopodial/lamellipodial structures were well-developed on all the bead substrates (See the Supporting Information; Figure S3). These results were in a good agreement with our previous report: the developmental acceleration occurred on a 400 nm-pitched anodized aluminum oxide substrate.^[8]

Scanning electron microscopy (SEM) was used to examine the physical interactions of neurons with silica beads. The bead size seemed to affect the filopodial contact with the substrate in a different fashion: filopodia of the neurons in Group-I extended straight regardless of the bead curvature, while filopodia in Group II did not grow straight and seemed to have many interactions with surface structures such as trenches, bead surfaces, and gaps between the beads (Figure 2b, insets). The SEM micrographs showed that different bead sizes caused the different spatial distribution of interaction spots between the filopodia and the substrate. These interactions were believed to be phenomenologically different from the ones with microstructures. Microtopographies, such as microchannels and micropillar array with feature sizes larger than 1 μm ,^[4–6] were reported to have influences on neuronal growth directly at cellular scale; entire neurites were trapped within the microstructures, which physically restricted the direction of neurite growth or generated cellular-scale tensions.^[5,6] In contrast, nanotopographies would have localized effects on subcellular structures such as filopodia, amplified to the cellular-scale responses such as developmental acceleration. Our results supports the idea that nanotopographical features trigger intracellular signaling pathways through mechanotransductions of cytoskeletal structures, and the discontinuous or jumping stimuli of

Effect of Mode of Binding Linkage on Monolayer Assembly of Zeolite

Jin Seok Lee

*Department of Chemistry, Sookmyung Women's University, Seoul 140-742, Korea. *E-mail: jinslee@sookmyung.ac.kr
Received November 1, 2011, Accepted November 22, 2011*

During the monolayer assembly of zeolite microcrystals using sonication with stacking (SS) method, the factors that govern the degree of close packing (DCP) between the microcrystals, the rate of attachment (r_A) of microcrystals onto the substrate, the degree of coverage (DOC), and the binding strength (BS) between each crystal and the substrate were investigated for each mode of binding linkage (MBL). The tested MBLs were covalent linkage (CL), ionic linkage (IL), and polyelectrolyte-mediated ionic linkage (p-IL). Unlike the monolayers of zeolite crystals assembled on glass with a covalent linkage, the strong BS, very high DOC, and very high DCP do not decrease during monolayer assembly on glass through ionic linkages. This results indicate that the surface migration of crystals undergo linkage-nondestructively when crystals were attached to the substrates through ionic linkages.

Key Words : Zeolite, Monolayer assembly, Sonication, Binding linkage, Surface migration

Introduction

The bottom-up approach for fabricating nano-structured materials, which could lead to highly integrated functional entities, has played an important roles in utilizing building blocks to advanced materials.^{1,2} Using several building blocks with different shape and size, we can design and fabricate multi-functional heterogeneous system on various substrates, and this has been the focus of interests during the last several decades.¹⁻⁹ The organization of building blocks have been intensively explored, and a number of assembly methods have been demonstrated to achieve ordered arrays on various substrates.

From the efforts to organize micro building blocks in the form of closely packed monolayers on various substrates, a number of attempts have been reported to improve the degree of close packing (DCP) between the microcrystals, the rate of attachment (r_A) of microcrystals onto the substrate, the degree of coverage (DOC), and the binding strength (BS) between each crystal and the substrate using large numbers of interconnecting molecular linkages.¹⁰⁻¹⁹ Recently, we discovered that sonication-induced strong agitation of the microcrystals and the interposition of a functional-group-tethering glass plate between two bare glass plates leads to very high DOC and DCP of the microcrystals on substrates within very short periods of time (< 2 min) during the monolayer assembly of microcrystals on substrates through covalent linkages.^{20,21} This method was named as 'sonication with stacking' (SS). Furthermore, under the SS mode, we found an interesting phenomenon that BS, DOC, and DCP gradually decrease with increasing the reaction time, despite of DOC reaches $\sim 100\%$ with a quite fast r_A . Conclusively, we ascribed the progressive decrease in DOC to the progressive weakening of BS between the crystals and glass substrate under the sonication conditions as a result of the diminution of the numbers of func-

tional groups on the surfaces of the crystals and glass substrates upon repeatedly undergoing of the cycle of bond-formation and bond-breakage during the courses of surface migration and replacement of the attached crystals with those in the solution.^{20,21}

As a means to settle the above critical limitation for applications, we try to change covalent linkages with ionic linkages during the monolayer assembly of microcrystals on glass using SS mode.²² For this purpose, we adopted the previously established methods of forming ionic linkages between the surface-bound full-fledged ionic centers or ionic linkages mediated by alternating layer of oppositely charged polyelectrolytes.¹⁴ Here, we report that indeed, the mode of binding linkage (MBL) sensitively affects to r_A , rate of the surface migration (r_{SM}), DCP, DOC-t profile, and BS-t profile. Furthermore, we have discovered that DOC and BS are governed with not only the mode of reaction promotion (MRP) but also the MBL undergoing of the cycle of bond-formation and bond-breakage during the courses of surface migration and replacement of the attached crystals with those in the solution.

Experimental Section

Materials. 3-Aminopropyltrimethoxysilane (AP-TMS, Aldrich) was distilled and kept in a Schlenk storage flask under argon. 3-Chloropropyltrimethoxysilane (CP-TMS, Aldrich) and 3-Cyanopropyltrichlorosilane (CNP-TCS, Merck) were used as received. For simplicity, 3-aminopropyl, 3-Chloropropyl, and 3-Cyanopropyl are denoted as AP, CP, and CNP, respectively. Poly(sodium 4-styrenesulfonate) (Na^+PSS^- , M_w 70,000, Aldrich) and poly(diallyldimethylammonium chloride) (PDDA^+Cl^- , 20 wt % in water, $\sim M_w$ 100,000-200,000, Aldrich) was used as received. Cover glasses ($18 \times 18 \text{ mm}^2$) were purchased from Marienfeld and treated in a piranha solution (H_2SO_4 : 30% H_2O_2 = 3:1) at 95-

100 °C for 1 h to remove organic residues on the surface. The acid-treated glasses were washed with distilled deionized water and dried at 120 °C for 30 min prior to treatment with AP-TMS or CP-TMS. Optical grade fused silica plates were purchased from CVI Laser Co. and the surfaces of them were similarly cleaned with piranha solution prior to treatment with CNP-TCS. All the organic solvents were purified according to the well-known literature procedures prior to use.

Mode of Binding Linkage. The types of linkage used in this report were (a) the covalent linkage achieved by reaction between bare silicalite crystals (Z) and chloropropyl-tethering glass (CP-G) plates, (b) the direct ionic linkage achieved by reaction between sodium butyrate-tethering silicalite (Z) crystals and trimethylpropylammonium iodide-tethering glass (G^+) plates, and (c) poly(sodium 4-styrenesulfonate) (Na^+PSS^-) and poly(diallyldimethylammonium chloride) ($PDDA^+Cl^-$)-mediated ionic bonding between Z and G^+ , respectively. Preparation of CP-G, Z, and G^+ was carried out according to the procedures described in previous report.¹⁴ The size of each glass plate was $1.8 \times 1.8 \text{ cm}^2$. The average size of Z was $1.3 \times 0.6 \times 1.8 \mu\text{m}^3$.

Assembly of Z^- on G^+ (or $PDDA^+/PSS^-/G^+$) using the SS Mode. The MRPs tested in this report were sonication with stacking between the glass plates (SS mode).^{20,21} Comb-shaped Teflon supports were used as such without attaching underlying supporting legs so that the glass plates inserted into the gaps can have direct contacts with the bottom of the round-bottomed flasks (50 mL). Typically, a comb-shaped Teflon support was placed at the bottom of a round-bottomed flask charged with 40 mL of toluene and 50 mg of Z^- . For the test of the effect of SS mode, a G^+ (or $PDDA^+/PSS^-/G^+$) plate was interposed between two clean bare glass (BG) plates, and three sets of BG/ G^+ (or $PDDA^+/PSS^-/G^+$)/BG stack were inserted, stack by stack, into each gap of the comb-shaped Teflon support. Sonication of the glass reactors was carried out using an ultrasonic bath with the dimension of $22 \times 12 \times 11 \text{ cm}^3$ equipped with two ultrasound generators (28 kHz, 95 W each) attached to the bath from the bottom. After reaction, the Z^- -attached G^+ (or $PDDA^+/PSS^-/G^+$) plates were mildly sonicated for 30 s in toluene to remove the microcrystals that were physically or very weakly adhered to the glass plates and to the silicalite monolayers that were chemically attached to the glass plates using the ultrasonic bath equipped with two 50-W ultrasound generators before conducting the measurements of the attached amounts (W_A) onto G^+ (or $PDDA^+/PSS^-/G^+$) plates on a microbalance, the analyses of the glass plates with scanning electron microscopy (SEM), and the sonication-induced silicalite detachment test. DOC was deduced from each W_A based on eq. (1),

$$\text{DOC} = (W_A/W_{A,\text{max}}) \times 100 \quad (1)$$

where, $W_{A,\text{max}}$ represents the maximum obtainable W_A on a glass plate used in our study (on both sides of $1.8 \times 1.8 \text{ cm}^2$). As for the $W_{A,\text{max}}$, the average W_A obtained from Z^-/G^+ (or $Z^-/PDDA^+/PSS^-/G^+$) plates prepared by carrying out the

attachment reaction under the condition of SS was used since the silicalite monolayers obtained under the above condition were nearly perfect in many respects such as DCP, coverage, negligible amounts of microcrystals on the monolayers.

Comparison of the Linkage-Dependent BS between Silicalite Crystals and Glass Plates. The linkage-dependent BS between silicalite crystals (Z or Z^-) and glass plates (CP-G, G^+ , or $PDDA^+/PSS^-/G^+$) were compared by monitoring the progressive DOC-t profile, and BS-t profile. The details about the procedure can be found from our previous report.^{20,21}

Sonication-Induced Detachment Test. Into an independent ultrasonic cleaning bath equipped with two 100 W ultrasound generators was added 2.5 L of water. Into a vial (i.d. = 2.4 cm, length = 5.4 cm) were added a Z-G ($Z^-/CP-G$, Z^-/G^+ or $Z^-/PDDA^+/PSS^-/G^+$) plate and 10 mL of fresh toluene. With the help of a clamp and a clamp holder placed outside the bath, the vial was placed in the center of the water pool, 9 cm above from the bottom of the bath. The water inside the bath was maintained at $\sim 20^\circ\text{C}$ using an immersion cooler. After 30 min of sonication, the substrate was removed from the vial, dried in an oven (120°C) for 3 min, and its weight was measured on a microbalance. Fresh toluene was charged into the vial after each weighing. The total weight of pure silicalite crystals bound to the substrate was deduced from the initial weight of the Z-G plate that was measured after removal of physisorbed crystals by mild sonication for 30 s. The W_{AS} from three Z-G plates were averaged for each data point.

Instrumentation. The SEM images of silicalite crystals and the Z-G plates were obtained from a FE-SEM (Hitachi S-4300) at an acceleration voltage of 10 to 20 kV. The SS method were tested using an ultrasonic cleaning bath equipped with two 95-W ultrasound generators operated at 28 kHz (Saehan Ultrasonic SH-2100). The weights of the glass plates were measured on a Mettler microbalance (MT5).

Results and Discussions

Basic Concept for the Linkage-Dependent Promotion Test. In the previously study,^{20,21} we have investigated the effects of MRP on the progress of the monolayer assembly of zeolite microcrystals on glass through large numbers of molecular covalent linkages between zeolite microcrystals and glass plates and the characteristic properties of the assembled monolayers. Again, by charged amount (A_C) of zeolite microcrystals into account, we have discussed how DOC and BS are governed with the MRP undergoing of the cycle of bond-formation and bond-breakage during the courses of surface migration and replacement of the attached crystals with those in the solution.

However, we now have fixed several factor, such as MRP (SS mode), charged amount (A_C) of zeolite microcrystals ($A_C = 50 \text{ mg}$) and reaction temperature (20°C), for the investigation of the only linkage-dependent attachment and detachment profile. As illustrated in Figure 1, the tested

의약과학과

Retroviral integration profiles: their determinants and implications for gene therapy

Kwang-il Lim*

Department of Medical and Pharmaceutical Sciences, College of Science, Sookmyung Women's University, Seoul 140-742, Korea

Retroviruses have often been used for gene therapy because of their capacity for the long-term expression of transgenes via stable integration into the host genome. However, retroviral integration can also result in the transformation of normal cells into cancer cells, as demonstrated by the incidence of leukemia in a recent retroviral gene therapy trial in Europe. This unfortunate outcome has led to the rapid initiation of studies examining various biological and pathological aspects of retroviral integration. This review summarizes recent findings from these studies, including the global integration patterns of various types of retroviruses, viral and cellular determinants of integration, implications of integration for gene therapy and retrovirus-mediated infectious diseases, and strategies to shift integration to safe host genomic loci. A more comprehensive and mechanistic understanding of retroviral integration processes will eventually make it possible to generate safer retroviral vector platforms in the near future. [BMB reports 2012; 45(4): 207-212]

INTRODUCTION

Retroviruses are RNA viruses that carry a diploid positive-sense RNA genome (1, 2). These viruses are classified into five retroviral genera (the alpha- to epsilon-retroviruses) and two additional genera, the spumaviruses and lentiviruses. Retroviruses cause various fatal human diseases, including acquired immunodeficiency syndrome (AIDS), inflammatory diseases and multiple types of cancer (3-6). Ironically, because of the unique ability of retroviruses to efficiently integrate genetic material within their genome into host chromosomes, these viruses have been used to reverse genetic defects in patients (2). This beneficial use of retroviruses was made possible by an early understanding of retroviral biology and the development of powerful molecular biology techniques (7, 8). After removing genes that are needed for viral replication (for safety reasons), the genomes

of retroviruses were further engineered to carry therapeutic transgenes between two long terminal repeats (LTRs) (7-9). These engineered, replication-defective forms of retroviruses, which have been designated retroviral vectors, have been mainly used in making genetic modifications to stem cells for cell-based therapy and in the gene therapy-mediated treatment of various human diseases, including blood disorders, diabetes, and neurological disorders (2, 10-15). Unfortunately, there have been unexpected side effects from retroviral gene therapy that largely stem from where the retroviral genome integrates into the host genome. When retroviral integration occurs near oncogenes, normal cells can be transformed into cancerous cells, as demonstrated by the incidence of leukemia in a gene therapy trial (11). This minireview summarizes the recent findings regarding the integration patterns of various types of retroviruses, the impact of these integration patterns on the initiation and progression of retrovirus-mediated infectious diseases and sporadic virus-initiated cancer, and the viral and host factors that determine these integration patterns. In addition, the ways in which both viral and host components have been engineered to develop safer retroviral vector platforms are briefly discussed.

RETROVIRAL INTEGRATION PREFERENCES FOR GENOMIC REGIONS RICH IN GENE REGULATORY ELEMENTS

Whereas retroviruses were previously thought to randomly integrate into the host genome, many of them display a non-random integration pattern. For example, murine leukemia virus (MLV), a gamma-retrovirus that has been well studied and is often used in gene therapy, shows a strong preference for integrating into genomic regions within or near transcriptional start sites (TSSs) and CpG islands (16-19), which are enriched in gene regulatory elements. MLV demonstrates equivalent levels of integration upstream and downstream of TSSs (16, 18). This integration pattern has been suggested to be closely linked to the incidence of leukemia in young patients who were treated with retrovirally modified CD34+ cells (11). The underlying mechanism by which cells are transformed during retroviral gene therapy is related to the ability of the promoter and enhancer components in the LTR and/or in the regions upstream of the therapeutic genes within the retroviral genome to activate host oncogenes, such as LMO2, when the viruses integrate near the TSSs of these genes (11). Activated oncogenes often turn normal cells into can-

*Corresponding author. Tel: +82-2-2077-7627; Fax: +82-303-0799-0396; E-mail: klim@sm.ac.kr
<http://dx.doi.org/10.5483/BMBRep.2012.45.4.207>

Received 6 February 2011

Keyword: Host genome, Integration patterns, Retroviral vectors, Safety, Tumorigenicity

cerous cells by perturbing the regulation of cell growth and survival. The retroviral promoter and enhancer components have recently been shown to affect cellular transcription over a long range, independent of vector type, thereby increasing the tumorigenicity of the viral vector (20). The retroviral propensity for frequently targeting oncogenes and cell growth/survival genes and the integration preference for TSSs and CpG islands synergistically increase the malignant potential of retroviral vectors (21).

In addition, MLV considerably prefers transcription factor binding sites (TFBSs) (22) and moderately genes, especially highly expressed genes, during integration (16, 21). Another gamma-retrovirus, porcine endogenous retrovirus (PERV), shows a similar integration pattern, favoring TSSs and CpG islands (23). Xenotropic murine leukemia virus-related virus (XMRV) has an even greater integration preference for TSSs and CpG islands than MLV (24). Relative to other genomic regions, the loci near gene regulatory elements are more deleterious sites of retroviral integration.

LENTIVIRAL INTEGRATION INTO ACTIVELY TRANSCRIBED GENES

HIV-1, a representative lentivirus, strongly prefers integrating into genes instead of genomic regions near TSSs and CpG islands (16-18, 25). In particular, HIV-1 targets actively transcribed genes, and this strong preference for genes has been consistently observed in different cell lines and multiple types of primary cells from different species (26). Furthermore, in both dividing and quiescent cells, HIV-1 demonstrates a strong preference for integration into highly transcribed genes. These findings suggest that the features of cellular physiology do not always affect the overall retroviral integration patterns (27). Whereas the extent of HIV-1's preference for genic regions positively depends upon the transcription level of the genes, interestingly, HIV-1 shows a reduced preference for the most highly expressed genes (21, 26). In contrast to HIV-1, the integration patterns of gamma-retroviruses, such as MLV and PERV, do not depend upon the expression level of the genes (23). HIV-1 normally integrates into entire transcriptional units (TUs) without a significant bias for a certain region of a gene. In addition, HIV-1 targets oncogenes and the genes involved in cell growth and survival (21). In contrast, HIV-1 disfavors CpG islands and does not show any preference for TFBSs (16, 22, 26). Interestingly, integrated lentiviral vectors can both up- and down-regulate host gene expression, in contrast to MLV (20). Another lentivirus, simian immunodeficiency virus (SIV), also shows a similar integration pattern of favoring actively transcribed genes with no preference for TSSs and CpG islands (19, 28, 29). Therefore, retroviruses and lentiviruses in the same genus often share similar integration preferences, as illustrated by MLV and PERV and HIV-1 and SIV (23).

The first molecular-level impact of viral integration on the host is the increased expression of genes that harbor retroviral integrants or are adjacent to retroviral integrants; this increased expression can be up to 3-fold compared with the control (26).

Such effects are particularly apparent in HIV-1-infected cells. The site at which retroviruses and lentiviruses integrate into the host genome can affect the initiation and progression of disease. A study using *Cdkn2a*^{-/-} mice, which develop tumors from cancer-triggering genetic changes because of the lack of the Rb1- and p53-dependent tumor-suppressor pathways, demonstrated that retroviral infection of the mice produced tumors following integration at oncogenes and cell cycle-controlling genes in a virus dose-dependent manner (30). In contrast, lentiviral infections did not significantly affect the extent of tumor formation even at high virus doses, indicating that lentiviral vectors may be less tumorigenic than other retroviral vectors (30).

SEMI-RANDOM INTEGRATION OF RETROVIRUSES

In contrast to MLV, PERV, HIV-1 and SIV, which have strong integration preferences for particular genomic regions, avian sarcoma leucosis virus (ASLV), an alpha-retrovirus, has an integration pattern that cannot be clearly distinguished from a hypothetical random pattern (26). ASLV has only a weak preference for actively expressed genes and no significant preference for TSSs. A spumavirus, foamy virus (FV), does not show strong preference toward either TSSs or active genes (31). Similarly, a delta-retrovirus, human T-cell leukemia virus type I (HTLV-I), does not have a strong preference for any host genomic features except aliphoid sequences and demonstrates an almost random integration pattern (3). The sequence similarity between the ASLV and HTLV-1 integrases, both of which have seemingly random-like integration patterns, suggests that the viral integrase plays an important role in the selection of sites during integration (32). Retroviruses with random integration patterns can be considered safer platforms for gene delivery vehicles compared with other retroviruses that demonstrate strong integration preferences for TSSs or transcribed genes. Random integration will have a lower chance of disturbing genes that regulate cell growth and survival than integrations targeted to genomic regions enriched in genes or gene regulatory elements.

Whereas integrated retroviral genomes affect the expression of host genes that harbor or are adjacent to the integrants, the host genomic environment surrounding the viral genomes also affects the expression of viral proteins and genomic RNA, ultimately regulating the production of viral progeny. Such mutual genetic effects between proviruses and host cells can lead to complications in the initiation and progression of infectious diseases and sporadic cancer, both of which are caused by retroviral infections. This scenario has been experimentally observed *in vivo* even in infections with retroviruses that display seemingly random integration patterns. For example, HTLV-I infection often results in an asymptomatic carrier disease state that can progress to adult T cell leukemia (ATL) (3, 33). In the carrier state, many infected cells harbor viral genomes at random loci without a strong bias for TSSs and TUs, but only cells in which the viral genome is integrated into the regions not actively transcribed, such as aliphoid sequences, can escape from the im-

Genome Sequence of *Vibrio* sp. Strain EJY3, an Agarolytic Marine Bacterium Metabolizing 3,6-Anhydro-L-Galactose as a Sole Carbon Source

Hanseong Roh,^a Eun Ju Yun,^b Saeyoung Lee,^a Hyeok-Jin Ko,^a Sujin Kim,^a Byung-Yong Kim,^c Heesang Song,^d Kwang-il Lim,^e Kyoung Heon Kim,^b and In-Geol Choi^a

Computational and Synthetic Biology Laboratory, College of Life Sciences and Biotechnology, Korea University,^a and School of Life Sciences and Biotechnology, Korea University,^b Seoul, South Korea; Department of Agricultural Microbiology, National Academy of Agricultural Science, Rural Development Administration (RDA), Suwon, South Korea^c; Department of Biochemistry and Molecular Biology, Chosun University School of Medicine, Gwangju, South Korea^d; and Department of Medical and Pharmaceutical Science, Sookmyung Women's University, Seoul, South Korea^e

The metabolic fate of 3,6-anhydro-L-galactose (L-AHG) is unknown in the global marine carbon cycle. *Vibrio* sp. strain EJY3 is an agarolytic marine bacterium that can utilize L-AHG as a sole carbon source. To elucidate the metabolic pathways of L-AHG, we have sequenced the complete genome of *Vibrio* sp. strain EJY3.

Microbial degradation of marine algal biomass contributes considerably to the global marine carbon cycle (1). The main cellular component of red algal biomass (e.g., *Rhodophyta*) is an agar polysaccharide that is composed of 3,6-anhydro-L-galactose (L-AHG) and D-galactose (11, 14). L-AHG is a rare anhydro-sugar whose metabolic fate and fermentability are completely unknown to date (4, 12, 16). We isolated a marine agarolytic bacterium using L-AHG as a sole carbon source from a decomposed grapsid crab (*Grapsidae*) collected from the coast of the West Sea (Incheon, South Korea). The phylogenetic and molecular analysis based on the 16S rRNA sequence indicated that the strain was closely related to *Vibrio* species, notably *Vibrio atypicus* HHS02^T (98.9%) (15), *Vibrio diabolus* HE800^T (98.9%), and *Vibrio mytili* LMG 19157^T (98.7%).

The genome sequence of *Vibrio* sp. strain EJY3 was assembled from two read libraries produced by the Roche 454 genome sequencer (GS-FLX) and the Solexa/Illumina genome analyzer (GA). The read library of GS-FLX is composed of 256,552 paired-end reads with 3-kb inserts (17-fold coverage), and the GA read library is comprised of 44,352,754 paired-end reads with about 350-bp inserts (821-fold coverage). Combining two read libraries with the Newbler gsAssembler 2.3 (454 Life Sciences, Branford, CT), we obtained 12 scaffolds, including 34 gaps. We arranged and verified the order of 12 scaffolds by comparing them to the other known genomes in the *Vibrio* genus. The 34 gaps in the scaffolds were filled by local reassembly using the Newbler gsAssembler and ABI 3730 sequencing of PCR products of the gap region. Gene finding and annotation were carried out by the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (<http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html>) and RAST server (2). Finally, the frameshift and homopolymer issues giving annotation discrepancies were manually inspected using a combination of BWA (6), GATK (4), IGV (10), SAMtools (7), and SSAHA2 (9) programs. The genome of *Vibrio* sp. strain EJY3 showed two circular chromosomes consisting of 3,478,307 bp and 1,974,339 bp with GC contents of 45.0% and 44.6%, respectively. The large chromosome (chromosome I) has 3,115 coding sequences (CDS), 110 tRNA genes, and 8 rRNA operons with one additional 5S rRNA gene. The small chromosome (chromosome II) has 1,731 CDS, 13 tRNA genes, and one rRNA operon.

Compared to known microbial agarolytic enzymes (3, 5), four β -agarase genes belonging to the glycoside hydrolase 50 (GH50) family were found in chromosome I of *Vibrio* sp. strain EJY3. The other typical β -agarases belonging to the GH16 and GH86 family were not identified in the genome. The agarases of AEX22317.1 (952 amino acids [aa]), AEX22316.1 (1,016 aa), and AEX22315.1 (971 aa) were clustered in the same locus of chromosome I. The agarase AEX22315.1 showed 97.5% sequence similarity to Aga41A of *Vibrio* sp. strain CN41 (8). By means of investigating the full genome, we anticipate unveiling the unknown metabolic pathway of L-AHG in *Vibrio* sp. strain EJY3.

Nucleotide sequence accession numbers. The complete genome sequence of *Vibrio* sp. strain EJY3 has been deposited in NCBI GenBank under accession numbers CP003241 and CP003242.

ACKNOWLEDGMENTS

This research was supported by the Intelligent Synthetic Biology Center (2011-0031953) and a National Research Foundation (NRF) grant (2011-0015629) funded by the Korean Government (MEST).

REFERENCES

1. Azam F, Malfatti F. 2007. Microbial structuring of marine ecosystems. *Nat. Rev. Microbiol.* 5:782–791.
2. Aziz RK, et al. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.
3. Cantarel BL, et al. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37: D233–D238.
4. Ekborg NA, et al. 2006. Genomic and proteomic analyses of the agarolytic system expressed by *Saccharophagus degradans* 2-40. *Appl. Environ. Microbiol.* 72:3396–3405.

Received 24 February 2012 Accepted 28 February 2012

Address correspondence to Kyoung Heon Kim, khekim@korea.ac.kr, or In-Geol Choi, igchoi@korea.ac.kr.

H.R. and E.J.Y. contributed equally to this work.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00303-12

5. Fu XT, Kim SM. 2010. Agarase: review of major sources, categories, purification method, enzyme characteristics and applications. *Mar. Drugs* 8:200–218.
6. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
7. Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
8. Liao L, et al. 2011. Cloning, expression, and characterization of a new β -agarase from *Vibrio* sp. strain CN41. *Appl. Environ. Microbiol.* 77: 7077–7079.
9. Ning Z. 2001. SSAHA: a fast search method for large DNA databases. *Genome Res.* 11:1725–1729.
10. Robinson JT, et al. 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29:24–26.
11. Sekkal M, et al. 1993. Direct structural identification of polysaccharides from red algae by FTIR microspectrometry I: localization of agar in *Gracilaria verrucosa* sections. *Mikrochim. Acta* 112:1–10.
12. Shin MH, et al. 2010. Global metabolite profiling of agarose degradation by *Saccharophagus degradans* 2-40. *New Biotechnol.* 27:156–168.
13. Reference deleted.
14. Stevenson TT, Furneaux RH. 1991. Chemical methods for the analysis of sulfated galactans from red algae. *Carbohydr. Res.* 210:277–298.
15. Wang Y, Zhang X-H, Yu M, Wang H, Austin B. 2010. *Vibrio atypicus* sp. nov., isolated from the digestive tract of the Chinese prawn (*Penaeus chinensis* O'sbeck). *Int. J. Syst. Evol. Microbiol.* 60:2517–2523.
16. Yun EJ, et al. 2011. Production of 3,6-anhydro-L-galactose from agarose by agarolytic enzymes of *Saccharophagus degradans* 2-40. *Process Biochem.* 46:88–93.

Retroviral Infection of hES Cells Produces Random-like Integration Patterns

Kwang-il Lim^{1,2,*}

Retroviral integration provides us with a powerful tool to realize prolonged gene expressions that are often critical to gene therapy. However, the perturbation of gene regulations in host cells by viral genome integration can lead to detrimental effects, yielding cancer. The oncogenic potential of retroviruses is linked to the preference of retroviruses to integrate into genomic regions that are enriched in gene regulatory elements. To better navigate the double-edged sword of retroviral integration we need to understand how retroviruses select their favored genomic loci during infections. In this study I showed that in addition to host proteins that tether retroviral pre-integration complexes to specific genomic regions, the epigenetic architecture of host genome might strongly affect retroviral integration patterns. Specifically, retroviruses showed their characteristic integration preference in differentiated somatic cells. In contrast, retroviral infections of hES cells, which are known to display decondensed chromatin, produced random-like integration patterns lacking of strong preference for regulatory-element-rich genomic regions. Better identification of the cellular and viral factors that determine retroviral integration patterns will facilitate the design of retroviral vectors for safer use in gene therapy.

INTRODUCTION

Retroviruses have been considered as efficient gene delivery vehicles based on their several advantageous characteristics. They can be engineered to carry large foreign genes (up to 8 kb) and induce only minor or no immune response. Most of all, retroviruses allow stable gene expression through integration of their genomes into host genomes. However, due to their oncogenic potential, as shown by the incidence of leukemia in a recent severe combined immune deficiency (SCID) clinical trial in Europe (Hacein-Bey-Abina et al., 2003), the therapeutic use of retroviruses has been strictly limited. The oncogenic propensity of retroviruses mainly stems from their tendency to selectively integrate into the transcriptional start sites (TSSs) of genes (Lewinski et al., 2006; Wu et al., 2003), including oncogenes, in host genomes. This integration preference can lead to the deregulation of cell growth through trans-activation of

oncogenes by retroviral promoters, ultimately transforming infected cells into cancer cells.

In parallel with attempts to shift such risky retroviral integration patterns to safer ones by molecular engineering (Lim et al., 2010), there have been significant efforts to elucidate how retroviruses, including lentiviruses, select integration spots in the host genome. It has been suggested that interactions among retroviral pre-integration complexes (PICs) and cellular proteins, such as transcription factors (TFs), determine viral integration sites via a mechanism in which cellular proteins tether PICs to certain genomic loci (Engelman and Cherepanov, 2008; Felice et al., 2009). However, only a fraction of host proteins associating with retroviral PICs have been identified, and other mechanisms that underlie such interactions are still poorly understood (Cattoglio et al., 2010). Moreover, cellular physiological states and cell-specific features, both of which determine the gene expression patterns of host cells as well as the epigenetic landscape of the host genome, may affect retroviral and lentiviral integrations, but the relevant studies have begun to be available only recently (Bartholomae et al., 2011; Biasco et al., 2011).

In this work I investigated how the overall chromatin architecture of host cells affected retroviral and lentiviral integration patterns by comparing retroviral and lentiviral infections of human embryonic stem (hES) cells with those of a differentiated somatic cell line. I found that the previously reported characteristic retroviral integration preference for gene regulatory regions enriched in TSSs and CpG islands was mostly abolished in hES cells, in which chromatin is known to be decondensed and highly plastic (Ang et al., 2011; Efroni et al., 2008). In addition, the inherent lentiviral integration preference for regions within genes was significantly reduced in hES cells compared with that in the somatic cell line, in which chromatin would be relatively much more condensed. I showed that direct interactions among viral PICs and cellular proteins were not the only factors determining integration preference; cellular physiological states and epigenetic features associated with cellular differentiation and development also affected viral genome integration patterns in host cells. Identification of the main determinants, both direct and indirect, of retroviral integration and a comprehensive understanding of the mechanisms underlying integration preference will aid in the effort to engineer retroviruses as therapeutic

¹Departments of Chemical Engineering and Bioengineering and The Helen Wills Neuroscience Institute, University of California, Berkeley, USA,

²Department of Medical and Pharmaceutical Sciences, College of Science, Sookmyung Women's University, Seoul 140-742, Korea

*Correspondence: klim@sm.ac.kr

Received February 6, 2012; revised March 11, 2012; accepted March 14, 2012; published online April 20, 2012

Keywords: gene therapy, human embryonic stem cells, oncogenic potential, overall chromatin structure, retroviral integration patterns

tics with less oncogenic potential.

MATERIALS AND METHODS

Cell culture

Human embryonic kidney (HEK) 293T cells were cultured in Iscove's modified Dulbecco's medium with 10% fetal bovine serum at 37°C and 5% CO₂. HSF6 human embryonic stem (hES) cells were kindly provided to the Schaffer lab by Susan Fisher at University of California at San Francisco. HSF6 cells were grown on a surface coated with hES cell-qualified Matrigel (BD Biosciences) in X-Vivo 10 Medium (Lonza) supplemented with 80 ng/ml FGF-2 (Peprotech) and 0.5 ng/ml TGF- β 1 (R&D Systems). The preparation and coating with Matrigel was performed following the manufacturer's instructions. This culture system has previously been shown to support extended self-renewal of hES cells (Li et al., 2005; Zhang et al., 2006).

Virus packaging and infections

The retroviral and lentiviral plasmids, pCLPIT GFP and pFUGW, were used for virus vector packaging (Lois et al., 2002). For the retroviral vectors, pCLPIT GFP, pCMV gag-pol and pcDNA3 IVS VSV-G plasmids were introduced into the packaging cell line (HEK 293T cells) via calcium phosphate transfection. For the lentiviral vectors, pFUGW, pMDLg/pRRE, pRSV-Rev and pcDNA3 IVS VSV-G plasmids (Dull et al., 1998; Lois et al., 2002) were similarly introduced into HEK 293T cells via calcium phosphate transfection. Virus supernatant was twice harvested at two and three days post transfection and concentrated by ultracentrifugation. The concentrated viruses were used to infect cells at multiplicity of infection less than 1.

Mapping of viral integration sites

The relevant molecular biological preparation to isolate virus and host genome junctions was performed by following and modifying an established method (Wu et al., 2003). The genomic DNA of infected 293T cells and hES cells was isolated five to seven days post infection using the DNA Mini Kit (Qiagen) and then digested with either *Mse*I or *Hpy*CH4III [New England Biolabs (NEB)]. Digested genomic DNA fragments were ligated to pre-annealed *Mse*I linker or *Hpy* linker DNA (*Mse*I linker+, 5'-GTAATACGACTCACTATAGGGCTCCGCTT AAGGGAC-3'; *Mse*I linker-, 5'-TAGTCCCTTAAGCGGAG-3'; *Hpy* linker+, 5'-GTAATACGACTCACTATAGGGCTCCGCTT AAGGGACN-3'; *Hpy* linker-, 5'-GTCCCTTAAGCGGAG-3'). The ligation products were used as templates for primary PCR with two primers annealing to the linkers and to either MLV LTR or HIV LTR (MLV-LTR, 5'-GTCTTGTGGTCTCGCTGTTTCCT TGG-3'; HIV-LTR, 5'-AGTGCTTCAAGTAGTGTGTGCCCG-3'; linker primer, 5'-GTAATACGACTCACTATAGGGC-3') under the following conditions: pre-incubation at 95°C for 2 min; 30 cycles each consisting of 30 s at 95°C, 30 s at 55°C and 2.5 min at 72°C; and a final extension step for 10 min at 72°C. The resulting primary PCR products were used for nested PCR with primers (MLV LTR nested, 5'-AAATTGGATCCGGTCTCCT CTGAGTGATTGACTACC-3'; HIV LTR nested, 5'-AAAAAG CATCCCGCTGTTGTGTGACTCTGGTAAC-3'; linker primer nested, 5'-AAATTAAGCTTAGGGCTCCGCTTAAGGGA C-3') under the same conditions as the primary PCR. The PCR-amplified virus-host genome junctions were cloned into the pBS SK SP plasmid after restriction enzyme digestion with *Bam*HI and *Hind*III (NEB), and then sequenced with a primer (5'-CGCGTTGGCCGATTCAATTAATG-3'). The host portions of the sequenced genome junctions were mapped to locations in the human genome, and the relevant genomic annotations

were made using the GTSG-Quickmap tool (Appelt et al., 2009).

Generation of random integrations

Hypothetical random integration positions in the human genome were computationally generated by GTSG-Quickmap tool, and the relevant genomic annotations were also made by the same tool (Appelt et al., 2009).

Statistical calculations

The P values for the frequency of integrations within specific genomic regions relative to the corresponding frequency of the hypothetical random integrations or to that of other viral infections were calculated by the chi-square test.

RESULTS

Retroviral vectors preferred to integrate into TSSs in the genomes of a somatic cell line

As observed in previous studies (Lewinski et al., 2006; Lim et al., 2010), retroviral vectors based on murine leukemia virus (MLV) strongly preferred TSSs and CpG islands for genome integration upon infection of somatic human embryonic kidney (HEK) 293T cells (see Table 1A for a comparison of this infection with a hypothetical random integration case; The numbers of analyzed integration events are mentioned in the table footnotes.). MLV vector integrations took place within 5 kb of TSSs 62.5% of the time and either inside or within 1kb of CpG islands 40.6% of the time (significance of 5.6×10^{-22} and 4.0×10^{-71} , respectively, compared with the random integration pattern). Such integration patterns can lead to perturbation of genetic circuits of cells, potentially causing dysregulation of cell growth, because the favored genomic regions are enriched in gene regulatory elements. However, MLV overall disfavored repeat regions as consistent with a recent study (Table 1A; Lim et al., 2010). The MLV vector only integrated into repeat regions 12.5% of the time, which was significantly lower than the 48.9% that was expected to occur by random chance. Long interspersed nuclear elements (LINEs) were especially not favored. Finally, the virus did not show any significant integration preference for the regions within genes (Table 1A), which is also consistent with the previous study (Lim et al., 2010).

The characteristic integration preference of retroviruses for TSSs and CpG islands was almost abolished in hES cells

In contrast with the above results in differentiated somatic cells, MLV vectors did not show a strong integration preference for any particular genomic region in hES cells (Table 1B). The viruses only displayed a moderate preference for the regions inside or within 5 kb of CpG islands (Table 1B). The observed level of integration preference for CpG islands, which was only 2.3-fold higher than what was expected by random chance, was dramatically less than that observed in infected 293T cells (9.1-fold higher than that expected by chance, Table 1A). In hES cells, the retrovirus was no longer biased against integration into genomic repeat regions at a significant level ($P < 0.05$) (Table 1B). These surprising shifts in preference for or against the particular genomic regions in hES cells suggest that development-specific characteristics of host cells, such as epigenetic landscape and chromatin organization, would affect retroviral integration patterns.

The integration preference of lentiviruses for the regions within genes was significantly reduced in hES cells

Lentiviral vectors that were based on human immunodeficiency

Resistive switching memory properties of layer-by-layer assembled enzyme multilayers

Hyunhee Baek^{1,4}, Chanwoo Lee^{2,4}, Kwang-il Lim³ and Jinhan Cho¹

¹ Department of Chemical and Biological Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea

² School of Advanced Materials Engineering, Kookmin University, Jeongneung-dong, Seongbuk-gu, Seoul 136-702, Korea

³ Department of Medical and Pharmaceutical Sciences, Sookmyung Women's University, Cheongpa-ro 47-Gil 100, Yongsan-gu, Seoul 140-742, Korea

E-mail: klim@sookmyung.ac.kr and jinhan71@korea.ac.kr

Received 14 December 2011, in final form 3 March 2012

Published 28 March 2012

Online at stacks.iop.org/Nano/23/155604

Abstract

The properties of enzymes, which can cause reversible changes in currents through redox reactions in solution, are of fundamental and practical importance in bio-electrochemical applications. These redox properties of enzymes are often associated with their charge-trap sites. Here, we demonstrate that reversible changes in resistance in dried lysozyme (LYS) films can be generated by an externally applied voltage as a result of charge trap/release. Based on such changes, LYS can be used as resistive switching active material for nonvolatile memory devices. In this study, cationic LYS and anionic poly(styrene sulfonate) (PSS) layers were alternately deposited onto Pt-coated silicon substrates using a layer-by-layer assembly method. Then, top electrodes were deposited onto the top of LYS/PSS multilayers to complete the fabrication of the memory-like device. The LYS/PSS multilayer devices exhibited typical resistive switching characteristics with an ON/OFF current ratio above 10^2 , a fast switching speed of 100 ns and stable performance. Furthermore, the insertion of insulating polyelectrolytes (PEs) between the respective LYS layers significantly enhanced the memory performance of the devices showing a high ON/OFF current ratio of $\sim 10^6$ and low levels of power consumption.

 Online supplementary data available from stacks.iop.org/Nano/23/155604/mmedia

(Some figures may appear in colour only in the online journal)

1. Introduction

The integration of enzymes with electronic elements to yield functional devices has attracted considerable attention because of the potential of these devices in applications such as biosensors and biofuel cells [1–6]. Recent research efforts have been directed toward the utilization of enzymes for the assembly of practical memory devices. For example,

it has been shown that ferritin with a large quantity of Fe ions in its inner cavity of approximately 8 nm could be converted into magnetite nanoparticles via pyrolysis of organic components. The resulting nanoparticles could be used as charge-trap elements in traditional complementary metal–oxide–semiconductor (CMOS) devices [7, 8]. Another study has showed how an electrochemical redox reaction within enzymes, monitored by cyclic voltammetry in an aqueous buffer solution, can be employed as a electric memory unit (i.e. oxidation involving charge release and

⁴ These authors contributed equally to this work.

reduction involving charge trap). Although modifications need to be made to improve their performance and simplify their fabrication processes to the level of conventional vacuum-deposited transition-metal-oxide-coated or spin-coated organic (i.e. synthesized organometallic compounds) memory devices [9–18], these findings from bio-inspired memory devices suggest intriguing new potentials of biomaterials for use in memory devices. Our study has been motivated based on the possibility that redox reactions within enzymes in solution can be utilized for nonvolatile memory in organic devices that operate via charge-trap/release mechanisms under externally applied voltages in an air atmosphere. Furthermore, instead of using single layer structures that are prepared from spin-coating processes, we relied on intelligently designed heterogeneous film structures to enhance the memory performance of devices.

Among the various deposition methods available, the layer-by-layer (LbL) assembly method, which is based on a solution dipping process and complementary interactions (i.e. electrostatic, hydrogen-bonding and covalent-bonding interactions), has been reported to be quite useful in the preparation of organic and/or inorganic nanocomposite films that have tailored electrical properties and layer thicknesses and have various functional components on substrates with different sizes and shapes [19–30]. The Advincula group has reported that the nanometer-scale charging of LbL films, based on conjugated polymers, can provide a write–read device using current-sensing atomic force microscopy [27]. Recently, the Pal group has reported that electrostatically assembled multilayers composed of CdSe quantum dots could exhibit electrical bistability according to an externally applied voltage sweep [28]. Our group has also reported that LbL multilayer films containing gold and iron oxide nanoparticles could be employed as active layers for charge-trap flash [29] and resistive switching nonvolatile memory devices [16], respectively.

Here, we demonstrate for the first time the use of lysozyme (LYS) proteins, in which charge release and trapping take place, as resistive switching active materials in nonvolatile memory devices. In this study multilayer films composed of cationic LYS and anionic polyelectrolyte (PE) were incorporated onto Pt-coated substrates using the electrostatic LbL assembly method. We ultimately aimed to use the LbL assembly method to modulate and significantly enhance the memory performance (i.e. to obtain a high ON/OFF current ratio greater than $\sim 10^6$ and a low level of current flow) of LYS multilayers additionally taking advantage of the additional insertion of insulating PEs of nanoscale thickness. To the best of our knowledge, the finding of nonvolatile memory functionalities of enzymes and the control of their memory performance via molecular-level manipulation of enzyme film architectures have been reported in detail for the first time in this study. We believe that our approach will be very effective for the preparation of bio-inspired memory devices, ultimately allowing easy modifications of electrical properties, low power consumption and tailored nanostructures under mild conditions (i.e. room temperature and an air atmosphere).

2. Materials and Methods

2.1. Preparation of multilayers

The concentrations of LYS (from hen egg white, Sigma Aldrich), poly(styrene sulfonate) (PSS; $M_w = 70\,000$, Sigma Aldrich), poly(acrylic acid) (PAA; $M_w = 1800$, Sigma Aldrich) and PAH ($M_w = 70\,000$, Sigma Aldrich) solutions used for all the experiments were 1 mg ml^{-1} . The pH levels of the LYS, PAH, PSS and PAA solutions were adjusted to 6. For these experiments, LYS and PAH were used as the cationic components, and PSS and PAA were used as the anionic components. The anionic surface of the Pt-coated Si substrates was prepared via irradiation with UV light. These substrates were dipped in the cationic LYS or PAH solution for 10 min, washed twice by submersion in water for 1 min, and then air-dried with a gentle stream of nitrogen. The anionic PSS or PAA was subsequently deposited onto the LYS (or PAH)-coated substrates using the same adsorption, washing, and drying procedures described above. This process was repeated until the desired number of layers had been deposited; the resultant multilayer films were then sufficiently dried to remove the residual moisture under an air atmosphere.

2.2. Quartz-crystal microgravimetry (QCM) measurements

A QCM device (QCM200, SRS) was used to quantify the mass of the material deposited in each adsorption step. The resonance frequency of the QCM electrodes was about 5 MHz. The adsorbed mass of the LYS and PSS, Δm , can be calculated from the change in the QCM frequency, ΔF , according to the Sauerbrey equation: $\Delta F\text{ (Hz)} = -56.6\Delta m_A$, where Δm_A is the mass change per quartz-crystal unit area in $\mu\text{g cm}^{-2}$. Because Kasemo and co-workers have reported that the correlation described in the Sauerbrey equation (i.e. the adsorbed mass and the frequency change) is difficult to observe when applied to viscoelastic, thick or hydrogel layers that contain water molecules in the solid–liquid interface [30], the QCM measurements in our study were made only after the adsorbed layer was sufficiently dried using nitrogen gas. In addition, the respective thicknesses of the polyelectrolytes and LYS were less than 2 nm per layer. Therefore, the polyelectrolytes/LYS multilayers that were adsorbed on the crystal surface can be regarded as rigid, but evenly distributed and thin, films that satisfy the Sauerbrey equation.

2.3. X-ray reflectometry

The internal structure and film thickness of LYS/PSS multilayer films were determined by x-ray reflectivity using a Cu K α ($\lambda = 1.54\text{ \AA}$) beam from a narrow line source of a 18 kW Rigaku Ru-300 rotating anode generator.

2.4. Measurement of catalytic properties of lysozymes using an Enzchek lysozyme assay kit

Fluorescence-labeled *Micrococcus lysodeikticus* cells were dissolved in a reaction buffer solution with pH of 7.5 at

Astrocytes regulate adult hippocampal neurogenesis through ephrin-B signaling

Randolph S Ashton^{1,3,4}, Anthony Conway^{1,4}, Chinmay Pangarkar², Jamie Bergen¹, Kwang-Il Lim^{1,3}, Priya Shah¹, Mina Bissell² & David V Schaffer¹

Neurogenesis in the adult hippocampus involves activation of quiescent neural stem cells (NSCs) to yield transiently amplifying NSCs, progenitors, and, ultimately, neurons that affect learning and memory. This process is tightly controlled by microenvironmental cues, although a few endogenous factors are known to regulate neuronal differentiation. Astrocytes have been implicated, but their role in juxtacrine (that is, cell-cell contact dependent) signaling in NSC niches has not been investigated. We found that ephrin-B2 presented from rodent hippocampal astrocytes regulated neurogenesis *in vivo*. Furthermore, clonal analysis in NSC fate-mapping studies revealed a previously unknown role for ephrin-B2 in instructing neuronal differentiation. In addition, ephrin-B2 signaling, transduced by EphB4 receptors on NSCs, activated β -catenin *in vitro* and *in vivo* independently of Wnt signaling and upregulated proneural transcription factors. Ephrin-B2⁺ astrocytes therefore promote neuronal differentiation of adult NSCs through juxtacrine signaling, findings that advance our understanding of adult neurogenesis and may have future regenerative medicine implications.

In the mammalian brain, neurogenesis persists throughout adulthood in the subgranular zone (SGZ) of the hippocampal dentate gyrus¹ and the subventricular zone (SVZ) of the lateral ventricles². In these niches, neural stem cell (NSC) maintenance, proliferation, and differentiation are orchestrated by a delicate balance of microenvironmental cues. The importance of such instructive signals in the niche is supported by the absence of significant neurogenesis in the adult mammalian CNS outside of these regions^{3,4} and by the development of aberrant neurogenesis in patients with pathologies that disrupt the natural chemistry of the brain, for example epilepsy⁵, inflammation⁶, and neurodegenerative diseases⁷.

In the SGZ, adult neurogenesis entails the activation of quiescent type 1 NSCs to produce mitotic, multipotent type 2a NSCs^{8,9}; the differentiation of the type 2a NSCs to lineage-committed, proliferative type 2b neuronal precursors and subsequently type 3 neuroblasts; and the survival, migration, and maturation of neuroblasts as they differentiate into granule cells that synaptically integrate into the existing neural network^{8–10}. Each of these stages is regulated by signals within the niche. For example, bone morphogenic protein¹¹ and Notch^{12,13} signaling modulate the balance between quiescent and proliferative NSCs, and Sonic hedgehog¹⁴, fibroblast growth factor-2 (ref. 15), vascular endothelial growth factor¹⁶ and Wnt7a¹⁷ regulate NSC proliferation.

Although there is increasing knowledge of niche factors that regulate NSC division, substantially less is known about key signals that instruct cells in the SGZ to undergo neuronal differentiation^{18,19}. GABA inputs from local neuronal circuitry²⁰ and systemic retinoic

acid levels²¹ modulate NSC neuronal fate commitment. Furthermore, while neuronal differentiation precedes gliogenesis during CNS development, adult hippocampal astrocytes directly induce neuronal differentiation of NSCs *in vitro* via both secreted and membrane-associated factors²², and the former have since been found to include Wnt3a²³ and potentially additional secreted signals¹⁹. However, the membrane-bound astrocytic components²² that may be critical for neuronal fate commitment *in vivo* remain unknown.

Ephrins are a diverse class of glycoposphatidylinositol-linked (ephrinA1–6) and transmembrane (ephrinB1–3) cell surface ligands that bind Eph receptors (EphA1–10 and EphB1–6) on opposing cell membranes to initiate bidirectional signaling²⁴. Ephrin-Eph signaling is traditionally known to control the spatial organization of cells and their projections by modulating intercellular attractive and repulsive forces²⁵. For example, ephrin-Eph signaling instructs topographical mapping of hippocampo-septal and entorhino-hippocampal projections during development²⁶ and regulates neuronal dendrite spine morphogenesis and synaptogenesis of adult hippocampal neurons²⁷.

Recent studies have also indicated that ephrin-Eph signaling has an earlier role in regulating stem cell behavior. For example, ephrin-A-EphA signaling promotes embryonic telencephalic NSC differentiation²⁸, and ephrin-B3-EphB3 signaling may exert an anti-proliferative effect on NSCs in the developing SVZ²⁹. In the adult CNS, infusion of ephrin-B2 or EphB2 ectodomains into the lateral ventricles induced SVZ NSC proliferation and disrupted neuroblast migration through the rostral migratory stream³⁰. In addition, signaling between ephrin-A2⁺ neural stem cells, and EphA7⁺ ependymal cells and putative stem cells,

¹Department of Chemical and Biomolecular Engineering, University of California Berkeley, Berkeley, California, USA. ²Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA. ³Present address: Department of Biomedical Engineering, Wisconsin Institutes for Discovery, University of Wisconsin Madison, Madison, Wisconsin, USA (R.S.A.), Department of Medical and Pharmaceutical Sciences, College of Science, Sookmyung Women's University, Seoul, Korea (K.-I.L.). ⁴These authors contributed equally to this work. Correspondence should be addressed to D.V.S. (schaffer@berkeley.edu).

Received 22 May; accepted 9 August; published online 16 September 2012; doi:10.1038/nn.3212

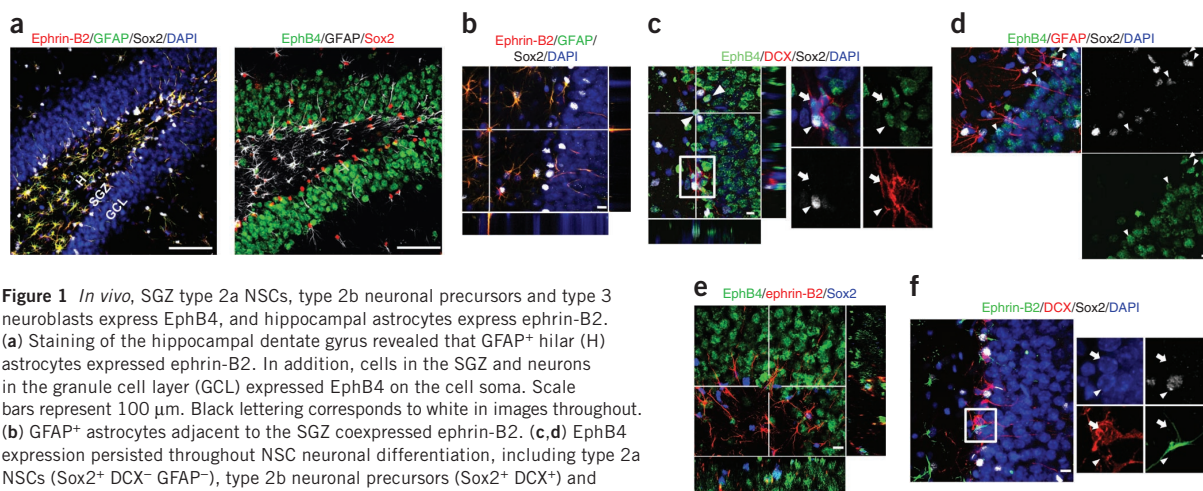


Figure 1 *In vivo*, SGZ type 2a NSCs, type 2b neuronal precursors and type 3 neuroblasts express EphB4, and hippocampal astrocytes express ephrin-B2. (a) Staining of the hippocampal dentate gyrus revealed that GFAP⁺ hilar (H) astrocytes expressed ephrin-B2. In addition, cells in the SGZ and neurons in the granule cell layer (GCL) expressed EphB4 on the cell soma. Scale bars represent 100 μ m. Black lettering corresponds to white in images throughout. (b) GFAP⁺ astrocytes adjacent to the SGZ coexpressed ephrin-B2. (c,d) EphB4 expression persisted throughout NSC neuronal differentiation, including type 2a NSCs (Sox2⁺ DCX⁻ GFAP⁻), type 2b neuronal precursors (Sox2⁺ DCX⁺) and type 3 neuroblasts (DCX⁺). Confocal images in c show EphB4 expression in Sox2⁺ DCX⁻ cells (large arrowheads) in the SGZ. The magnified region depicts the presence of EphB4 expression on a Sox2⁺ DCX⁺ type 2b neuronal precursor (small arrowheads) and a DCX⁺ type 3 neuroblast (arrows). Confocal images in d show EphB4 expression by Sox2⁺ GFAP⁻ cells (small arrowheads); thus, Sox2⁺ DCX⁻ GFAP⁻ type 2a NSCs also expressed EphB4. (e,f) Given the close proximity of ephrin-B2⁺ astrocytes to EphB4⁺ cells in the SGZ, ephrin-B2-EphB4 juxtacrine signaling is in a position to induce NSC differentiation into Sox2⁺ DCX⁺ type 2b neuronal precursors (small arrowheads) and subsequently Sox2⁻ DCX⁺ type 3 neuroblasts (arrows). The scale bars represent 10 μ m (b-f). Crosshairs represent the position of x-z (right panel) and y-z (bottom panel) orthogonal slices of the confocal stack depicted in the image (b,d,e).

was shown to suppress proliferation of NSCs in the adult SVZ³¹. In the adult SGZ, ephrin-B3⁺ neurons in the granule cell layer have been proposed to regulate EphB1⁺ NSC polarity, SGZ positioning, and proliferation³². Similarly, in *EfnA5*^{-/-} mice, a decrease in both the proliferation of NSCs and the survival and maturation of newborn neurons in the adult SGZ was observed³³. Thus, ephrin-Eph signaling has been shown to affect the proliferation, migration and survival of adult NSCs; however, its potential regulation of NSC fate commitment remains unknown.

We found that ephrin-B2 presented by hippocampal astrocytes instructed neuronal differentiation of NSCs in the SGZ of the adult hippocampus. Furthermore, ephrin-B2-EphB4 forward signaling induced neuronal differentiation of NSCs by activating β -catenin, independent of Wnt signaling, and inducing transcription of proneural transcription factors. Our findings describe a juxtacrine signaling mechanism by which astrocytes actively regulate neuronal differentiation of NSCs during adult neurogenesis.

RESULTS

Ephrin-B2 and EphB4 expression in the SGZ

In the SGZ, type 1 NSCs express the SRY-box 2 (Sox2) transcription factor and stain for glial fibrillary acid protein (GFAP) and Nestin along radial granule cell layer (GCL)-spanning or horizontal processes¹³. Type 2a NSCs remain Sox2⁺ and Nestin⁺, but downregulate GFAP expression, and fate-restricted, proliferative type 2b neuronal precursors coexpress Sox2 and doublecortin (DCX) while downregulating Nestin expression. The latter then mature into (Sox2⁻ DCX⁺) type 3 neuroblasts, which migrate into the GCL and become neuronal-specific nuclear protein (NeuN)⁺ DCX⁻ granule neurons^{8,10,11,13}. Alternatively, type 1 NSCs can differentiate into stellate, Sox2⁻ GFAP⁺ hippocampal astrocytes, which are primarily located in the hilus adjacent to the SGZ¹⁰.

As part of a candidate screen, we investigated ephrin-B2 expression in the hippocampal dentate gyrus. Antibody to Ephrin-B2 consistently labeled GFAP⁺ astrocytes in the hilus, SGZ and molecular layer

(Fig. 1a,b). In contrast, EphB4, an ephrin-B2 receptor, was expressed by Sox2⁺ DCX⁻, Sox2⁺ DCX⁺, Sox2⁻ DCX⁺ and Sox2⁺ GFAP⁻ cells in the SGZ, as well as by the majority of cells in the GCL (Fig. 1a,c,d). Thus, EphB4 appeared to be expressed in type 2a (Sox2⁺ DCX⁻ GFAP⁻) NSCs and persisted as these cells became type 2b neuronal precursors (Sox2⁺ DCX⁺), type 3 neuroblasts (Sox2⁻ DCX⁺) and eventually granule neurons. Some type 1 NSCs were EphB4⁺, although EphB4 was not expressed at levels detectable by immunostaining in many cells (data not shown). Overall, these results indicate that astrocytes are a source of ephrin-B2 ligand in the dentate gyrus and suggest that EphB4-expressing NSCs, neuronal precursors and neuroblasts contact these ligand-expressing astrocytes (Fig. 1e,f).

Ephrin-B2 increases neurogenesis *in vitro* and *in vivo*

Ephrin-Eph signaling begins with clustering of multiple ligand-receptor complexes at sites of cell-cell contact²⁴. Thus, to initially explore a possible role for ephrin-B2 in regulating NSC fate, we added antibody-clustered ephrin-B2-Fc fusion molecules (Fc-ephrin-B2)³⁴ to adult hippocampus-derived NSCs in culture and found that they induced a strong, dose-dependent increase in NSC neuronal differentiation (Fig. 2a,b). No biological activity was observed with monomeric, unclustered ephrin-B2 and Fc, consistent with previous findings^{30,34}, and no proliferative effect was observed under any conditions (data not shown). To investigate which receptor mediated Fc-ephrin-B2's activity, we pre-incubated the NSCs with an antibody to either EphB2 or EphB4, two known ephrin-B2 receptors, before adding Fc-ephrin-B2. Inhibition of EphB4, but not EphB2, significantly reduced Fc-ephrin-B2 induction of neuronal differentiation ($P = 0.002$; Fig. 2c), indicating that EphB4 receptors mediate the proneuronal effect of Fc-ephrin-B2. Reverse transcription PCR (RT-PCR) and immunocytochemistry results confirmed that hippocampus-derived NSCs express EphB4 (Fig. 2d), in contrast with SVZ NSCs, which reportedly do not express this receptor (ref. 30).

Given our staining and *in vitro* results, we hypothesized that astrocytic ephrin-B2 actively promotes neuronal differentiation of NSCs

Tamoxifen Resistance in Breast Cancer

Minsun Chang*

Department of Medical and Pharmaceutical Science, College of Science, Sookmyung Women's University, Seoul 140-742, Republic of Korea

Abstract

Tamoxifen is a central component of the treatment of estrogen receptor (ER)-positive breast cancer as a partial agonist of ER. It has been clinically used for the last 30 years and is currently available as a chemopreventive agent in women with high risk for breast cancer. The most challenging issue with tamoxifen use is the development of resistance in an initially responsive breast tumor. This review summarizes the roles of ER as the therapeutic target of tamoxifen in cancer treatment, clinical values and issues of tamoxifen use, and molecular mechanisms of tamoxifen resistance. Emerging knowledge on the molecular mechanisms of tamoxifen resistance will provide insight into the design of regimens to overcome tamoxifen resistance and discovery of novel therapeutic agents with a decreased chance of developing resistance as well as establishing more efficient treatment strategies.

Key Words: Breast cancer, Estrogen receptor, Tamoxifen, Antiestrogen resistance

INTRODUCTION

Breast cancer is one of the most common cancers and the second leading cause of cancer death in American women according to the estimation by the American Cancer Society (ACS). Over 200,000 new cases of breast cancer are diagnosed each year and more than 40,000 deaths have been reported in women in the USA in 2010 alone (Jemal *et al.*, 2010). Despite advances in early detection, therapeutic regimens, and understanding the molecular basis of breast cancer biology, nearly 30% of all patients with early stage breast cancer have recurring disease, which is metastatic in most cases. Breast cancer is a hormone-dependent tumor and estrogen is known to play a major role in the initiation and progression of the disease. Biological actions of estrogen are mediated with the estrogen receptor (ER) and nearly 70% of breast tumors express the ER, progesterone receptor (PR), and/or estrogen-responsive and ER-dependent gene products. Therefore, targeting the ER using ER antagonists or antiestrogens has been a reliable therapeutic measure for all stages of the disease in both pre- and postmenopausal women.

Tamoxifen is a triphenylethylene derivative pharmacologically classified as a selective ER modulator (SERM) that acts as an agonist in the uterus but as an antagonist in the breast. Tamoxifen is the most commonly used chemotherapeutic agent for patients with ER-positive breast cancer, which represents almost 70% of all cases. In the hormone-sensitive breast, tamoxifen acts as a partial antagonist, impairing ER

function by competing with estrogen for binding to the receptor (Banerjee *et al.*, 2003). Tamoxifen bound ER complex prevents the genes from being switched on by estrogen, leading to inhibition of estrogenic effects, which are responsible for cancer cell growth or proliferation (Chang *et al.*, 2007). Many ER-positive patients regardless of the high level of ER demonstrate intrinsic resistance to hormonal therapies. A substantial proportion of patients with localized disease and all nearly with advanced disease who initially respond to tamoxifen develop *de novo* or acquired resistance (EBCTCG, 2005). Interestingly, many patients who relapse on tamoxifen therapy will respond to different types of hormonal manipulation including either ER downregulators/ER antagonists or aromatase inhibitors (AIs), implying that ER continues to play a major role in breast cancer progression (Pike *et al.*, 1993; Forbes *et al.*, 2008; Mouridsen *et al.*, 2009).

Although the molecular mechanisms underlying resistance to tamoxifen remains unclear, various mechanisms have been proposed; for example, differential metabolic activation of tamoxifen, loss of ER function/expression, alterations in crosstalk between ER and growth factor-mediated signaling pathways, the presence of ER-negative cancer stem cells, and dynamic responses to oxidative stress.

This review will focus the molecular basis of the ER-mediated mammary carcinogenesis, clinical use of tamoxifen in breast cancer treatment targeting the ER, and various molecular mechanisms of tamoxifen resistance. A better understanding of the mechanisms for tamoxifen resistance may provide

www.biomolther.org

Open Access <http://dx.doi.org/10.4062/biomolther.2012.20.3.256>

pISSN: 1976-9148 eISSN: 2005-4483

Copyright © 2012 The Korean Society of Applied Pharmacology

Received Mar 6, 2012 Revised Mar 29, 2012 Accepted Apr 2, 2012

***Corresponding Author**

E-mail: minsunchang@sookmyung.ac.kr

Tel: +82-2-2077-7626, Fax: +82-303-0801-1074

novel strategies to overcome or bypass tamoxifen resistance and make further improvements in cancer therapeutics. Finally, this review will suggest future directions on tamoxifen-associated endocrine therapies.

THE BIOLOGY OF ESTROGEN RECEPTOR-MEDIATED SIGNALING IN MAMMARY CARCINOGENESIS

Mechanisms of action of the estrogen receptors

Two ERs, ER α and ER β , belong to the nuclear receptor (NR) family of transcription factors (Greene *et al.*, 1986). Both receptors are encoded by unique genes composed of eight exons, which are located on different chromosomes with ER α found on the long arm of chromosome 6 and ER β on chromosome 14 (Fig. 1) (Chang *et al.*, 2002). They contain the evolutionarily conserved modular structure of functional domains characteristic of many other NR families (Mangelsdorf *et al.*, 1995). The central and most conserved domain, the DNA-binding domain (DBD), is involved in DNA recognition and binding, and exhibits a high degree of homology between both ER α and ER β (Fig. 1). Another functional domain includes the ligand-binding domain (LBD), which occurs near the C-terminal with 60% homology between two ERs. The N-terminal domain represents the most variable domain where homology between two ERs was found to be only 18%. In addition, hinge (E region) and F regions are also not well conserved (Mosselman *et al.*, 1996). Transcriptional activation is facilitated by two distinct activation functions, AF-1 and AF-2. From the structural and functional points of view, it is reasonable to conclude that the two ERs interact with the same DNA response elements and have similar binding affinities for 17 β -estradiol (E2) (Kuiper *et al.*, 1997).

Binding of E2 to its receptor, ER α/β , amplifies signals in either the genomic, nongenomic, or mitochondrial ER-mediated signaling pathways that lead to increased cell proliferation and inhibition of apoptosis, uncontrolled cell division or growth and

tumor promotion (Björnström and Sjöberg, 2005). Molecular pathways for ER-mediated signaling transduction are found elsewhere (Heldring *et al.*, 2007) and therefore, they will not be discussed in this review.

Tamoxifen binds to the ER and undergoes conformational change in a similar fashion in which E₂ interacts with the ER, except for the repositioning of helix 12 (Shiau *et al.*, 1998). Analysis of the E₂-ER LBD complex showed that helix 12 is packed against helix 3, 5/6, and helix 11, sealing the ligand binding pocket like a lid (Brzozowski *et al.*, 1997). This conformation allows the specific amino acids in the AF-2 region to interact with coactivators for transcriptional activation of the target genes. In the 4-OHT [4-hydroxytamoxifen is the active metabolite of tamoxifen *in vivo* (Jordan, 1977; Jordan *et al.*, 1977)]-ER LBD complex, helix 12 binds to and occludes the coactivator recognition site. The binding mode of the side chain of 4-OHT results in repositioning of helix 12 that prevents the binding of coactivators and thus AF-2 transcription is blocked. Therefore, tamoxifen acts as an ER antagonist on genes which depend on AF-2 activation for ER-mediated transcription. On the other hand, tamoxifen may function as an agonist in the transcription of genes where the AF-1 domain, instead of AF-2 plays a critical role in transcription (McGuire *et al.*, 1977; Tzukerman *et al.*, 1994). The presence of cell or tissue-specific AF-1 or AF-2-activated genes may explain the differential agonistic or antagonistic activity of tissue-specific tamoxifen. The potential for the agonist activity of tamoxifen to outweigh its antagonist effects may serve as one of the molecular ER-mediated mechanisms of tamoxifen resistance. Since the therapeutic target for breast cancer intervention is the ER α , ER refers to ER β throughout this review unless specified otherwise.

Estrogen and ER-mediated carcinogenesis

Estrogen is implicated in the development of breast cancer, based on the data from the relationship between life-time estrogen exposure and cancer incidence (Henderson and Feigelson, 2000). The greater increase in the risk of developing breast cancer, the longer women are exposed to estrogen either through early onset of menstruation, nulliparity, delayed first childbirth, short duration of breast feeding, late menopause, and estrogen replacement therapy. Estrogens are considered to play a critical role in promoting the proliferation of both the normal and the neoplastic breast epithelium. The most biologically active estrogen, E₂, acts locally in the mammary gland, stimulating DNA synthesis and promoting bud formation through an ER-mediated mechanism (Russo *et al.*, 2000). The fact that the normal epithelium contains receptors for estrogen lends support to the receptor-mediated mechanism as a major player in the hormonal regulation of breast development. The role of estrogen on the proliferative activity of the breast, which is indispensable for its normal growth and development, has been for a long time the subject of controversy (Russo and Russo, 2006). Experimental animal systems and clinical observations, however, support the carcinogenicity of estrogens. There are three major mechanisms that have been considered as the molecular mechanisms underlying the estrogen-induced carcinogenesis; 1) ER-mediated hormonal activity, which has been related to stimulation of cell proliferation and rates of cell division, resulting in more opportunities for accumulation of genetic damages (Russo *et al.*, 1988; Habbel and Stanford, 1993; Russo *et al.*, 1999), 2) direct geno-

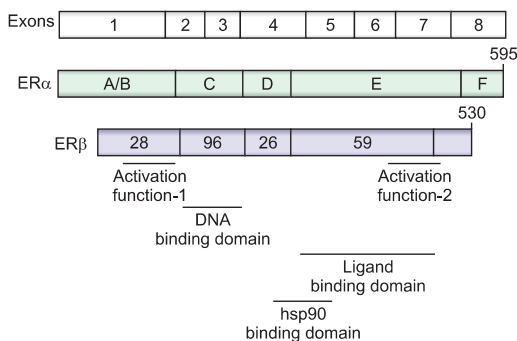


Fig. 1. Schematic representation of the common structural and functional domains of ER α and ER β . The cDNA coding for ER is composed of 8 exons. The functional domains of the receptors are indicated (A-F), as are the regions responsible for transcriptional activation, DNA binding, ligand binding, and heat shock protein (hsp) 90 binding. The percentage of amino acid homology between regions A through F in ER α and ER β is indicated by the numbers written on the structure of ER β . Two transcriptional activation domains, activation function (AF) domains are present in the A/B and E region of the ERs.



Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Preclinical metabolism of LB42908, a novel farnesyl transferase inhibitor, and its effects on the cytochrome P450 isozyme activities

Minsun Chang^{a,†,||}, Sung-Hak Lee^{a,||}, Ho Jun Kim^{a,‡}, Jong-Sung Koh^{b,§}, Aeri Kim^{c,¶,*}^a Drug Metabolism and Pharmacokinetics Department, LG Life Sciences R&D Center, Daejeon, Republic of Korea^b Medicinal Chemistry Department, LG Life Sciences R&D Center, Daejeon, Republic of Korea^c Division of Business Development, LG Life Sciences, Seoul, Republic of Korea

ARTICLE INFO

Article history:

Received 29 December 2011

Revised 18 March 2012

Accepted 19 March 2012

Available online 27 March 2012

Keywords:

Farnesyl transferase inhibitor

Preclinical hepatic metabolism

Drug–drug interaction

Cytochrome P450

ABSTRACT

Metabolism of LB42908, a novel farnesyl transferase inhibitor, was investigated for preclinical development. In vitro hepatic metabolism of LB42908 gave rise to at least 9 metabolites via phase I biotransformation pathways, which were characterized by HPLC–UV, LC–MS, and LC–MS/MS analyses. N-Dealkylation was shown to be a major phase I metabolic pathway. Species-specific in vitro metabolism of LB42908 was studied in liver fractions of rat, dog, monkey, and human. Order of metabolic stability is human $\approx$ dog $>$ rat $\approx$ monkey in both S9 and microsomal fractions. Tissue-specific metabolism of LB42908 in various tissue homogenates of rats demonstrated that the liver was the major organ responsible for phase I metabolism of LB42908. The results from both qualitative and quantitative metabolism studies such as metabolic profiling and metabolic clearance indicated that dog would be the animal model of choice for preclinical toxicology studies. In addition, LB42908 was a potent CYP3A4 inhibitor in human liver microsomes and induced the activities of several CYP isozymes, implying that it has the potential for drug–drug interactions. Repeated dosing of LB42908 in rats did not significantly affect its own metabolism, indicating that long-term administration of LB42908 would not alter its pharmacokinetic profiles.

© 2012 Elsevier Ltd. All rights reserved.

Farnesyl transferase is a molecular target of interest in cancer therapy.¹ It catalyzes a key step in coupling of an aliphatic isoprenoid group to a wide range of target proteins. Farnesyl transferase inhibitors (FTIs) are a new class of biologically active anticancer agents that specifically inhibit FT and target Ras, a G-protein with 4 isoforms (H-, N- and K-RasA/K-Ras B), that requires prenylation for its activity.^{2,3} Ras plays a pivotal role in relaying cell growth-stimulating signals, and mutation of the *ras* gene may lead to constant activation of the protein, ultimately resulting in uncontrolled cell proliferation. The antiproliferative and antitumor activities of FTIs have been demonstrated in preclinical and clinical studies. However, it has become clear in many instances that Ras may not be the critical or only target of FTIs.^{4,5} These observations suggest the possibility that many other protein targets containing

CAAX are subject to farnesylation by FT and may be potential targets for FTI.⁶ Clinical trials of FTI monotherapy produced just modest efficacy, which was contributed to the alternative prenylation of Ras and some other proteins by GGTase-1.^{7,8} Recently observed synergistic efficacy in both preclinical and clinical studies reignited interest in FTI as an anticancer agent.⁹

LB42908, [1-[[1-(1,3-benzodioxole-5-ylmethyl)-1H-imidazole-5-yl]methyl]-4-(1-naphthyl)-1H-pyrrole-3-yl](4-methyl-1-piperazinyl)methanone, is a novel pyrrole-based FTI developed as an anticancer drug (Structure in Fig. 1). Its antitumor activity is believed to be derived from suppressing the growth of tumors with an activated ras oncogene. It displayed excellent antitumor activities in both in vitro and in vivo animal models and had a high oral bioavailability in preclinical models.^{10,11} However, systemic exposure level and clearance in male Sprague–Dawley (SD) rats following oral administration of LB42908 showed significant gender differences. In addition, pretreatment of female rats with dexamethasone (known as a CYP3A inducer) caused elevated clearance and a decrease in oral absorption. Since the pharmacokinetic profile of a compound is the primary determinant of its efficacy and safety during the drug development process, preclinical studies of a candidate compound provide a great understanding and opportunity to evaluate its potential as a candidate for clinical development. Our preclinical investigation of LB42908, including pharmacology, absorption, and distribution, prompted us to investigate the

* Corresponding author. Tel.: +82 2 3468 3678; fax: +82 2 555 6689.

E-mail addresses: arkim512@gmail.com, arkim@cha.ac.kr (A. Kim).[†] Present address: Department of Medical and Pharmaceutical Science, College of Science, Sookmyung Women's University, 52 Hyochangwon Road, Yongsan-Gu, 140-742 Seoul, Republic of Korea.[‡] Present address: Center for Doping Control, Korea Institute of Science and Technology, Hwarang-No, Sungbuk-Gu, 136-791 Seoul, Republic of Korea.[§] Present address: GENOSCO, 2nd Floor, 767C Concord Avenue, Cambridge, MA 02138, USA.[¶] Present address: College of Pharmacy, CHA University, 606-16 Yeoksam 1-dong, Gangnam-gu, 135-081 Seoul, Republic of Korea.^{||} These two authors contributed equally to this work.

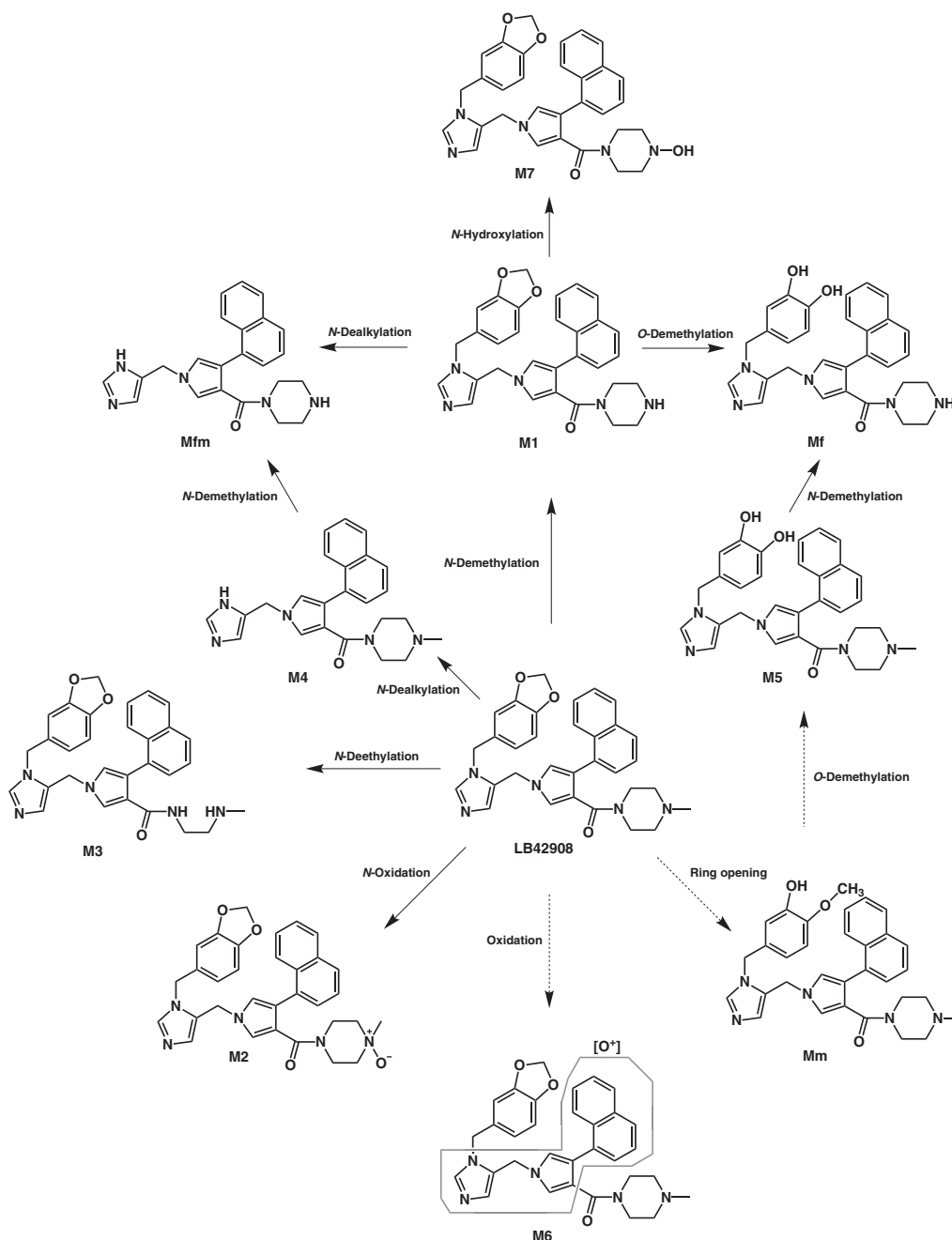


Figure 1. The proposed metabolic pathways for LB42908 in liver S9 and microsomal fractions of four species. Structures of M1 to M7, Mf, and Mfm were elucidated using LC-MS/MS analysis and their authentic forms were synthesized for further confirmation. The LC-MS/MS chromatogram to represent the mass fragmentation for each metabolite is available in [Supplementary data](#) (B). Solid arrows represent the pathways confirmed by the authentic forms whereas dotted arrows represent proposed ones.

metabolism of LB42908 to determine whether to proceed further with preclinical and/or clinical development of LB42908. In particular, it was expected that species-specific *in vitro* or *in vivo* metabolism studies would provide meaningful information on the metabolic clearance pattern of LB42908 and that the study results would help us choose preclinical species for further preclinical development prior to a first-in-man study. Inhibitory and inductive effects of a specific cytochrome P450 (CYP) isozyme by LB42908

were examined to explore the potential of LB42908 in drug–drug interactions as part of a preclinical investigation of drug safety. In this Letter, we describe the metabolic pathways, metabolite identification, and metabolic profiling of LB42908 in four different species, namely, rat, dog, monkey, and human. The effects of LB42908 on CYP activities *in vivo* will also be discussed.

NADPH-dependent phase I metabolism of LB42908 was confirmed in liver fractions including S9 and microsomes of rat

GATA-binding protein 1 is a novel transcription regulator of peroxiredoxin 5 in human breast cancer cells

MIN JI SEO¹, XIYUAN LIU¹, MINSUN CHANG² and JONG HOON PARK¹

Departments of ¹Biological Science and ²Medical and Pharmaceutical Science, Sookmyung Women's University, 52 Hyochangwon Road, Yongsan-gu, Seoul 140-742, Republic of Korea

Received August 8, 2011; Accepted September 28, 2011

DOI: 10.3892/ijo.2011.1236

Abstract. Peroxiredoxin 5 (Prx5) is a member of a family consisting of six antioxidant enzymes. Prx5 is ubiquitously expressed in various tissues including mitochondria and peroxisomes, implying that Prx5 functions as a regulator of the cellular oxidation state. Prx5 plays a critical role in protecting cells from oxidative stress by inhibiting the accumulation of reactive oxygen species and cell death. Overexpression of Prx5 in mammary tissue is associated with poor prognosis of breast cancer. The present study was conducted to elucidate the regulatory mechanisms of Prx5 gene expression and the physiological relevance of Prx5 in breast cancer cell survival. Analysis of the promoter region of the Prx5 gene and reporter gene assays revealed the promoter region critical for Prx5 gene regulation to which the novel negative transcription regulator, GATA1 binds in human breast cancer cell lines. GATA1 was further confirmed as a potential transcription factor for Prx5 gene expression in another set of transfection assays using a reporter gene construct that contained a mutated GATA1 binding site. Direct binding of GATA1 to the Prx5 promoter was demonstrated using chromatin immunoprecipitation and electrophoretic mobility shift assays. Finally, functions of Prx5 associated with apoptosis were examined in two human breast cancer cell lines displaying high levels of either Prx5 or GATA1 expression. Knockdown of GATA1 led to increased expression of Prx5 and inhibition of apoptosis. Our data suggest that Prx5 may protect cells from oxidative stress-mediated apoptosis in a GATA1-regulated manner.

Introduction

Peroxiredoxins (Prxs) are a family of antioxidant enzymes that regulate the levels of reactive oxygen species (ROS) in cells (1). Since ROS are involved in necrotic or apoptotic cell death as well as the control of signal transduction, critical roles of Prxs include protecting cells against oxidative stress as well as serving as redox regulators of signal transduction (2,3). Knock-out of Prxs 1, 2, 3, or 6 in various animal or cell models results in abnormal physiological conditions such as anemia, malignancies, disruption of hematologic homeostasis, increased susceptibility to pro-inflammatory responses, shortened life span, and accelerated senescence (4-8). It has been noted that such effects are rescued upon overexpression of Prxs or catalase, implying a role of antioxidant Prxs and hydrogen peroxide (H₂O₂) (4,9,10).

There are six mammalian Prxs which all contain the conserved cysteine residue(s) responsible for peroxide scavenging. Prx1 through Prx5 have two catalytically-active cysteine residues whereas Prx6 has only one cysteine residue. The human Prx5 gene was cloned in 1997 and the full cDNA of this gene was isolated and sequenced in 1999 (11,12). Crystal structure studies and site-directed mutagenesis of human Prx5 suggested the functional importance of two cysteines (Cys47 and Cys151) as well as the role of Prx5 as an antioxidant (2,13). Human Prx5 gene organization as well as its promoter have been studied (14) and the expression of Prx5 in concert with Prx1 is regulated by interaction between the Ets transcription factor and high mobility protein group B1 in prostate cancer cells (15). However, the promoter-driven regulatory mechanisms of Prx5 gene and endogenous factors that modulate Prx5 gene expression remain unclear.

Prx5 is found in various cell compartments including the cytosol as well as the major intracellular sources of ROS, namely mitochondria and peroxisomes, suggesting that it may have an important role as a signaling molecule and antioxidant in organelles (14). The biological roles of Prx5 have been defined by its ability to prevent p53-induced apoptosis (3), inhibit various types of DNA damage (16,17), and suppress TNF- α -mediated accumulation of H₂O₂ (2). Overexpression of Prx5 in lung cancer cells has been associated with the protection of cells from apoptosis, implying that Prx5 prevents apoptosis or oxidative damage (18). In addition, the physiological roles of Prx5 were examined in *Drosophila* where Prx5 is thought to

Correspondence to: Dr Jong Hoon Park, Department of Biological Science, Sookmyung Women's University, Seoul 140-742, Republic of Korea

E-mail: parkjh@sookmyung.ac.kr

Dr Minsun Chang, Department of Medical and Pharmaceutical Science, Sookmyung Women's University, Seoul 140-742, Republic of Korea

E-mail: minsunchang@sookmyung.ac.kr

Key words: peroxiredoxin 5, GATA1, breast cancer, apoptosis, transcription factor

be involved in the modulation of immune responses and protection against oxidative stress that may lead to longevity (7,19). Interestingly, expression of Prx5 was shown to be dependent on tissue type and grade of breast and its expression was induced in response to ROS, indicating that Prx5 is involved in mammary tumorigenesis (20). The mechanistic roles and pathways of Prx5 in gene regulation, apoptosis, and carcinogenesis should be studied further to provide insights into the pathophysiological roles of Prx5 in breast cancer cells and tissues.

Our present study investigated the role of Prx5 in cell survival and apoptosis in human breast cancer cell lines and a novel regulatory mechanism of Prx5 gene expression. Our results showed that expression Prx5 gene was breast cancer cell line-specific. MCF7 cells resistant to the chemotherapeutic agent adriamycin MCF7-Adr showed high expression levels of Prx5 whereas MDA-MB-231 cells had low Prx5 expression levels. Both cell types are associated with multi-drug resistance and serious malignancies compared to other breast cancer cell lines (21). Analysis of the Prx5 promoter region revealed that the 500 bp core region was responsible for gene regulation and identified as the transcription regulator of the Prx5 gene in these two cell lines. We demonstrated that Prx5 may help protect cells against oxidative stress-mediated apoptosis in a DNA binding factor (GATA1)-dependent manner. In addition, we suggest that GATA1 acts as a transcription repressor of the Prx5 gene and that GATA1 is associated with Prx5 during apoptosis in a human breast cancer cell model.

Materials and methods

Cell culture and reagents. MCF7s cell line is a steroid-sensitive sub-line developed from MCF7 human breast cancer cells and was provided by Dr Robert Clarke (Lombardi Cancer Center, Georgetown University, Washington D.C. USA) (22). It is grown in cultured in RPMI-1640 supplemented with 10% fetal bovine serum (FBS) and 1% antibiotic-antimycotic solution. MCF7-Adr cell line, a MCF7-derived line resistant to adriamycin, has been previously characterized (23). Both MCF7-Adr and MDA-MB-231 (American Tissue Culture Collection, GA, USA) cells were grown in DMEM supplemented with 10% FBS and 1% antibiotic-antimycotic solution. Cells were maintained in a humidified 5% CO₂ atmosphere at 37°C. All cell culture reagents were purchased from WelGENE (Daejeon, South Korea). Restriction endonucleases and T4 DNA ligase were purchased from Fermentas MBI (Amherst, NY, USA). *Pfu* polymerase was obtained from Takara Bio Inc. (Shiga, Japan). Oligonucleotides were synthesized by GenoTech (Daejeon, South Korea). Other molecular biology reagents were purchased from Sigma-Aldrich (St. Louis, MO, USA) unless stated otherwise.

Plasmid construction. The 3000 bp Prx5 promoter region was amplified by polymerase chain reaction (PCR) using the forward primer 5'-GAAGATCTGACAGCCGACGCAACTACAAATG-3' and reverse primer 5'-CCCAAGCTTTCGCTGTGCTGTCTGACCAGATGTC-3'. The PCR product was isolated by gel electrophoresis and digested with *Bgl*III and *Hind*III (sites underlined in the primers) to generate the sticky ends for ligation with the expression vector. The final product was subcloned into an upstream portion of the luciferase reporter gene in the pGL3-basic vector (Promega,

Madison, WI, USA). The resulting construct was designated as pGL3-p3000. A series of reporter gene plasmids containing various portion of the 3000 bp Prx5 promoter region were generated using pGL3-p3000 as a template. The forward PCR primer sequences used for each promoter-reporter construct are listed in a Table I. The number shown in the designation of each plasmid represents the approximate distance from the translation start site in the Prx5 gene. In addition, a pGL3-F500 construct which contains 500 bp upstream of the translation start site in the Prx5 gene was generated using pGL3-p3000 as template plasmid DNA with the following primers; 5'-GAAGATCTACATGCGAGCTCAGCAGATTGTGGG-3' (forward) and 5'-GGGAAGCTTCGCTCCACTGCGCAGGC GCGAGC-3' (reverse).

Site-directed mutagenesis. The promoter regions of three target genes, Ets-like gene 1 (*ELK1*), NF-E2-related factor 2 (*Nrf2*), and GATA-binding protein 1 (*GATA1*), were subjected to site-directed mutagenesis to study the function of each gene as potential Prx5 gene regulators. Site-directed mutagenesis was performed using Quikchange site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA) according to the manufacturer's instructions. Three constructs containing mutated sequences in the transcription binding sites were generated to prevent each transcription factor from binding to the specific region of the Prx5 promoter region using the primers listed in Table II. The mutated sites are underlined as follows: mtELK1 (AACCGGAAGT-AACCTTAAGT), mtNrf2 (CCGGAAGCGTCT-CCTTCCGCGTCT), and mtGATA1 (GGGATAC-GGGATTCAC).

Transfection and luciferase reporter assay. Luciferase transcription was quantitated by transient transfection of various pGL3 constructs containing the respective DNA sequences corresponding to the promoter regions of Prx5 gene as well as the firefly luciferase reporter gene. Cells were seeded at 4x10⁵ cells/well in a 6-well plate and the appropriate pGL3 plasmid (2 µg/well) was transfected into the cells 24 h after seeding using FuGENE HD transfection reagent (Roche, Basel, Switzerland) according to the manufacturer's instructions. The cells were cotransfected with a pRL-CMV plasmid (Promega) encoding Renilla luciferase to normalize luciferase activity. Luciferase activity 24 h after transfection was measured using a Dual Luciferase Assay system (Promega) according to the manufacturer's instructions with a Victor3™ (PerkinElmer, Waltham, MA, USA) in a 96-well plate format. Data are reported as the relative luciferase activity calculated by dividing firefly luciferase activity by Renilla luciferase activity.

Reverse transcription polymerase chain reaction (RT-PCR). Total cellular RNA was extracted and reverse transcribed as described above. The primers used for semi-quantification of the human-specific genes were as follows: GATA1, 5'-CTTTCA GGTGTACCCATTGCT-3' (forward) and 5'-TAGGTAGTG GCCTGTCCTGTC-3' (reverse); Prx5, 5'-ACGGTGCACTGA AGGAGAGT-3' (forward) and 5'-ACATTCAGGGCCTTCACT ATG-3' (reverse); and GAPDH, 5'-CCATCCAATCGGTAG TAGCG-3' (forward) and 5'-GTAACCCGTTGAACCCATT-3' (reverse). PCR was performed with a GeneAmp PCR system

수학과

HARNACK-TYPE INEQUALITIES FOR THE POROUS MEDIUM EQUATION ON A MANIFOLD WITH NON-NEGATIVE RICCI CURVATURE

JEONGWOOK CHANG

*Department of Mathematics Education
 Dankook University
 Jukjeon-Dong, Gyeonggi 448-701, Korea
 jchang@dankook.ac.kr*

JINHO LEE

*Department of Mathematics
 Sookmyung Women's University
 Seoul 140-742, Korea
 jhlee@sookmyung.ac.kr*

Received 18 December 2010

Accepted 1 June 2011

Published 23 February 2012

We derive Harnack-type inequalities for non-negative solutions of the porous medium equation on a complete Riemannian manifold with non-negative Ricci curvature. Along with gradient estimates, reparametrization of a geodesic and time rescaling of a solution are key tools to get the results.

Keywords: Porous medium equation; Harnack inequality; geodesic; Ricci curvature.

Mathematics Subject Classification 2010: 53C21, 58J35

1. Introduction

Harnack inequalities are well known not only for the elliptic equations but also for parabolic equations. For the heat equation

$$\frac{\partial}{\partial t} u = \Delta u, \quad t > 0,$$

it is known that a positive solution satisfies the inequality

$$u(x_2, t_2) \geq \left(\frac{t_1}{t_2}\right)^{\frac{n}{2}} u(x_1, t_1) \exp\left(-\frac{\|x_2 - x_1\|^2}{4(t_2 - t_1)}\right), \quad (1.1)$$

where $0 < t_1 < t_2$, and $x_i \in \mathbb{R}^n$. In the case of complete Riemannian manifolds with non-negative Ricci curvature, a nice extension of (1.1) was studied by Li and Yau [6].

Furthermore, there has been important works on Harnack inequalities for various types of nonlinear parabolic equations (e.g. [3–5]). Di Benedetto studied Harnack inequalities for solutions of degenerate parabolic equations (see [4]), especially he showed the following intrinsic Harnack-type inequality for the porous medium equation.

Theorem 1.1 ([4]). *For an open subset $\Omega \subset \mathbb{R}^n$, let $u \geq 0$ be any local weak solution of the porous medium equation $u_t(x, t) = \Delta u^m(x, t)$ in $\Omega \times (0, T]$. Assume that $u(\bar{x}, \bar{t}) > 0$ for some $(\bar{x}, \bar{t}) \in \Omega \times (0, T]$. Then there exist two constants C_0, C_1 depending only upon m and n such that*

$$u(\bar{x}, \bar{t}) \leq C_0 \inf_{x \in B_R(\bar{x})} u(x, \bar{t} + \theta),$$

where

$$\theta = \frac{C_1 R^2}{[u(\bar{x}, \bar{t})]^{m-1}},$$

provided $B_{2R}(\bar{x}) \times (\bar{t} - \theta, \bar{t} + \theta)$ is contained in $\Omega \times (0, T]$.

Also, Auchmuty and Bao studied pointwise bounds of solutions of parabolic equations which could be degenerate (see [3]). Precisely, they showed that for positive solutions of the porous medium equation

$$\frac{\partial u}{\partial t} = \Delta(u^M), \quad t > 0, \quad x \in \mathbb{R}^n, \quad M > \max\left\{0, 1 - \frac{2}{n}\right\}, \quad M \neq 1,$$

the following estimate holds:

$$f(x_2, t_2) \geq \left(\frac{t_1}{t_2}\right)^\mu \left[f(x_1, t_1) - \frac{\delta}{4} \frac{|x_2 - x_1|^2}{t_2^\delta - t_1^\delta} \frac{1}{t_1^\mu} \right],$$

where $f = \left(\frac{M}{M-1}\right) u^{M-1}$, $\mu = \frac{M-1}{M-1+\frac{2}{n}}$, $\delta = 1 - \mu$.

In this paper, we mainly deal with intrinsic Harnack-type inequalities for the porous medium equation on a complete Riemannian manifold with non-negative Ricci curvature. Especially, we extend Theorem 1.1 to the case of Riemannian manifold with non-negative Ricci curvature by using gradient estimates similar to [3, 6].

We refer [1] for detail on the porous medium equation. Also throughout this paper, M is assumed to be an n -dimensional complete Riemannian manifold with the non-negative Ricci curvature unless otherwise stated. Also, we denote ρ , ∇ , and Δ by the Riemannian distance function, the Riemannian connection, and the Laplace–Beltrami operator on M , respectively.

RESEARCH

Open Access

Some properties of Chebyshev polynomials

Seon-Hong Kim*

*Correspondence:
shkim17@sookmyung.ac.kr
Department of Mathematics,
Sookmyung Women's University,
Seoul, 140-742, Korea

Abstract

In this paper we obtain some new bounds for Chebyshev polynomials and their analogues. They lead to the results about zero distributions of certain sums of Chebyshev polynomials and their analogues. Also we get an interesting property about the integrals of certain sums of Chebyshev polynomials.

Keywords: Chebyshev polynomials; bounds; sums; zeros

1 Introduction

Let $T_n(x)$ and $U_n(x)$ be the Chebyshev polynomials of the first kind and of the second kind, respectively. These polynomials satisfy the recurrence relations

$$\begin{aligned} T_0(x) &= 1, & T_1(x) &= x, & T_{n+1}(x) &= 2xT_n(x) - T_{n-1}(x) \quad (n \geq 1) \\ U_0(x) &= 1, & U_1(x) &= 2x, & U_{n+1}(x) &= 2xU_n(x) - U_{n-1}(x) \quad (n \geq 1). \end{aligned} \quad (1)$$

Chebyshev polynomials are of great importance in many areas of mathematics, particularly approximation theory. Many papers and books [3, 4] have been written about these polynomials. Chebyshev polynomials defined on $[-1, 1]$ are well understood, but the polynomials of complex arguments are less so. Reported here are several bounds for Chebyshev polynomials defined on $\mathbb{C}$ including zero distributions of certain sums of Chebyshev polynomials. Moreover, we will introduce certain analogues of Chebyshev polynomials and study their properties. Also we get an interesting property about the integrals of certain sums of Chebyshev polynomials.

Other generalized Chebyshev polynomials (known as Shabat polynomials) have been introduced in [5] and they are studied in the theory of graphs on surfaces and curves over number fields. For a survey in this area, see [6].

For $n \geq 1$ and $2 - \epsilon > 1 + \epsilon \geq 1$, i.e., $0 \leq \epsilon < 1/2$, we let

$$\begin{aligned} T_{n,\epsilon}(z) &:= (1 + \epsilon)(T_n(z) - T_n(0)) + T_n(0), \\ U_{n,\epsilon}(z) &:= (2 - \epsilon)(U_n(z) - U_n(0)) + U_n(0). \end{aligned}$$

In detail,

$$\begin{aligned} T_n(0) = U_n(0) &= \begin{cases} 0 & \text{if } n \text{ is odd,} \\ (-1)^{n/2} & \text{if } n \text{ is even,} \end{cases} \\ T_{0,\epsilon}(z) = U_{0,\epsilon}(z) &= T_0(z) = U_0(z) = 1, \end{aligned}$$

and it is easy to show that for an odd integer n ,

$$\begin{aligned}T_{n+1,\epsilon}(z) &= 2zT_{n,\epsilon}(z) - T_{n-1,\epsilon}(z), \\U_{n+1,\epsilon}(z) &= 2zU_{n,\epsilon}(z) - U_{n-1,\epsilon}(z).\end{aligned}$$

Since $T_{n,0}(z) = T_n(z)$, we may apply results about $T_{n,\epsilon}(z)$ to those about $T_n(z)$. But $U_{n,1}(z) = U_n(z)$ and ϵ was a nonnegative real number less than $1/2$, and so properties of $U_n(z)$ will be investigated separately from those of $U_{n,\epsilon}(z)$.

2 New results

In this section we list some new results related to the bounds of Chebyshev polynomials $T_n(z)$, $U_n(z)$ and their analogues $T_{n,\epsilon}(z)$, $U_{n,\epsilon}(z)$ defined on $\mathbb{C}$ including zero distributions of certain sums of Chebyshev polynomials and their analogues. And we will get an interesting property about the integrals of certain sums of Chebyshev polynomials. We first begin with properties about bounds of $T_n(z)$, $U_n(z)$, $T_{n,\epsilon}(z)$ and $U_{n,\epsilon}(z)$. We may compute that for $n \geq 1$,

$$T_{n,\epsilon}(z) = \begin{cases} (1+\epsilon)T_n(z) & \text{if } n \text{ is odd,} \\ (1+\epsilon)T_n(z) - \epsilon & \text{if } n \text{ is even and } 4 \mid n, \\ (1+\epsilon)T_n(z) + \epsilon & \text{if } n \text{ is even and } 4 \nmid n, \end{cases}$$

and

$$U_{n,\epsilon}(z) = \begin{cases} (2-\epsilon)U_n(z) & \text{if } n \text{ is odd,} \\ (2-\epsilon)U_n(z) - 1 + \epsilon & \text{if } n \text{ is even and } 4 \mid n, \\ (2-\epsilon)U_n(z) + 1 - \epsilon & \text{if } n \text{ is even and } 4 \nmid n. \end{cases}$$

Proposition 1 Suppose that z is a complex number satisfying $|z| \geq 1$. Then for $n \geq 1$,

$$|U_n(z)| - |U_{n-1}(z)| \geq 1$$

and

$$|U_n(z)| - |T_n(z)| \geq 1. \quad (2)$$

Also

$$|U_{n,\epsilon}(z)| - |U_{n-1,\epsilon}(z)| \geq 1 \quad (3)$$

and

$$|U_{n,\epsilon}(z)| - |T_{n,\epsilon}(z)| > \begin{cases} 1+\epsilon & \text{if } n \text{ is odd,} \\ \epsilon & \text{if } n \text{ is even.} \end{cases} \quad (4)$$

Proposition 1 will be used in the proofs of Theorems 4 and 6.

On Two Sequences of Polynomials Satisfying Certain Recurrence

Seon-Hong Kim[†]

Abstract

Bae and Kim displayed a sequence of 4th degree self-reciprocal polynomials whose maximal zeros are related in a very nice and far from obvious way. Kim showed that the auxiliary polynomials in their results are related to Chebyshev polynomials. In this paper, we study two sequences of polynomials satisfying the recurrence of the auxiliary polynomials with generalized initial conditions. We obtain same results with the auxiliary polynomials from a sequence, and some interesting conjectural properties about resultants and discriminants from another sequence.

Key words : Polynomials, Chebyshev Polynomials, Sequences

1. Introduction

Bae and Kim^[2] displayed a sequence of 4th degree self-reciprocal polynomials whose maximal zeros are related in a very nice and far from obvious way. In fact, they proved

Theorem 1 (Bae and Kim). For each real number $n > 6$, there is a sequence $\{p_k(n, z)\}_{k=1}^{\infty}$ of fourth degree self-reciprocal polynomials such that the zeros of $p_k(n, z)$ are all simple and real, and every $p_{k+1}(n, z)$ has the largest (in modulus) zero $\alpha\beta$ where α and β are the first and the second largest (in modulus) zeros of $p_k(n, z)$, respectively. One such sequence is given by $p_k(n, z)$ so that

$$p_k(n, z) = z^4 - q_{k-1}(n)z^3 + (q_k(n) + 2)z^2 - q_{k-1}(n)z + 1,$$

where $q_0(n) = 1$ and other $q_k(n)$'s are polynomials in n defined by the severely nonlinear recurrence

$$4q_{2m-1}(n) = q_{2m-2}^2(n) - (4n+1) \prod_{j=0}^{m-2} q_{2j}^2(n)$$

$$4q_{2m}(n) = q_{2m-1}^2(n) - (n-2)(n-6) \prod_{j=0}^{m-2} q_{2j+1}^2(n)$$

for $m \geq 1$, with the usual empty product conventions,

Department of Mathematics, Sookmyung Women's University, Seoul, 140-742 Korea

[†]Corresponding author : shkim17@sookmyung.ac.kr
(Received : June 15, 2012, Revised : June 20, 2012,
Accepted : June 24, 2012)

$$\text{i.e. } \prod_{j=0}^{-1} b_j = 1.$$

For notational convenience, write q_k instead of $q_k(n)$ in the theorem. The auxiliary polynomials q_k that parametrize their coefficients are of significant independent interest and they satisfy good recurrence relations (see Lemma 3 (i) of^[2])

$$q_{k+1} = q_{k-1}^2 - 2q_k - 4, \quad k \geq 1 \quad (1)$$

Kim^[3] showed the polynomials q_k are some sort of Chebyshev polynomials.

Proposition 2 (Kim). Define

$$H(n) := 2T_{2^n}\left(\frac{1}{2}\sqrt{4-x}\right),$$

where $T_n(x)$ is the n th degree Chebyshev polynomial of the first kind. Then the sequence

$$H(1)^2, 2H(2), H(2)^2, 2H(3), H(3)^2, 2H(4), \dots \quad (2)$$

comes from the recurrence

$$Q_{k+1} = Q_{k-1}^2 - 2Q_k - 4 \quad (3)$$

with initial conditions

$$\{Q_0, Q_1, Q_2\} = \{0, -x, 2x-4\} \quad (4)$$

where the sequence (2) begins with Q_3 .

The recurrence (3) for polynomials Q_k is in the same form as (1) for polynomials q_k . Proposition 2 asserts that the polynomials q_k are some sort of Chebyshev polynomials. In this paper, we study two sequences $\{Q_k\}$ of polynomials satisfying the recurrence (3) from different generalized initial conditions

$$\{Q_0, Q_1, Q_2\} = \left\{0, -x + a - \frac{3}{2}, 2x - 2a - 1\right\} \quad (5)$$

and

$$\{Q_0, Q_1, Q_2\} = \{0, -x, 2x - 2a - 1\} \quad (6)$$

where the sequence $\{Q_k\}$ from the latter initial conditions begins with Q_1 . In (5), we see that $Q_2 = Q_0^2 - 2Q_1 - 4$, but $Q_2 \neq Q_0^2 - 2Q_1 - 4$ in (6). However, in both, $Q_{k+1} = Q_k^2 - 2Q_{k-1} - 4$ if $k \geq 3$, and for $a = 3/2$, (5) is the same as (4) and Q_0, Q_2 in (6) are same as those in (4). The sequence $\{Q_k\}$ from (5) will deduce same results with Proposition 2 by replacing $H(n)$ in Proposition 2 with $H(n) = 2T_{2^n}\left(\frac{1}{2}\sqrt{a + \frac{5}{2} - x}\right)$. The proof of this easily follows from that of Proposition 2. We will provide it in Proposition 3 in Section 2.

A resultant is a scalar function of two polynomials which is non-zero if and only if the polynomials are relatively prime. The theory of resultants is an old and much studied topic in what used to be called the theory of equations. The resultant of two polynomials is in general a rather complicated function of their coefficients. However some polynomials such as any two cyclotomic polynomials have elegant formulas^[1]. The discriminant of a polynomial can be computed using resultants. More specifically,

$$Disc_x(p(x)) = \frac{(-1)^{n(n-1)/2}}{a_n} Res_x(p(x), p'(x))$$

where $\deg p(x) = n$ and a_n is the leading coefficient of $p(x)$. It seems that the sequence $\{Q_k\}$ from (6) has very interesting properties about factorizations and zero distributions. We will obtain some interesting conjectural properties about resultants and discriminants in Section 2.

2. Two Sequences of Polynomials from Different Initial Conditions

We begin with the sequence $\{Q_k\}$ from (5).

Proposition 3. Define

$$H(n) := 2T_{2^n}\left(\frac{1}{2}\sqrt{a + \frac{5}{2} - x}\right),$$

where $T_n(x)$ is the n th degree Chebyshev polynomial of the first kind. Then

$$H(n+1) = H(n)^2 - 2$$

and the sequence

$$H(1)^2, 2H(2), H(2)^2, 2H(3), H(3)^2, 2H(4), \dots \quad (7)$$

comes from the recurrence

$$Q_{k+1} = Q_k^2 - 2Q_{k-1} - 4$$

with initial conditions

$$\{Q_0, Q_1, Q_2\} = \{0, -x + a - 3/2, 2x - 2a - 1\}$$

where the sequence (7) begins with Q_3 .

Proof. Using an identity

$$T_{2n}(x) = 2T_n^2(x) - 1$$

followed from double angle formula, we may compute

$$H(n+1) = H(n)^2 - 2 \quad (8)$$

We use induction on $k \geq 3$. For $k = 3$,

$$\begin{aligned} Q_3 &= Q_1^2 - 2Q_2 - 4 \\ &= \left(-x + a - \frac{3}{2}\right)^2 - 2(2x - 2a - 1) - 4 \\ &= \frac{1}{4}(2x - 2a - 1)^2 \end{aligned}$$

and

$$\begin{aligned} H(1) &= 2T_2\left(\frac{1}{2}\sqrt{a + \frac{5}{2} - x}\right) \\ &= 2\left[2\left(\frac{1}{2}\sqrt{a + \frac{5}{2} - x}\right)^2 - 1\right] \\ &= \frac{1}{2}(-2x + 2a + 1) \end{aligned}$$

ON SOME COMBINATIONS OF SELF-RECIPROCAL POLYNOMIALS

SEON-HONG KIM

ABSTRACT. Let $\mathcal{P}_n$ be the set of all monic integral self-reciprocal polynomials of degree n whose all zeros lie on the unit circle. In this paper we study the following question: For $P(z), Q(z) \in \mathcal{P}_n$, does there exist a continuous mapping $r \rightarrow G_r(z) \in \mathcal{P}_n$ on $[0, 1]$ such that $G_0(z) = P(z)$ and $G_1(z) = Q(z)$?

1. Introduction

Throughout this paper, U denotes the unit circle and n is a positive integer. If all zeros of $P(z) \in \mathbb{R}[z]$ of degree n lie on U , it is well known that $P(z) = \pm z^n P(1/z)$, and $P(z)$ is called self-reciprocal when $P(z) = z^n P(1/z)$. In what follows we denote by $\mathcal{P}_n$ the set of all monic integral self-reciprocal polynomials of degree n whose all zeros lie on U . Our basic goal in this paper is the study of the following question.

Question A. For given polynomials $P(z), Q(z) \in \mathcal{P}_n$, does there exist a continuous mapping $r \rightarrow G_r(z) \in \mathcal{P}_n$ on $[0, 1]$ such that

$$G_0(z) = P(z), \quad G_1(z) = Q(z)?$$

The condition $P(z)$ and $Q(z)$ self-reciprocal in Question A seems to be necessary because $z + 1$ and $z - 1$ are the only monic integral polynomials of degree 1 with all zeros on U and so no $G_r(z)$ in Question A exists. One may ask naturally whether $G_r(z)$ in Question A is the convex combination of $P(z)$ and $Q(z)$, that is $G_r(z) = (1 - r)P(z) + rQ(z)$ where $0 \leq r \leq 1$. However $G_r(z)$ is not always the convex combination, for example the polynomials

$$\begin{aligned} P(z) &= (z^2 - z + 1)(z^4 + 1)(z^6 + z^5 + z^4 + z^3 + z^2 + z + 1), \\ Q(z) &= (z^4 - z^3 + z^2 - z + 1)(z^8 - z^6 + z^4 - z^2 + 1) \end{aligned}$$

Received October 15, 2010.

2010 *Mathematics Subject Classification.* Primary 30C15; Secondary 26C10.

Key words and phrases. convex combination, polynomials, self-reciprocal polynomials, unit circle, zeros.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (No.2009-0075759).

have all their zeros on U , but $(P(z) + Q(z))/2$ has four zeros of modulus $1.15963\dots$. This yields a question about the existence of $G_r(z)$ which is not the convex combination. If Question A is true, we will write

$$P \sim Q,$$

and specially if $G_r(z)$ is the convex combination, we will write

$$P \hookrightarrow Q.$$

Preliminary attempts using computer algebra lead to the various unproved examples including

$$\frac{z^{2n+1} - 1}{z - 1} \hookrightarrow \left(\frac{z^{n+1} - 1}{z - 1} \right)^2$$

and

$$\left(\frac{z^n - 1}{z - 1} \right)^3 \hookrightarrow z^{3(n-1)} + z^{3(n-1)-1} + \dots + 1.$$

And Kim [2] showed

$$(z^{2n} + 1)^2 \sim (z^2 - 1)(z^{4n-2} - 1)$$

for some $G_r(z)$ that is not the convex combination. But it seems to be hard to find such $G_r(z)$.

In Section 2, we will give examples of self-reciprocal polynomials $P(z)$ and $Q(z)$ satisfying $P \sim Q$ in different two ways and one of them is $P \hookrightarrow Q$. This generalizes Theorem 1 of [2]. In Section 3, we will consider self-reciprocal polynomials that are products of finite geometric series. Let

$$F_{a_1, \dots, a_u}(z) = \prod_{j=1}^u \frac{z^{a_j} - 1}{z - 1},$$

$$\mathcal{F}_{u,n} = \left\{ F_{a_1, \dots, a_u}(z) : \text{all } a_j \in \mathbb{N} - \{1\} \ (1 \leq j \leq u) \text{ are distinct, } \sum_{j=1}^u a_j = n \right\}.$$

We will prove a conjecture given in [2] that $f_1 \hookrightarrow f_2$ for any $f_1, f_2 \in \mathcal{F}_{2,n}$. In case of $u \geq 3$, not all $g_1, g_2 \in \mathcal{F}_{u,n}$ satisfy $g_1 \hookrightarrow g_2$. But we will see that, using the case $u = 2$, for any $g_1, g_3 \in \mathcal{F}_{3,n}$, there exists $g_2 \in \mathcal{F}_{3,n}$ such that $g_1 \hookrightarrow g_2 \hookrightarrow g_3$, and such g_2 is not unique. Also we will study how to show $g_1 \hookrightarrow g_2$ where $g_1, g_2 \in \mathcal{F}_{3,n}$ are specifically given. The main tool for this is Fell's lemma [1]. Finally in Section 4, we will discuss whether " $\sim$ " is an equivalence relation or not over the set $\mathcal{F}_{u,n}$. It is obvious that " $\sim$ " is reflexive and symmetric. But we don't give an answer for transitivity but at least we will see that a natural candidate for this does not work.

EIGENVALUE ESTIMATES FOR STABLE MINIMAL HYPERSURFACES

KEOMKYO SEO

(Communicated by H. Martini)

Abstract. In this article we provide estimates on the first eigenvalue of stable minimal hypersurfaces in a Riemannian manifold with sectional curvature bounded from below and above by negative constants. We also obtain a lower bound of the total scalar curvature of a stable minimal hypersurface if the scalar curvature of the ambient space is positive.

1. Introduction

As it is well known, the first Dirichlet eigenvalue of a Riemannian manifold Σ with boundary is characterized as

$$\lambda_1 = \inf_f \frac{\int_{\Sigma} |\nabla f|^2}{\int_{\Sigma} f^2},$$

where the infimum is taken over all piecewise smooth functions in Σ vanishing on the boundary $\partial\Sigma$. Recently Candel [1] gave an upper bound for the first eigenvalue of a stable minimal surface in $\mathbb{H}^3$. We say that a minimal hypersurface Σ in an $(n+1)$ -dimensional Riemannian manifold M is *stable* if the second variation of its volume is always nonnegative for every compactly supported deformation of Σ in M^{n+1} . More precisely, an n -dimensional minimal hypersurface Σ in a Riemannian manifold M is called *stable* if for any compactly supported Lipschitz function f on M

$$\int_{\Sigma} |\nabla f|^2 - (\overline{\text{Ric}}(e_{n+1}) + |A|^2)f^2 \geq 0 \quad (1.1)$$

holds, where $\overline{\text{Ric}}(e_{n+1})$ is the Ricci curvature of M in the direction of e_{n+1} , e_{n+1} is the unit normal vector of Σ in M , and $|A|^2$ is the squared length of the second fundamental form of Σ .

Recall that the Yamabe invariant of the conformal class $[g]$ of an n -dimensional Riemannian manifold Σ is defined by

$$Y(g) = \inf\{E(\bar{g}) : \bar{g} = u(x)^{\frac{2n}{n-2}}g, u(x) > 0, u \in H^1(M)\},$$

Mathematics subject classification (2010): 53C21, 53A10.

Keywords and phrases: Stability, minimal hypersurface, first eigenvalue, Yamabe invariant.

This research was supported by the Sookmyung Women's University Research Grants 2009.

where $E(\bar{g})$ is defined by

$$E(\bar{g}) = \frac{\int_{\Sigma} \left(\frac{4(n-1)}{n-2} |\nabla u|^2 + R_g u^2 \right) dv_g}{\left(\int_{\Sigma} u^{\frac{2n}{n-2}} dv_g \right)^{\frac{n-2}{n}}}.$$

Here R_g and dv_g denote the scalar curvature and the volume form of the metric g on Σ , respectively. Then the *Yamabe invariant* of Σ is defined by

$$\sigma(\Sigma) = \sup\{Y(g) : g \text{ is a smooth metric on } \Sigma\}.$$

Ho [4] generalized Candel's result to higher-dimensional cases. He gave estimates on the first eigenvalue of a stable minimal hypersurface Σ in hyperbolic space $\mathbb{H}^{n+1}$ for $n \geq 3$ under the assumption that the Yamabe invariant $\sigma(\Sigma)$ satisfies that $\sigma(\Sigma) < 0$. Actually he proved

THEOREM 1.1. *Let $\Sigma \subset \mathbb{H}^{n+1}$ be a compact stable minimal hypersurface with boundary. If $\sigma(\Sigma) < 0$, then the first eigenvalue of Σ satisfies*

$$\frac{1}{4}(n-1)^2 \leq \lambda_1(\Sigma) < \frac{n^2(n-2)}{7n-6}.$$

In this paper we extend Ho's result to stable minimal hypersurfaces in a Riemannian manifold of variable curvature. We provide the upper bound of the first eigenvalue when the ambient space has negative scalar curvature. More precisely, we prove

THEOREM 1.2. *Let M be an $(n+1)$ -dimensional Riemannian manifold with scalar curvature S satisfying that $-S_0 \leq S < 0$ for some positive constant S_0 , $n \geq 3$. Let $\Sigma \subset M$ be a compact stable minimal hypersurface with boundary. If $\sigma(\Sigma) < 0$, then the first eigenvalue of Σ satisfies*

$$\lambda_1(\Sigma) < \frac{S_0(n-2)}{2n}.$$

If the ambient space M has sectional curvature bounded from below and above by negative constants, we can extend Theorem 1.1 as follows.

THEOREM 1.3. *Let M be an $(n+1)$ -dimensional Riemannian manifold with sectional curvature K_M satisfying $-b \leq K_M \leq -a < 0$ for some positive constants a and b , $n \geq 3$. Let $\Sigma \subset M$ be a compact stable minimal hypersurface with boundary. Assume that $\sigma(\Sigma) < C_\sigma$ for some constant $C_\sigma \in [0, \frac{3n-2}{n-2} \frac{1}{C_s})$, where C_s is a Sobolev constant in [5]. Then the first eigenvalue of Σ satisfies*

$$\frac{(n-1)^2}{4}a \leq \lambda_1(\Sigma) < \frac{n(n+1)b - na}{\frac{3n-2}{n-2} - C_\sigma C_s}.$$

When the scalar curvature of the ambient space is positive, we estimate the total scalar curvature of a stable minimal hypersurface Σ and the Yamabe invariant $\sigma(\Sigma)$ of Σ .

Stable minimal hypersurfaces in a Riemannian manifold with pinched negative sectional curvature

Nguyen Thac Dung · Keomkyo Seo

Received: 4 June 2011 / Accepted: 2 September 2011 / Published online: 17 September 2011
© Springer Science+Business Media B.V. 2011

Abstract We give an estimate of the smallest spectral value of the Laplace operator on a complete noncompact stable minimal hypersurface M in a complete simply connected Riemannian manifold with pinched negative sectional curvature. In the same ambient space, we prove that if a complete minimal hypersurface M has sufficiently small total scalar curvature then M has only one end. We also obtain a vanishing theorem for L^2 harmonic 1-forms on minimal hypersurfaces in a Riemannian manifold with sectional curvature bounded below by a negative constant. Moreover, we provide sufficient conditions for a minimal hypersurface in a Riemannian manifold with nonpositive sectional curvature to be stable.

Keywords Minimal hypersurface · Stability · First eigenvalue

Mathematics Subject Classification (2000) 53C42 · 58C40

1 Introduction

Let M be a complete noncompact Riemannian manifold and let Ω be a compact domain in M . Let $\lambda_1(\Omega) > 0$ denote the first eigenvalue of the Dirichlet boundary value problem

$$\begin{cases} \Delta f + \lambda f = 0 & \text{in } \Omega \\ f = 0 & \text{on } \partial\Omega \end{cases}$$

N. T. Dung
Department of Mathematics, National Tsinghua University,
No. 101, Sec. 2, Kuangfu Road, Hsinchu, Taiwan, R.O.C.
e-mail: dungmath@yahoo.co.uk

K. Seo (✉)
Department of Mathematics, Sookmyung Women's University,
Hyochoangwongil 52, Yongsan-ku, Seoul 140-742, Korea
e-mail: kseo@sookmyung.ac.kr

where Δ denotes the Laplace operator on M . Then, by the domain monotonicity principle, the first eigenvalue $\lambda_1(M)$ of a complete noncompact manifold M is defined by

$$\lambda_1(M) = \inf_{\Omega} \lambda_1(\Omega),$$

where the infimum is taken over all compact domains in M .

In this article, we shall find an estimate of the smallest spectral value of the Laplace operator on minimal hypersurfaces. Let us briefly introduce previous results in this direction. In [4], Cheung and Leung obtained the first eigenvalue estimate for a complete noncompact submanifold with bounded mean curvature in hyperbolic space. In particular, they proved that for a n -dimensional complete minimal submanifold M in the m -dimensional hyperbolic space $\mathbb{H}^m$

$$\frac{1}{4}(n-1)^2 \leq \lambda_1(M).$$

Here this inequality is sharp because equality holds when M is totally geodesic [7]. Bessa and Montenegro [1] extended this result to complete noncompact submanifolds in a complete simply connected Riemannian manifold with sectional curvature bounded above by a negative constant. Indeed they proved

Theorem 1 ([1]) *Let N be an n -dimensional complete simply connected Riemannian manifold with sectional curvature K_N satisfying $K_N \leq -a^2 < 0$ for a positive constant $a > 0$. Let M be a m -dimensional complete noncompact submanifold with bounded mean curvature H in N satisfying $|H| \leq b < (m-1)a$. Then*

$$\frac{[(m-1)a-b]^2}{4} \leq \lambda_1(M).$$

Candel [2] gave an upper bound for the smallest spectral value of the universal cover of a complete stable minimal surface in the 3-dimensional hyperbolic space $\mathbb{H}^3$. More precisely, it was proved

Theorem 2 ([2]) *Let M be a complete simply connected stable minimal surface in $\mathbb{H}^3$. Then the first eigenvalue of M satisfies*

$$\frac{1}{4} \leq \lambda_1(M) \leq \frac{4}{3}.$$

Recall that an n -dimensional minimal hypersurface M in a Riemannian manifold N is called *stable* if it holds that for any compactly supported Lipschitz function f on M

$$\int_{\Sigma} |\nabla f|^2 - (\overline{\text{Ric}}(e_{n+1}) + |A|^2) f^2 \geq 0,$$

where e_{n+1} is the unit normal vector of M in N , $\overline{\text{Ric}}(e_{n+1})$ is the Ricci curvature of N in the direction of e_{n+1} , and $|A|$ is the length of the second fundamental form of Σ . Recently, the second author [13] extended the above theorem to higher-dimensional cases as follows:

Theorem 3 ([13]) *Let M be a complete stable minimal hypersurface in $\mathbb{H}^{n+1}$ with finite L^2 -norm of the second fundamental form A (i.e., $\int_M |A|^2 dv < \infty$). Then we have*

$$\frac{(n-1)^2}{4} \leq \lambda_1(M) \leq n^2.$$

Isoperimetric inequalities for submanifolds with bounded mean curvature

Keomkyo Seo

Received: 18 October 2010 / Accepted: 8 July 2011
© Springer-Verlag 2011

Abstract In this paper, we provide various Sobolev-type inequalities for smooth nonnegative functions with compact support on a submanifold with variable mean curvature in a Riemannian manifold whose sectional curvature is bounded above by a constant. We further obtain the corresponding linear isoperimetric inequalities involving mean curvature. We also provide various first Dirichlet eigenvalue estimates for submanifolds with bounded mean curvature.

Keywords Isoperimetric inequality · Sobolev-type inequality · First eigenvalue · Mean curvature

Mathematics Subject Classification (2000) 58E35 · 58C40

1 Introduction

The classical isoperimetric inequality says that for a given bounded domain D in $\mathbb{R}^n$, if ω_n is the volume of a unit ball in $\mathbb{R}^n$, then

$$n^n \omega_n \text{Vol}(D)^{n-1} \leq \text{Vol}(\partial D)^n$$

and equality holds if and only if D is a ball. As an extension of the above classical isoperimetric inequality, it is conjectured that any n -dimensional compact minimal submanifold M of $\mathbb{R}^m$ still satisfies the above inequality. Moreover, equality holds

Communicated by Andreas Cap.

This research was supported by the Sookmyung Women's University Research Grants 2011.

K. Seo (✉)
Department of Mathematics, Sookmyung Women's University,
Hyochoangwongil 52, Yongsan-ku, Seoul 140-742, Korea
e-mail: kseo@sookmyung.ac.kr

Published online: 23 July 2011

 Springer

if and only if M is an n -dimensional ball. When $n = 2$, there are several partial results ([9, 16, 18], see also [15] for more references). In particular, for any minimal surface Σ in $\mathbb{R}^n$ with at most two boundary components, Σ satisfies the following sharp isoperimetric inequality (see [6, 12]):

$$4\pi \text{Area}(\Sigma) \leq \text{Length}(\partial \Sigma)^2,$$

where equality holds if and only if Σ is a disk. When $n \geq 3$, the only known result was due to Almgren [2] who proved that the above conjecture is true for n -dimensional area-minimizing submanifolds in $\mathbb{R}^m$. Even though the conjecture is still open, one may try to extend the classical isoperimetric inequality to more general submanifold with variable mean curvature vector in a Riemannian manifold. In this direction, Allard [1] proved an isoperimetric inequality which involves mean curvature term. More precisely, he showed that given n -dimensional compact submanifold M in $\mathbb{R}^m$,

$$c(n) \text{Vol}(M)^{n-1} \leq \left(\text{Vol}(\partial M) + \int_M |H| \right)^n, \quad (1.1)$$

where $c(n)$ is a constant which depends only on n and H is the mean curvature vector of M in $\mathbb{R}^m$. In [14], Michael and Simon also obtained the similar isoperimetric inequality by proving the corresponding Sobolev inequality on a submanifold in a Euclidean space. One year later, Hoffman and Spruck [10] generalized the inequality on a submanifold in a Riemannian manifold.

In this paper, we study linear isoperimetric inequalities and the first (Dirichlet) eigenvalue estimates for submanifolds with variable mean curvature in a Riemannian manifold whose sectional curvature is bounded above. In Sect. 2, we derive some useful inequalities on the Laplacian of functions of distance by virtue of Hessian comparison theorem. For minimal surface cases, such inequalities were obtained by Choe [7]. We modify his arguments to obtain more general inequalities (Lemmas 2.3–2.5) for submanifolds (not necessarily minimal) with variable mean curvature.

Let Σ be a domain in a hyperbolic space $\mathbb{H}^n$ or an n -dimensional minimal submanifold of $\mathbb{H}^m$. Then the following linear isoperimetric inequality holds [8, 19]:

$$(n-1)\text{Vol}(\Sigma) \leq \text{Vol}(\partial \Sigma).$$

In Sect. 3, we extend the above result and obtain more general linear isoperimetric inequalities for submanifolds in a complete simply connected Riemannian manifold with sectional curvature bounded above by a constant. We further obtain the corresponding Sobolev-type inequalities involving mean curvature for smooth nonnegative functions with compact support on submanifolds. In Sect. 4, we provide various first (Dirichlet) eigenvalue estimates for submanifolds with bounded mean curvature. The first (Dirichlet) eigenvalue of a Riemannian manifold Σ with boundary is characterized as

$$\lambda_1 = \inf_f \frac{\int_{\Sigma} |\nabla f|^2}{\int_{\Sigma} f^2},$$

Relative isoperimetric inequalities for minimal submanifolds outside a convex set

Keomkyo Seo*

Department of Mathematics, Sookmyung Women's University, Hyochangwongil 52, Yongsan-ku, Seoul, 140-742, Korea

Received 22 March 2011, revised 4 July 2011, accepted 5 December 2011

Published online 6 March 2012

Key words Isoperimetric inequality, minimal submanifold, convex set

MSC (2010) 58E35, 49Q20

Let C be a closed convex set in a complete simply connected Riemannian manifold M with sectional curvature bounded above by a positive constant K . Assume that Σ is a compact minimal surface outside C such that Σ is orthogonal to ∂C along $\partial\Sigma \cap \partial C$ and $\partial\Sigma \sim \partial C$ is radially connected from a point $p \in \partial\Sigma \cap \partial C$. We introduce a modified volume $M_p(\Sigma)$ of Σ and obtain a sharp isoperimetric inequality

$$2\pi M_p(\Sigma) \leq \text{Length}(\partial\Sigma \sim \partial C)^2,$$

where equality holds if and only if Σ is a geodesic half disk with constant Gaussian curvature K . We also prove higher dimensional isoperimetric inequalities for minimal submanifolds outside a closed convex set in a Riemannian manifold using the modified volume.

© 2012 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

1 Introduction

Let Σ be a domain in a complete simply connected surface with constant Gaussian curvature K . The classical isoperimetric inequality says that

$$4\pi \text{Area}(\Sigma) - K \text{Area}(\Sigma)^2 \leq \text{Length}(\partial\Sigma)^2,$$

where equality holds if and only if Σ is a geodesic disk. One natural way to extend this optimal inequality is to find the corresponding relative isoperimetric inequality. Let C be a closed convex set in a complete simply connected surface S with constant Gaussian curvature $K \leq 0$. It has been known that if Σ is a relatively compact subset in $S \sim C$, then

$$2\pi \text{Area}(\Sigma) - K \text{Area}(\Sigma)^2 \leq \text{Length}(\partial\Sigma \sim \partial C)^2, \quad (1.1)$$

where equality holds if and only if Σ is a geodesic half disk [2]. Here $\sim$ denotes the set minus operator. The inequality (1.1) is called the *relative isoperimetric inequality* for Σ . Recently the author [8] extended the inequality (1.1) for a relatively compact subset Σ in $S \sim C$ of Gaussian curvature bounded above by a nonpositive constant. Surprisingly, the relative isoperimetric inequality (1.1) still holds for a compact minimal surface Σ with partially free boundary in a Riemannian manifold M with sectional curvature bounded above by a nonpositive constant, if the relative boundary $\partial\Sigma \sim \partial C$ is radially connected [7], [9]. Note that a curve Γ is said to be *radially connected* from a point $p \in M$ if $\{\text{dist}(p, q) : q \in \Gamma\}$ is a connected interval. Although we do not know whether the inequality (1.1) holds for Σ in a Riemannian manifold with sectional curvature bounded above by a positive constant or not, we introduce a concept of the modified volume $M_p(\Sigma)$ to obtain the following isoperimetric

* e-mail: kseo@sookmyung.ac.kr, Phone: +82 2 2077 7465, Fax: +82 2 2077 7323

inequality (Theorem 2.6) without the additional curvature term:

$$2\pi M_p(\Sigma) \leq \text{Length}(\partial\Sigma \sim \partial C)^2.$$

Notice that there is no curvature term of the ambient space in the above inequality.

In Section 3, we study relative isoperimetric inequalities for an n -dimensional minimal submanifold Σ outside a closed convex set C in a Riemannian manifold with sectional curvature bounded above by a constant. Under the strong assumption that the relative boundary $\partial\Sigma \sim \partial C$ lies on a geodesic sphere centered at $p \in \partial\Sigma \cap \partial C$, we prove the following relative isoperimetric inequality (Theorem 3.2)

$$\frac{1}{2}n^n\omega_n M_p(\Sigma)^{n-1} \leq \text{Vol}(\partial\Sigma \sim \partial C)^n,$$

where ω_n is the volume of a unit ball in $\mathbb{R}^n$.

For the higher dimensional case, let Σ be a domain in hyperbolic space $\mathbb{H}^n$ or an n -dimensional minimal submanifold of $\mathbb{H}^n$. Then the following linear isoperimetric inequality is well-known [5], [10]:

$$(n-1)\text{Vol}(\Sigma) \leq \text{Vol}(\partial\Sigma).$$

However, it is not yet known whether this linear isoperimetric inequality holds for the ambient space with sectional curvature bounded above by a positive constant. Using the concept of the modified volume, we obtain the modified linear isoperimetric inequality (Theorem 3.3).

Let Γ be a submanifold in a Riemannian manifold M . Given $p \in M$, we define a *cone* $p \rtimes \Gamma$ from p over Γ to be the union of the geodesic segments from p to the points of Γ . Denote by $S_t(p)$ the geodesic sphere of radius t with center at p . Then, in case that Γ is a curve, the *angle* of a curve Γ viewed from p is defined as follows and denoted by $\text{Angle}(\Gamma, p)$

$$\text{Angle}(\Gamma, p) = \lim_{t \rightarrow 0} \frac{1}{t} \text{Length}((p \rtimes \Gamma) \cap S_t(p)).$$

2 Two-dimensional relative isoperimetric inequalities

Let p be a point in the m -dimensional sphere $\mathbb{S}^m \subset \mathbb{R}^{m+1}$ and let $r(x)$ be the distance from p to x in $\mathbb{S}^m$. Choe and Gulliver [5] defined the *modified volume* $M_p(\Sigma)$ of Σ with center at p as

$$M_p(\Sigma) = \int_{\Sigma} \cos r.$$

Similarly for Σ in the m -dimensional hyperbolic space $\mathbb{H}^m$, they defined the *modified volume* of Σ by

$$M_p(\Sigma) = \int_{\Sigma} \cosh r.$$

Using the concept of the modified volume, they were able to prove the isoperimetric inequalities for minimal submanifolds in $\mathbb{S}^m$ or $\mathbb{H}^m$. Motivated by this, we introduce the modified volume of a submanifold in a Riemannian manifold of variable curvature.

Definition 2.1 Let M be a complete simply connected Riemannian manifold with sectional curvature bounded above by a positive constant $K = k^2$ for $k > 0$. Let p be a point in M and let $r(x)$ be the distance from p to x in M . For a given n -dimensional submanifold Σ in M , the *modified volume* $M_p(\Sigma)$ of Σ with center at p is defined as

$$M_p(\Sigma) = \int_{\Sigma} \cos kr.$$

Similarly, when M is a complete simply connected Riemannian manifold with sectional curvature bounded above by a negative constant $K = -k^2$ for $k > 0$, we define the *modified volume* of Σ by

$$M_p(\Sigma) = \int_{\Sigma} \cosh kr.$$



Contents lists available at SciVerse ScienceDirect

Finite Fields and Their Applications

www.elsevier.com/locate/ffa



Delta sets for divisors supported in two points

Iwan M. Duursma^{a,*}, Seungkook Park^{b,1}

^a Department of Mathematics, University of Illinois at Urbana-Champaign, United States

^b Department of Mathematical Sciences, University of Cincinnati, United States

ARTICLE INFO

Article history:

Received 24 March 2009

Revised 18 June 2012

Accepted 26 June 2012

Available online 31 July 2012

Communicated by Michael Tsfasman

MSC:

94B27

14G50

94A62

94B05

Keywords:

Algebraic geometric code

Geometric Goppa code

Minimum distance bound

Order bound

Linear secret sharing scheme

Hermitian curve

Suzuki curve

ABSTRACT

In Duursma and Park (2010) [7], the authors formulate new coset bounds for algebraic geometric codes. The bounds give improved lower bounds for the minimum distance of algebraic geometric codes as well as improved thresholds for algebraic geometric linear secret sharing schemes. The bounds depend on the delta set of a coset and on the choice of a sequence of divisors inside the delta set. In this paper we give general properties of delta sets and we analyze sequences of divisors supported in two points on Hermitian and Suzuki curves.

© 2012 Elsevier Inc. All rights reserved.

0. Introduction

The best known lower bounds for the minimum distance of an algebraic geometric code are the floor bound and the order bound. In [7], the authors improve both the Lundell–McCullough floor bound [12] and the Beelen order bound [1]. The improvements are formulated separately but are such that the improved order bound is always at least equal to the improved floor bound. The order

* Corresponding author.

E-mail addresses: duursma@math.uiuc.edu (I.M. Duursma), skpark@sookmyung.ac.kr (S. Park).

¹ Current address: Department of Mathematics, Sookmyung Women's University, Republic of Korea.

bound considers a subcode of a code and gives lower bounds for the weights in a coset of the subcode. In [7], to each coset is associated a set of divisors called a delta set such that for any word in the coset and for any increasing sequence of divisors in the delta set the weight of the word is at least the length of the sequence.

This paper and the paper [7] were originally written together as one preprint [3]. The preprint contained improved bounds, that have now appeared in [7], as well as a discussion on how to apply the bounds effectively in concrete cases such as for Hermitian and Suzuki codes. This paper grew out of the second part and is meant to facilitate using the bounds in the setting of two-point codes. Our main tool to analyze sequences of divisors supported in two points is the notion of discrepancy, a special case of a delta set. We explicitly describe discrepancies for the Hermitian and the Suzuki curves and illustrate their role in establishing lower bounds for the minimum distance of two-point codes.

We follow the methods and notation introduced in [7]. In Section 1, we recall the main definitions and basic results. For full details we refer to Sections 2 to 4 on algebraic geometric codes, cosets of algebraic geometric codes, and semigroup ideals in [7]. The main theorem of [7] is the lower bound d_{DP} for the distance between cosets of a given algebraic geometric code. In Section 2 we recall the bound d_{DP} and basic properties of delta sets. In Section 3, we define discrepancies as a special type of delta set. In Section 4, we give two proofs, one due to Beelen [1] and Park [15], and one new, for lower bounds for the minimum distance of Hermitian two-point codes. The lower bounds meet the actual minimum distances in [10]. In Section 5, we determine discrepancies for Suzuki curves. In Section 6 we use discrepancies to compare delta sets that are ordered by inclusion, and we describe relations between delta sets and dual delta sets. Section 7 and Section 8 describe ways to construct delta sets and how to compare sequences inside a given delta set. In examples we use these methods to show how the bounds in [7] improve known bounds for Suzuki two-point codes. After the paper was written, the bound d_{DP} was improved in [4]. A general description of the semigroup ideal method that includes all known order bounds is presented in [5, Sections 5–8].

1. Algebraic geometric codes and semigroup ideals

Let $X/\mathbb{F}$ be an algebraic curve (absolutely irreducible, smooth, projective) of genus g over a finite field $\mathbb{F}$. Let $\mathbb{F}(X)$ be the function field of $X/\mathbb{F}$ and let $\Omega(X)$ be the module of rational differentials of $X/\mathbb{F}$. Given a divisor E on X defined over $\mathbb{F}$, let $L(E) = \{f \in \mathbb{F}(X) \setminus \{0\} : (f) + E \geq 0\} \cup \{0\}$ and let $\Omega(E) = \{\omega \in \Omega(X) \setminus \{0\} : (\omega) \geq E\} \cup \{0\}$. Let K represent the canonical divisor class. For n distinct rational points $P_1, \dots, P_n$ on X and for disjoint divisors $D = P_1 + \dots + P_n$ and G , the geometric Goppa codes $C_L(D, G)$ and $C_\Omega(D, G)$ are defined as the images of the maps

$$\begin{aligned} \alpha_L : L(G) &\longrightarrow \mathbb{F}^n, & f &\mapsto (f(P_1), \dots, f(P_n)), \\ \alpha_\Omega : \Omega(G - D) &\longrightarrow \mathbb{F}^n, & \omega &\mapsto (\text{res}_{P_1}(\omega), \dots, \text{res}_{P_n}(\omega)). \end{aligned}$$

The maps establish isomorphisms $L(G)/L(G - D) \simeq C_L(D, G)$ and $\Omega(G - D)/\Omega(G) \simeq C_\Omega(D, G)$. With the Residue theorem, the images are orthogonal subspaces of $\mathbb{F}^n$. With the Riemann–Roch theorem they are maximal orthogonal subspaces. The Goppa lower bound for the minimum distance of the codes is

$$\begin{aligned} d(C_L(D, G)) &\geq \max\{0, \deg(D - G)\}, \\ d(C_\Omega(D, G)) &\geq \max\{0, \deg(G - K)\}. \end{aligned}$$

To analyze the minimum distance of the codes we use the following characterization.

FINITELY GENERATED gr-MULTIPLICATION MODULES

SEUNGKOOK PARK*

ABSTRACT. In this paper, we investigate when gr-multiplication modules are finitely generated and show that if M is a finitely generated gr-multiplication R -module then there is a lattice isomorphism between the lattice of all graded ideals I of R containing $\text{ann}(M)$ and the lattice of all graded submodules of M .

1. Introduction

Let R be a commutative ring with identity $1 \neq 0$ and M a unital R -module. M is called a *multiplication module* provided for each submodule N of M , there exists an ideal I of R such that $N = IM$ [2]. Let G be a multiplicative group with identity e . A ring R is said to be a *graded ring of type G* if there is a family of additive subgroups of R , say $\{R_i \mid i \in G\}$, such that $R = \bigoplus_{i \in G} R_i$ and $R_i R_j \subseteq R_{ij}$ for all $i, j \in G$, where $R_i R_j$ is the set of all finite sums of products $r_i r_j$ with $r_i \in R_i$ and $r_j \in R_j$. The elements of $h(R) = \bigcup_{i \in G} R_i$ are called the homogeneous elements of R . Any nonzero $r \in R$ has a unique expression as a sum of homogeneous elements, that is, $r = \sum_{i \in G} r_i$ where r_i is nonzero for a finite number of i in G . The nonzero elements r_i in the decomposition of r are called the homogeneous components of r . Let R be a graded ring of type G . An R -module M is said to be a *graded R -module* if there is a family $\{M_i \mid i \in G\}$ of additive subgroups of M such that $M = \bigoplus_{i \in G} M_i$ and $R_i M_j \subseteq M_{ij}$ for all $i, j \in G$. Elements of $h(M) = \bigcup_{i \in G} M_i$ are called the homogeneous elements of M . A submodule N of M is a graded submodule if $N = \bigoplus_{i \in G} (N \cap M_i)$, or equivalently, if for any

Received August 03, 2012; Accepted October 10, 2012.

2010 Mathematics Subject Classification: Primary 16W50; Secondary 13A02.

Key words and phrases: finitely generated gr-multiplication module, multiplication module.

*This research was supported by the Sookmyung Women's University Research Grants 2011.

$x \in N$, the homogeneous components of x are again in N . Properties of finitely generated multiplication module have been studied in [2], [3], [6], and [7]. In this paper, we study the properties of finitely generated gr-multiplication modules and investigate when gr-multiplication modules are finitely generated.

2. Main results

DEFINITION 2.1. Let R be a graded ring and let M be a graded R -module. Then M is called a gr-multiplication module if for any graded submodule N of M , there exists a graded ideal I of R such that $N = IM$.

REMARK 2.2. [4] Let R be a graded ring. If M is a graded R -module and N is a submodule of M , then $(N : M)$ is a graded ideal of R , where $(N : M) = \{r \in R \mid rM \subseteq N\}$.

For any graded submodule N of M , we denote $(N : M)_g$ the graded ideal of R generated by $(h(N) : h(M)) = \{r \in h(R) \mid rh(M) \subseteq h(N)\}$. Note that $(N : M)_g$ is the graded ideal of R generated by $(N : M) \cap h(R)$ and that $(N : M)_g = (N : M)$.

The following lemma can be found in [1] and also in [8].

LEMMA 2.3. Let R be a graded ring. Let M be a finitely generated graded R -module and let I be a graded ideal of R such that $M = IM$. Then there exists $q \in I$ such that $(1 - q)M = 0$.

Proof. See Lemma 2.1 in [8] □

DEFINITION 2.4. An R -module M is faithful if, whenever $r \in R$ is such that $rM = 0$, then $r = 0$.

THEOREM 2.5. Let R be a graded ring and M a faithful gr-multiplication R -module. Then the following statements are equivalent.

- (i) M is finitely generated.
- (ii) If A and B are graded ideals of R such that $AM \subseteq BM$ then $A \subseteq B$.
- (iii) For each graded submodule N of M , there exists a unique graded ideal I of R such that $N = IM$.
- (iv) $M \neq AM$ for any proper graded ideal A of R .
- (v) $M \neq PM$ for any gr-maximal ideal P of R .

Proof. (i) $\implies$ (ii).

Suppose that M is finitely generated. Let A and B be graded ideals of

MODULE-THEORETIC CHARACTERIZATIONS OF GENERALIZED GCD DOMAINS, III

HWANKOO KIM* AND SEUNGKOOK PARK**

ABSTRACT. In terms of divisible-like modules, some equivalent conditions for an integral domain R to be a generalized GCD domain are given.

1. Introduction

Generalized GCD domains (for short, GGCD domains) were first introduced in [5], further investigated in [6], and since then, have played important roles in multiplicative ideal theory. Several ring- or ideal-theoretic characterizations of GGCD domains were given in the literature ([1, 2, 3, 4, 7]). The purpose of this note is to give another module-theoretic characterizations of GGCD domains, as a continuation of the study of module-theoretic characterizations of certain integral domains ([9, 10, 11, 12]).

We first introduce some definitions and notations. Let R be an integral domain with quotient field K . Let I be a nonzero fractional ideal of R . Then $I^{-1} := \{x \in K \mid xI \subseteq R\}$, $I_v := (I^{-1})^{-1}$, and $I_t := \bigcup \{J_v \mid J \subseteq I \text{ a nonzero finitely generated (f.g.) subideal of } I\}$. An ideal J of R is called a *GV-ideal*, denoted by $J \in GV(R)$, if J is a f.g. ideal of R with $J^{-1} = R$. A fractional ideal I of R is said to be *invertible* (resp., *t -invertible*) if $II^{-1} = R$ (resp., $(II^{-1})_t = R$).

For a torsion-free R -module M , Wang and McCasland defined the *w-envelope* of M as $M_w := \{x \in M \otimes_R K \mid Jx \subseteq M \text{ for some } J \in GV(R)\}$ ([17], cf., [9]). A torsion-free R -module is called a *w-module* (or

Received July 12, 2012; Accepted October 10, 2012.

2010 Mathematics Subject Classification: Primary 13A15; Secondary 13E99, 13F05.

Key words and phrases: generalized GCD domain, divisible module, injective module, pseudo-Dedekind domain.

Correspondence should be addressed to Hwankoo Kim, hkkim@hoseo.edu.

semidivisorial) if $M_w = M$. We say that a torsion-free R -module M is *w-finite* if $M = N_w$, for some f.g. submodule N of M .

Following [5], an integral domain R is called a *generalized GCD domain* (*GGCD domain*) if the intersection of two (integral) invertible ideals is invertible. It is well known that R is a GGCD domain if and only if I_v (equivalently I^{-1}) is invertible for every nonzero f.g. ideal I of R ([6, Theorem 1]). Recall that an integral domain R is called a *Prüfer v -multiplication domain* (for short, *PvMD*) if I_v (equivalently I^{-1}) is t -invertible for every nonzero f.g. ideal I of R . Thus the class of GGCD domains is contained in the class of PvMDs. It is also well known that in a PvMD, $t = w$. Therefore, R is a GGCD domain if and only if every w -finite w -ideal is invertible. For any undefined terminologies, we refer to [8] or [16].

2. Main results

Let $\mathcal{F}$ be a set of ideals of the integral domain R . An R -module M is said to be $\mathcal{F}$ -*injective* if for every ideal $I \in \mathcal{F}$, every R -homomorphism from I into M can be extended to an R -homomorphism from R into M . Denote by $\mathcal{F}_w(R)$ (resp., $\mathcal{F}_{w,f}(R)$) the set of all w -ideals (resp., w -finite w -ideals) of R .

In [14], Matlis introduced the notion of h -divisible modules. Recall that an R -module is said to be h -*divisible* if it is a homomorphic image of an injective R -module. In [13], Lee defined the notion of weak-injective modules: An R -module M is called *weak-injective* if $\text{Ext}_R^1(N, M) = 0$ for all R -modules N of weak dimension ≤ 1 . In [15], Nikandish introduced the concept of hw -divisible modules. Recall that an R -module is hw -*divisible* if it is a homomorphic image of a weak-injective R -module. In [18, Theorem 4], it was shown that an integral domain R is Prüfer if and only if for any divisible R -module M and any f.g. ideal I of R , $\text{Ext}_R^1(R/I, M) = 0$. We generalize this result to GGCD domains using the technique in the proof of [18, Theorem 4] in the following:

THEOREM 2.1. *Let R be an integral domain. Then the following are equivalent:*

- (1) R is a GGCD domain;
- (2) every divisible R -module is $\mathcal{F}_{w,f}(R)$ -injective;
- (3) every h -divisible R -module is $\mathcal{F}_{w,f}(R)$ -injective;
- (4) every hw -divisible R -module is $\mathcal{F}_{w,f}(R)$ -injective.

RELATION BETWEEN THE FLOOR BOUND AND THE ORDER BOUND

SEUNGKOOK PARK*

ABSTRACT. The purpose of this paper is to show that the coset bound can be used to prove the floor bound. Our proof provides a natural relation between the floor bound and the order bound.

1. Introduction

For algebraic geometric codes, it is in general hard to determine the actual minimum distance. The main methods for finding the lower bound for the minimum distance of an algebraic geometric code can be divided into two categories which are Lundell-McCullough floor bound [4] and Beelen order bound [1]. The connection between the two bounds were given in [2, 3]. In [3], ABZ bound for codes and ABZ bound for cosets are formulated. The authors prove that ABZ bound for codes improves the floor bound. Then by showing that the order bound, obtained with the ABZ bound for cosets, is at least the ABZ bound for codes, they prove that the order type bound improves the floor bound. In this paper, we give a direct proof that the floor bound can be obtained by using the decompositions from two different order sequences and merging them into one new sequence. Our proof establishes a natural relation between the floor bound and the order bound. The paper is organized as follows: In Section 2 we give the definition of an algebraic geometric code and recall the method for finding the lower bound for the minimum distance based on the order bound. In Section 3 we prove the Lundell-McCullough floor bound using the order bound, which provides a better understanding of the relation between the two bounds. In

Received December 26, 2011; Accepted July 19, 2012.

2010 Mathematics Subject Classification: Primary 11T71; Secondary 14G50.

Key words and phrases: algebraic-geometric code, floor bound, order bound.

*This research was supported by the Sookmyung Women's University Research Grants 2012.

Section 3, we give an example for finding the minimum distance using the floor bound and the order bound.

2. Algebraic geometric codes and order bound

Let $X/\mathbb{F}$ be an algebraic curve (absolutely irreducible, smooth, projective) of genus g over a finite field $\mathbb{F}$. Let $\mathbb{F}(X)$ be the function field of $X/\mathbb{F}$ and let $\Omega(X)$ be the module of rational differentials of $X/\mathbb{F}$. Given a divisor E on X defined over $\mathbb{F}$, let $L(E) = \{f \in \mathbb{F}(X) \setminus \{0\} : (f) + E \geq 0\} \cup \{0\}$ and let $\Omega(E) = \{\omega \in \Omega(X) \setminus \{0\} : (\omega) \geq E\} \cup \{0\}$. Let K represent the canonical divisor class. For n distinct rational points $P_1, \dots, P_n$ on X and for disjoint divisors $D = P_1 + \dots + P_n$ and G , the geometric Goppa codes $C_L(D, G)$ and $C_\Omega(D, G)$ are defined as the images of the maps

$$\begin{aligned}\alpha_L : L(G) &\longrightarrow \mathbb{F}^n, f \mapsto (f(P_1), \dots, f(P_n)), \\ \alpha_\Omega : \Omega(G - D) &\longrightarrow \mathbb{F}^n, \omega \mapsto (\text{res}_{P_1}(\omega), \dots, \text{res}_{P_n}(\omega)).\end{aligned}$$

The maps establish isomorphisms $L(G)/L(G - D) \simeq C_L(D, G)$ and $\Omega(G - D)/\Omega(G) \simeq C_\Omega(D, G)$. The codes $C_L(D, G)$ and $C_\Omega(D, G)$ are dual to each other. The Hamming distance between two vectors $x, y \in \mathbb{F}^n$ is $d(x, y) = |\{i : x_i \neq y_i\}|$. The minimum distance of a non-trivial linear code $\mathcal{C}$ is

$$\begin{aligned}d(\mathcal{C}) &= \min \{d(x, y) : x, y \in \mathcal{C}, x \neq y\} \\ &= \min \{d(x, 0) : x \in \mathcal{C}, x \neq 0\}.\end{aligned}$$

The Hamming distance between two nonempty subsets $X, Y \subset \mathbb{F}^n$ is the minimum of $\{d(x, y) : x \in X, y \in Y\}$. For a proper subcode $\mathcal{C}' \subset \mathcal{C}$, the minimum distance of the collection of cosets $\mathcal{C}/\mathcal{C}'$ is

$$\begin{aligned}d(\mathcal{C}/\mathcal{C}') &= \min \{d(x + \mathcal{C}', y + \mathcal{C}') : x, y \in \mathcal{C}, x - y \notin \mathcal{C}'\} \\ &= \min \{d(x, 0) : x \in \mathcal{C}, x \notin \mathcal{C}'\}.\end{aligned}$$

For two vectors $x, y \in \mathbb{F}^n$, let $x * y \in \mathbb{F}^n$ denote the Hadamard or coordinate-wise product of the two vectors. We state the theorem of coset bound from ([3], Theorem 1.2.) without proof.

THEOREM 2.1 (Coset bound). *Let $\mathcal{C}/\mathcal{C}_1$ be an extension of $\mathbb{F}$ -linear code with corresponding extension of dual codes $\mathcal{D}_1/\mathcal{D}$ such that $\dim \mathcal{C}/\mathcal{C}_1 = \dim \mathcal{D}_1/\mathcal{D} = 1$. If there exist vectors $a_1, \dots, a_w$ and $b_1, \dots, b_w$ such*

SOME PROPERTIES OF GR-MULTIPLICATION MODULES

SEUNGKOOK PARK

ABSTRACT. In this paper, we provide the necessary and sufficient conditions for a faithful graded module to be a graded multiplication module and for a graded submodule of a faithful gr-multiplication to be gr-essential.

1. Introduction

Let R be a commutative ring with identity $1 \neq 0$ and M a unital R -module. M is called a *multiplication module* provided for each submodule N of M , there exists an ideal I of R such that $N = IM$ [2]. Let G be a multiplicative group with identity e . A ring R is said to be a *graded ring of type G* if there is a family of additive subgroups of R , say $\{R_i \mid i \in G\}$, such that $R = \bigoplus_{i \in G} R_i$ and $R_i R_j \subseteq R_{ij}$ for all $i, j \in G$, where $R_i R_j$ is the set of all finite sums of products $r_i r_j$ with $r_i \in R_i$ and $r_j \in R_j$. The elements of $h(R) = \bigcup_{i \in G} R_i$ are called the homogeneous elements of R . Any nonzero $r \in R$ has a unique expression as a sum of homogeneous elements, that is, $r = \sum_{i \in G} r_i$ where r_i is nonzero for a finite number of i in G . The nonzero elements r_i in the decomposition of r are called the homogeneous components of r . Let R be a graded ring of type G then R -module M is said to be a *graded R -module* if there is a family $\{M_i \mid i \in G\}$ of additive subgroups of M such that $M = \bigoplus_{i \in G} M_i$ and $R_i M_j \subseteq M_{ij}$ for all $i, j \in G$. Elements of $h(M) = \bigcup_{i \in G} M_i$ are called the homogeneous elements of M . A submodule N of M is a graded submodule if $N = \bigoplus_{i \in G} (N \cap M_i)$, or equivalently, if for any $x \in N$, the homogeneous components of x are again in N . Properties

Received July 22, 2012. Revised August 22, 2012. Accepted September 5, 2012.
2010 Mathematics Subject Classification: 16W50, 13A02.

Key words and phrases: gr-multiplication module, multiplication module.

This Research was supported by the Sookmyung Women's University Research Grants 2012.

of multiplication module have been studied by many mathematicians [1], [2], [3], [5], [6], [7], [8], [9], [10]. In this paper, we generalize some of the properties of the multiplication modules to graded multiplication modules.

2. Gr-multiplication modules

In this Section we state the definition of the gr-multiplication module and introduce a basic theorem which will be a main tool used to provide proofs of the theorems in the following sections.

DEFINITION 2.1. Let R be a graded ring and let M be a graded R -module. Then M is called a *gr-multiplication module* if for any graded submodule N of M , there exists a graded ideal I of R such that $N = IM$.

For any graded submodule N of M , we denote $(N : M)_g$ the graded ideal of R generated by $(h(N) : h(M)) = \{r \in h(R) \mid rh(M) \subseteq h(N)\}$. Note that $(N : M)_g$ is the graded ideal of R generated by $(N : M) \cap h(R)$ and that $(N : M)_g = (N : M)$, where $(N : M) = \{r \in R \mid rM \subseteq N\}$. Note that if M is a graded R -module and N is a submodule of M , then $(N : M)$ is a graded ideal of R [4].

PROPOSITION 2.2. Let R be a graded ring and let M be a graded R -module. Then M is a gr-multiplication R -module if and only if for any graded submodule N of M , $N = (N : M)_g M$.

Proof. Suppose that M is a gr-multiplication module and let N be a graded submodule. Then $N = IM$ for some graded ideal I of R . Since $I \subseteq (N : M) = (N : M)_g$, $N = IM \subseteq (N : M)_g M \subseteq N$. Thus $N = (N : M)_g M$. The other direction of the proof is clear by taking $(N : M)_g = I$. This completes the proof. $\square$

REMARK. If M is a graded module and a multiplication module, then M is a gr-multiplication module. However, a gr-multiplication module may not be a multiplication module. An example of a gr-multiplication module which is not a multiplication module is given in [4].

PROPOSITION 2.3. Let R be a graded ring and let M be a graded R -module. Then M is a gr-multiplication module if and only if for each $m \in h(M)$, there exists a graded ideal I of R such that $Rm = IM$.



Voluntary retirement and portfolio selection: Dynamic programming approaches

Yong Hyun Shin^{*}

Department of Mathematics, Sookmyung Women's University, Seoul 140742, Republic of Korea

ARTICLE INFO

Article history:

Received 27 October 2011

Received in revised form 13 March 2012

Accepted 16 March 2012

Keywords:

Voluntary retirement

Cobb–Douglas utility

Dynamic programming method

Portfolio selection

ABSTRACT

I consider a continuous-time optimal consumption and portfolio selection problem with voluntary retirement. When the agent's utility of consumption and leisure are of Cobb–Douglas form, I use the dynamic programming method to derive the value function and optimal strategies in closed-form. These coincide with the solutions of Farhi and Panageas (2007) [7], who have solved the problem using a martingale method.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Following the pioneering works of Merton [1,2], there have been many studies on the problem of continuous-time optimal consumption and portfolio selection. One of the main issues considered within the context of portfolio selection problems is the voluntary retirement problem. (See [3–7].) The literature regarding voluntary retirement can be largely divided into the following three categories:

- Labor–leisure model: Farhi and Panageas [7], Choi et al. [5], Barucci and Marazzinay [3]
- Labor–disutility model: Choi and Shim [4]
- Stochastic labor income model: Dybvig and Liu [6], Barucci and Marazzinay [3].

Under the labor–leisure framework, I consider a problem when the agent's utility of consumption and leisure are of Cobb–Douglas form. The setup is similar to that of Farhi and Panageas [7], who have solved the problem using a martingale method. My main contributions are the following.

- Methodological contribution: I use the dynamic programming method based on Karatzas et al. [8] to solve a voluntary retirement problem when the agent has Cobb–Douglas utility of consumption and leisure.
- I derive the value function and optimal strategies in closed-form: these coincide with those found by Farhi and Panageas [7].
- I provide a comparison of the dynamic programming solution and the martingale solution in this framework.

2. The financial market

It is assumed that there are two assets in the financial market: a riskless asset and a risky asset. It is also assumed that the interest rate for the riskless asset is a constant r , and the risky asset S_t follows the stochastic differential equation

^{*} Fax: +82 2 2077 7323.

E-mail address: yhshin@sookmyung.ac.kr.

(SDE) $dS_t/S_t = \mu dt + \sigma dB_t$, where B_t is a standard Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with the augmentation $\{\mathcal{F}_t\}_{t \geq 0}$ generated by the standard Brownian motion $\{B_t\}_{t \geq 0}$. (See Section 1.7 of Karatzas and Shreve [9] for a detailed mathematical construction in the infinite-horizon setting of the probability space.) It is assumed that parameters of the risky asset μ and σ are constants.

The market price of risk is denoted by $\theta \triangleq \frac{\mu-r}{\sigma}$. Let π_t be the amount of money invested in the risky asset S_t at time t , c_t be the consumption rate process at time t and τ be an $\mathcal{F}_t$ -stopping time considered as voluntary retirement time from labor. The portfolio process π_t is an $\mathcal{F}_t$ -measurable, adapted process such that, for all $t \geq 0$, almost surely (a.s.),

$$\int_0^t \pi_s^2 ds < \infty.$$

The consumption rate process c_t is a progressively $\mathcal{F}_t$ -measurable, adapted, nonnegative process such that, for all $t \geq 0$, a.s.,

$$\int_0^t c_s ds < \infty.$$

The agent in this model receives constant labor income $y_0 > 0$ before retirement. So the agent's wealth process X_t at time t follows the SDE

$$dX_t = [rX_t + \pi_t(\mu - r) - c_t + y_0 \mathbf{1}_{\{0 \leq t < \tau\}}]dt + \sigma \pi_t dB_t,$$

with an initial endowment $X_0 = x > -y_0/r$.

3. Cobb–Douglas utility

First I introduce the Cobb–Douglas utility function $u(c_t, l_t)$ of the form:

$$u(c_t, l_t) \triangleq \frac{1}{\alpha} \cdot \frac{(c_t^\alpha l_t^{1-\alpha})^{1-\gamma}}{1-\gamma}, \quad 0 < \alpha < 1 \text{ and } \gamma > 0 (\gamma \neq 1), \quad (3.1)$$

where γ is the agent's coefficient of relative risk aversion and α is a constant parameter. c_t is a consumption rate process at time t and l_t is a leisure rate process at time t . I assume that l_t is a binary choice variable. That is, l_t can only take two values l_1 or $\bar{L}$; $l_t = l_1$ before retirement, and $l_t = \bar{L}$ after retirement. I further assume that $l_1 < \bar{L}$. The labor income stream is defined as $y_0 \triangleq w(\bar{L} - l_1) > 0$, where w is the constant wage rate. Without loss of generality I normalize $l_1 = 1$ and define $\gamma_1 \triangleq 1 - \alpha(1 - \gamma)$. Then the Cobb–Douglas utility function $u(\cdot, \cdot)$ in (3.1) can be rewritten as follows:

$$u(c_t, 1) = \frac{c_t^{1-\gamma_1}}{1-\gamma_1}, \quad \text{if } 0 \leq t < \tau, \quad \text{and} \quad u(c_t, \bar{L}) = \bar{L}^{\gamma_1-\gamma} \frac{c_t^{1-\gamma_1}}{1-\gamma_1}, \quad \text{if } t \geq \tau.$$

Remark 3.1. If $\gamma > 1$, I obtain that $\gamma > \gamma_1 > 1$ and $\bar{L}^{\frac{\gamma-\gamma_1}{\gamma_1}} > 1$. If $0 < \gamma < 1$, then $0 < \gamma < \gamma_1 < 1$ and $\bar{L}^{\frac{\gamma-\gamma_1}{\gamma_1}} < 1$.

4. The optimization problem

The agent's optimization problem is to maximize her expected utility

$$\begin{aligned} V(x) &= \sup_{(c, \pi, \tau) \in \mathcal{A}(x)} \mathbb{E} \left[\int_0^\infty e^{-\beta t} \left(\frac{c_t^{1-\gamma_1}}{1-\gamma_1} \mathbf{1}_{\{0 \leq t < \tau\}} + \bar{L}^{\gamma_1-\gamma} \frac{c_t^{1-\gamma_1}}{1-\gamma_1} \mathbf{1}_{\{t \geq \tau\}} \right) dt \right] \\ &= \sup_{(c, \pi, \tau) \in \mathcal{A}(x)} \mathbb{E} \left[\int_0^\tau e^{-\beta t} \frac{c_t^{1-\gamma_1}}{1-\gamma_1} dt + \bar{L}^{\gamma_1-\gamma} \int_\tau^\infty e^{-\beta t} \frac{c_t^{1-\gamma_1}}{1-\gamma_1} dt \right] \\ &= \sup_{(c, \pi, \tau) \in \mathcal{A}(x)} \mathbb{E} \left[\int_0^\tau e^{-\beta t} \frac{c_t^{1-\gamma_1}}{1-\gamma_1} dt + e^{-\beta \tau} U(X_\tau) \right], \end{aligned} \quad (4.1)$$

where $\beta > 0$ is a subjective discount factor and $\mathcal{A}(x)$ is an admissible class of triples (c, π, τ) .

Assumption 4.1.

$$K_1 \triangleq r + \frac{\beta - r}{\gamma_1} + \frac{\gamma_1 - 1}{2\gamma_1^2} \theta^2 > 0.$$

Remark 4.1. I use the solution of the classical Merton problem to derive the function $U(\cdot)$ in (4.1) as follows:

$$U(x) = \frac{\bar{L}^{\gamma_1-\gamma}}{K_1^{\gamma_1}} \frac{1}{1-\gamma_1} x^{1-\gamma_1}.$$

PORTFOLIO SELECTION WITH REGIME-SWITCHING: DYNAMIC PROGRAMMING APPROACHES

YONG HYUN SHIN*

ABSTRACT. I study an optimal consumption and portfolio selection problem with regime-switching using a dynamic programming method. With constant relative risk aversion (CRRA) utility I obtain optimal solutions in closed-form.

1. Introduction

Motivated by Merton's seminal works [5, 6], much scholarly work has been conducted on portfolio optimization problems. Especially Karatzas *et al.* [4] have derived general explicit solutions of a consumption / investment optimization problem using a dynamic programming method. The regime-switching model, in which uncertainty is affected not only by the Brownian motion of a stock, but also by the long-term business cycle which is mathematically modeled as a continuous-time Markov chain, is becoming widely considered in mathematical finance (option pricing [2], real options [1], portfolio selection problems [3, 7]).

In this paper I investigate an optimal consumption and portfolio selection problem with regime-switching under the framework of Karatzas *et al.* [4]. With constant relative risk aversion (CRRA) utility I derive optimal solutions in closed-form.

2. The financial market

For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, I define a standard Brownian motion B_t and a continuous-time two-state Markov chain ϵ_t . It is assumed that B_t and ϵ_t are independent and the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ is generated by the Brownian motion B_t and the Markov chain ϵ_t .

Received January 25, 2012; Revised April 22, 2012; Accepted April 24, 2012.

2010 Mathematics Subject Classification: Primary 91G10.

Key words and phrases: Regime-switching, CRRA utility, portfolio selection.

In this financial market, two assets are traded: One is a money market account and the other is a stock. It is assumed that there are two regimes 1, 2 in the market and regime i switches into regime j at the first jump time of an independent Poisson process with intensity λ_i , for $i, j \in \{1, 2\}$. So, in regime $i \in \{1, 2\}$, the interest rate is r_i , and the stock price process is governed by $dS_t/S_t = \mu_i dt + \sigma_i dB_t$. The market price of risk is defined by $\theta_i \triangleq \frac{\mu_i - r_i}{\sigma_i}$ $i = 1, 2$. Let π_t be the $\mathcal{F}_t$ -progressively measurable portfolio process, the amount of the agent's wealth invested in the risky asset at time t and c_t be the nonnegative $\mathcal{F}_t$ -progressively measurable consumption rate process at time t . It is assumed that they satisfy the following conditions:

$$\int_0^t c_s ds < \infty \quad \text{and} \quad \int_0^t \pi_s^2 ds < \infty, \quad \text{for all } t \geq 0, \text{ almost surely (a.s.).}$$

The agent's wealth process X_t at time t follows the stochastic differential equation (SDE)

$$dX_t = [r_i X_t + \pi_t(\mu_i - r_i) - c_t] dt + \sigma_i \pi_t dB_t, \quad X_0 = x > 0, \quad i = 1, 2.$$

3. The optimization problem

The agent's optimization problem is to maximize her expected utility

$$(3.1) \quad V_i(x) = \sup_{(c, \pi) \in \mathcal{A}(x)} \mathbb{E} \left[\int_0^{T_i} e^{-\beta t} \frac{c_t^{1-\gamma}}{1-\gamma} dt + e^{-\beta T_i} V_j(X_{T_i}) \right],$$

where T_i is the first jump time from i -th state to j -th state, $\beta > 0$ is a subjective discount factor, and $\mathcal{A}(x)$ is an admissible class of pair (c, π) at x , where $i, j \in \{1, 2\}$ and $i \neq j$.

In order to guarantee the agent's problem to be well-defined, we have the following assumption.

ASSUMPTION 1.

$$K_i \triangleq r_i + \frac{\beta - r_i}{\gamma} + \frac{\gamma - 1}{2\gamma^2} \theta_i^2 > 0, \quad i = 1, 2.$$

The next theorem gives my main results.

THEOREM 1. *The value function is given by*

$$V_i(x) = M_i \frac{x^{1-\gamma}}{1-\gamma},$$

통계학과

New Design of Choice Sets for Choice-based Conjoint Analysis

Bu-Yong Kim¹

¹Department of Statistics, Sookmyung Women's University

(Received May 3, 2012; Revised July 5, 2012; Accepted September 15, 2012)

Abstract

This article is concerned with choice-based conjoint analysis versus rating-based and ranking-based conjoint analysis. Choice-based conjoint analysis has a definite advantage in that the respondent's task of choosing the most preferred profile from several competing profiles adequately mimics consumer marketplace behavior. It is crucial to design the choice sets appropriate for the choice-based conjoint. Thus, this article suggests a new method to design the choice sets that are well-balanced. It augments the balanced incomplete block design and then obtains the dual design of the result to accommodate various numbers of profiles. In consequence, the choice sets designed by the new method have the desirable characteristics that each profile is presented to the same number of respondents, and pairs of any two distinct profiles occur together in the same number of choice sets. The balancing of the design increases the efficiency of the conjoint analysis. In addition, the pair-comparison scheme can improve the quality of data through the identification of contradictory responses.

Keywords: Choice-based conjoint analysis, choice set, balanced incomplete block design, dual design.

1. 서론

컨조인트분석은 학계와 산업계에서 마케팅조사기법 중의 하나로 발전되어 왔는데, 최근 컨조인트분석을 위한 소프트웨어들의 보급에 힘입어 시장분석 도구로서의 유용성이 크게 증가하였다. 제품(서비스, 전략, 정책, 디자인, 프로그램 등도 포함)을 구성하는 주요 속성과 수준을 파악하여 제품프로파일을 설계하고, 각 프로파일에 대한 소비자들의 종합적인 선호도를 측정하여 각 수준에 소비자가 부여하는 부분가치를 추정하고, 각 속성의 상대적 중요도를 평가하며, 소비자들이 시장에서 실제로 선택할 제품을 예측하는데 컨조인트분석이 적용된다. 이 기법은 주로 제품의 속성 및 수준과 가격 등의 제품특성을 최적화하기 위해 활용되며, 시장세분화, 포지셔닝과 판매촉진을 위한 기준설정, 시장점유율 예측 등의 마케팅활동에 광범위하게 활용되고 있다. Green과 Srinivasan (1990), Marshall과 Bradlow (2002), Giancrisofaro (2003)는 컨조인트분석에 관한 이론과 방법론 그리고 응용에 관하여 연구하였다. 한편, Green 등 (1981)은 제품을 구성하는 속성들의 최적수준을 알아내기 위하여, Kamakura (1988)는 시장세분화를 위하여, Barone과 Lombardo (2004)는 대중교통 서비스의 품질향상을 위하여, Kim (2005)은

This research was supported by Sookmyung Women's University Research Grants (2011).

¹Professor, Department of Statistics, Sookmyung Women's University, Yongsan-gu, Seoul 140-742, Korea.
E-mail: buykim@sookmyung.ac.kr

휴대전화기 신제품개발을 위하여, Shin 등 (2007)은 담뱃갑 표지의 경고문구가 금연의지에 미치는 영향을 파악하기 위하여 컨조인트분석을 활용하였다.

컨조인트분석의 초기단계에서는 분석대상 제품의 주요 속성과 수준을 결정하고 가능한 제품프로파일들을 정의한다. 그리고 응답자들에게 제품프로파일들을 제시하고 선호도를 평가하도록 하는데, 기존의 컨조인트분석 기법들은 점수나 순위의 형태로 평가하는 방식을 채택하고 있다. 그러나 이러한 선호도 평가방법들은 제품프로파일의 개수가 많은 경우에는 응답자들에게 막대한 부담을 지우고 선호도를 정확하게 평가하기 어렵게 한다는 단점을 가지고 있다. 더욱이 이 방법들은 응답자들에게 실제시장에서 그들이 취하는 구매행동과 다른 평가방식을 강요한다는 중대한 결함을 가지고 있다. 이러한 문제점들을 극복하기 위하여, 적절한 규모의 선택집합(choice set)을 구성하고 응답자들이 각 선택집합에 포함된 소수의 프로파일들 중에서 가장 선호하는 것을 선택하거나, 그 중에서 어느 것도 선택하지 않을 수 있도록 하는 평가방식을 채택할 수 있다. 일반적으로 각 선택집합에는 소수의 프로파일들이 포함되므로 응답자가 가장 선호하는 것을 선택하는데 어려움이 없고, 이러한 선택방식은 실제시장에서 소비자들이 취하는 구매행동과 매우 유사하기 때문에 정확한 선호도를 얻을 수 있다는 장점을 갖는다. 이와 같은 선호도 평가방식을 채택하는 컨조인트분석을 선택기반 컨조인트분석(choice-based conjoint analysis; CBCA)라고 하는데, 관광산업을 비롯하여 다양한 분야에서 많이 활용되고 있다 (Lazari와 Anderson, 1994). 특히, 1990년대부터는 마케팅조사 분야에서 CBCA를 적용하게 되었는데 (Kuhfeld와 Tobias, 2005), Desarbo 등 (1995)은 시장세분화에 CBCA를 활용하였다. 한편, Elrod 등 (1992)과 Chakraborty 등 (2002)은 점수기반 컨조인트분석과 CBCA의 시장점유율 예측력을 비교하였으며, Moore (2004)는 점수기반 컨조인트분석과 CBCA의 교차타당성을 비교하였고, Lim 등 (2006)은 컨조인트분석 기법들의 예측력을 비교하고 전통적인 방식과 CBCA를 결합한 방식을 제안하였다.

CBCA를 실행하기 위해서는 응답자들에게 제시할 선택집합들을 적절한 형태로 구성하는 것이 매우 중요하다. 그런데 흔히 사용되는 임의배정방법은 각 프로파일이 선택집합에 동일한 횟수씩 배정된다는 보장을 할 수 없어서 CBCA의 효율성을 낮출 수 있다는 단점을 가지고 있다. 한편, 직교설계방법은 이상적이긴 하지만 선택집합의 수가 많거나 선택집합의 크기가 지나치게 큰 경우가 많아서 응답자에게 부담을 줄 수 있고, 연구자가 원하는 특정 규모의 선택집합을 구성할 수 없기 때문에 현실적으로 적용하기 어렵다는 한계를 가지고 있다. 따라서 본 연구에서는 선택집합을 설계하는 방법의 대안을 제안하고자 하는데, 실험계획법의 하나인 균형불완비블록설계(balanced incomplete block design; BIBD)를 바탕으로 선택집합을 설계하는 방법을 제시하고자 한다. 우선 제 2장에서는 CBCA의 전반적인 조사실행과정과 분석방법을 요약하고, 제 3장에서는 선택집합을 설계하는 새로운 방법에 관하여 구체적으로 기술하고자 한다.

2. 선택기반 컨조인트분석

CBCA가 다른 방식의 컨조인트분석들과 가장 상이한 점은 제품프로파일들에 대한 응답자들의 선호도를 점수나 순위로 측정하지 않고, 프로파일들을 적절한 개수의 선택집합으로 묶어 응답자에게 제시하고 각 선택집합에 포함된 몇 개의 프로파일들 중에서 가장 선호하는 것을 선택하거나 그것들 중에서 아무 것도 선택하지 않을 수 있도록 한다는 것이다. 이러한 CBCA는 다음과 같은 장점들을 갖는다. 첫째, 선택집합에 포함된 3-6개 정도의 프로파일들 중에서 가장 선호하는 것을 선택하는 작업은 실제로 시장에서 소비자들이 취하는 구매행동과 매우 유사하기 때문에 응답자들의 선호도 측정결과가 매우 현실적인 것이 된다. 둘째, 몇 개의 제품 중에서 가장 선호하는 제품 하나를 선택하는 작업은 누구나 쉽게 할 수 있고 자연스러운 일 (Moore, 2004)이기 때문에 정확한 선호도측정이 가능하다. 셋째, 선택집합에는 제품프로파일들과 함께 ‘이것들 중에서 아무 것도 사지 않겠다(no-choice)’라는 선택지를 포함시키기 때

Analysis of Consumer Preferences for Cosmetic Essence-for-Men via Choice-Based Conjoint with New Design of Choice Sets

Bu-Yong Kim¹ · Jiyoung Kim² · Yu-Yeong Kan³

¹Department of Statistics, Sookmyung Women's University

²Department of Merchandising and Digital Retailing, University of North Texas

³Research Division, Research & Research, Inc.

(Received October 10, 2012; Revised November 17, 2012; Accepted November 27, 2012)

Abstract

The sales volume of men's cosmetics has drastically increased in Korea. In recent years, men's needs for cosmetics have been diversified and the consumer demand for functional cosmetics has greatly risen. In particular, male consumers have become more interested in essence product that is a light and concentrated treatment to correct skin problems. This research analyzes consumer preferences for essence-for-men through the use of choice-based conjoint analysis. This approach is adopted since the task of respondents to choose the most preferred option from several alternatives closely mimics actual marketplace purchasing behavior by consumers. New technique for the construction of choice sets is suggested based on the balanced incomplete block design, to accommodate a larger number of product profiles. The proposed design for choice sets is balanced and provides a tool to filter the contradictory choices. Conjoint analyses are performed to assess the relative importance of attributes and identify the most preferred profile of essence-for-men with respect to attributes such as emphasized function, price, type of content, and design of container. Some differences are indicated in the analysis results between age brackets as well as between groups classified by the amount of fashion item expenditures.

Keywords: Consumer preference, essence-for-men, choice-based conjoint analysis, choice set.

1. Introduction

Annual sales of the cosmetics in Korea amounted to 10 trillion won in 2011 and are expected to continue to increase in volume. In recent years, men have become more interested in enhancing and maintaining their facial appearance. It has resulted in a drastic growth in the men's cosmetics market. The men's cosmetics market (which includes products for facial-care, hair-care, and body-care) grew from 530 billion won in 2007 to 800 billion won in 2010, and the volume was forecast at 1 trillion won for 2011 (Jeon and Jae, 2009; Economic Review, Aug. 23, 2011). The demand

This research was supported by Sookmyung Women's University Research Grants (2012).

¹Corresponding author: Professor, Department of Statistics, Sookmyung Women's University, Chungpa-dong, Yongsan-gu, Seoul 140-742, Korea. E-mail: buykim@sookmyung.ac.kr

for cosmetics with enhanced functions has highly increased due to the diversified needs of male consumers. In particular, male consumers' interests in the concentrated solution product ('essence') are supported by various study results. The survey results of Hong (2008) indicate that 69.7% of men recognize the needs for functional cosmetics such as lotion and essence; in addition, Park (2008) indicates that the proportion of men using essence is the highest after sunscreen-cream among the usage of functional cosmetics. To keep pace with these trends, leading cosmetics companies are expected to focus on the development of new essence products that cater to the needs of men. Thus, this research attempts to analyze the preferences of male consumers for essence by applying a conjoint analysis approach.

Conjoint analysis has received considerable academic and industry attention as an approach that most realistically portrays consumer decisions. In particular, it has been widely employed in marketing researches that encompass new product development, pricing, positioning, and market share prediction. Moreover, with recent advances in computer software, conjoint analysis has been increasingly used in marketing research as an effective analysis tool. For example, Green and Srinivasan (1990), Green *et al.* (2001), Marshall and Bradlow (2002), Giancristofaro (2003) introduced the methodologies and applications of conjoint analysis. Wittink *et al.* (1994) investigated the commercial use of conjoint analysis in Europe. Kim (2005) conducted conjoint analysis to study consumer preferences for cellular phones. Shin *et al.* (2007) employed conjoint analysis to investigate the effects of cigarette warning labels and packaging on user's intention to quit smoking. Cassab (2009) applied conjoint analysis to mobile phone service to investigate the dynamics of multi-channel service attributes and their influence on consumers' loyalty intentions.

In the earlier step of the conjoint analysis, a set of product profiles are defined by the factorial design after attributes and levels of the product are identified. Then the preference measures of consumers are elicited usually by asking respondents to rate or rank their preferences for the profiles. Those rating-based and ranking-based conjoint analyses have been widely employed in marketing research since it is relatively convenient to measure consumer preferences by conducting a personal interview with a questionnaire. However, it can be very difficult for respondents to accurately measure preferences when they are asked to rate or rank a large number of profiles. In addition, the respondents could feel burdened and might even reject to cooperate with the interview process. Moreover, rating and ranking tasks have a serious flaw since decision making process by respondents in those tasks is vastly different from their actual purchasing behavior when they do not necessarily exhaust the entire list of products to give ratings or rankings. In contrary, the task of choosing one profile from a small group of profiles closely resembles the actual marketplace purchasing behavior of consumers. Moreover, the choice task is less cumbersome for consumers as it asks them to choose the most preferred one among three to five profiles. For these reasons, the choice-based conjoint analysis(CBCA) has been widely adopted in various areas of research and industry. Over the past several years, the marketing researchers became interested in using CBCA to predict consumer choices. For example, Elrod *et al.* (1992) and Chakraborty *et al.* (2002) compared the rating-based conjoint analysis and CBCA in their ability to correctly predict market shares under varying conditions. Moore (2004) compared cross-validity of rating-based conjoint analysis and CBCA. In addition, DeSarbo *et al.* (1995) performed the market segmentation via CBCA. Johnson and Orme (1996) reanalyzed data from 21 commercial studies conducted by the CBCA. However, CBCA is not without limitations since respondents are supposed to choose only the most preferred profile and it is impossible to measure how strong that preference is relative to other options. In this context, the statistical efficiency of CBCA might be lower than rating-based or ranking-based conjoint analysis;

Preliminary test of fit in a general class of conditionally heteroscedastic nonlinear time series

J.S. Baek^a, J.A. Park^b and S.Y. Hwang^{b*}

^aStatistical Research Institute, Daejeon, South Korea; ^bDepartment of Statistics, Sookmyung Women's University, Seoul, South Korea

(Received 4 May 2010; final version received 24 January 2011)

This article is concerned with a general class of conditionally heteroscedastic time series including possibly nonlinear and asymmetric autoregressive conditional heteroscedastic (ARCH) and generalized ARCH models. A problem of *preliminary test of fit* (PTF, hereafter) within the broad class under consideration is discussed. It is noted that contrary to usual tests in the literature of conditionally heteroscedastic time series, PTF does not require any specification of the conditional variance in advance. Based on the joint limit distributions of sample autocorrelations, a certain Portmanteau-type statistic for PTF is proposed, and its limit is shown to be a chi-square distribution. In addition, some simulation studies, under various innovations, are reported to support our theoretical results.

Keywords: conditional heteroscedasticity; nonlinear GARCH; nonlinear time series; preliminary test of fit

AMS 2000 subject classification: Primary 62M10

1. Motivation

In a seminal paper by Engle [1], the conditional variance (denoted by h_t throughout) of a time series is believed to be time varying rather than constant, especially in the field of financial time series. Autoregressive conditional heteroscedastic (ARCH) models of Engle [1] and a class of generalized ARCH (called GARCH) processes due to [2] have served as useful tools for analysing time-varying conditional variances inherent in a time series. Both ARCH and GARCH models are usually referred to as *linear-symmetric* models, in which, h_t is formulated as a linear function of past squared residuals and h_t features a symmetric pattern irrespective of the signs of residuals. In the last two decades, however, in response to various empirical evidences of asymmetric and nonlinear h_t arising mainly from financial time series, many successful efforts for nonlinear and/or asymmetric modelling for h_t have been made. Among others, we refer to [3–5] with references therein for diverse nonlinear and asymmetric ARCH and GARCH models. Consequently, at an early stage of data analysis, we frequently encounter a broad class of conditionally heteroscedastic models to choose from. This article deals with a class of conditionally heteroscedastic time series

*Corresponding author. Email: shwang@sookmyung.ac.kr

which is rich enough to include nonlinear and asymmetric ARCH/GARCH models. This class (denoted by G_A in Section 3) also accommodates various nonlinear autoregressive (AR) processes such as threshold models and exponential AR (EAR) processes as special cases (see Sections 2 and 3). Refer to, for instance, [6] for various nonlinear AR models and their applications.

We will first derive joint limit distributions of the sample autocorrelation functions (SACFs, for short) for this broad class. The limit distribution of SACFs enables us to construct a consistent estimator (*valid for all models* belonging to the class) of the limiting variance–covariance matrix of SACFs. We are then interested in checking for model adequacy of the time series belonging to the general class G_A . We refer to recent monograph by Li [7] for an excellent comprehensive review on various tests for fit in conditionally heteroscedastic time series. Most of the research on this problem has been directed to the case where the functional form of h_t is known. Accordingly, tests of fit usually require a specification of h_t . Engle's [1] ARCH test and Li and Mak's [8] test are typically of this kind, calling for the specification of h_t . See, for instance, [9] for a comparative simulation study on tests for heteroscedasticity.

On the other hand, this article is concerned with a test of fit when h_t is unspecified in the class G_A . Due to the broad structure of h_t being unspecified, our test can be regarded as a *preliminary test of fit* (PTF, hereafter) within the class G_A . The proposed PTF will be useful at an early stage of time series data analysis. If PTF is not rejected, then one may go ahead and select an appropriate heteroscedastic time series in the class G_A . Thus, the term 'preliminary' in PTF seems to be appropriate. Based on the limit distributions of SACFs for this general class, a certain Portmanteau-type statistic is proposed for the PTF. The proposed test is investigated through simulation studies under various innovations including normal distribution, Student's t -distribution and logistic distribution. Simulation experiment shows that the proposed statistic for PTF follows chi-square distributions for all time series models in the class, and that it produces empirical sizes successfully matching with the nominal size (see Sections 2 and 3). The proposed test is compared with a modified Portmanteau test, the so called Ljung and Box (LB, for short) test (cf. [7, p. 10]). Recently, Francq *et al.* [10] discussed tests for fit in a class of ARMA models with uncorrelated errors. They also observed poor performances of the LB test of fit when errors are nonindependent (but uncorrelated) as we have obtained in the simulation studies in Sections 2 and 3. In addition, we have conducted simulation experiments in Section 4 concerning the effect of misspecification on the tests of fit for a general class of conditionally heteroscedastic nonlinear time series.

2. A test of fit for time series with unspecified heteroscedasticity

Consider the following conditionally heteroscedastic time series defined by

$$\epsilon_t = \sqrt{h_t} e_t, \quad (1)$$

where $\{e_t\}$ is a sequence of (not necessarily Gaussian) i.i.d. random variates with mean zero and unit variance. Throughout, h_t stands for the conditional variance of ϵ_t , given the past, that is,

$$h_t = \text{Var}(\epsilon_t | F_{t-1}), \quad (2)$$

where F_{t-1} denotes the sigma field generated by $\epsilon_{t-1}, \epsilon_{t-2}, \dots$. It is assumed that

(C1) $\{\epsilon_t\}$ is a strictly stationary and ergodic time series with finite fourth-order stationary moment, that is, $E\epsilon_t^4 < \infty$.

Consider the following collection G of conditionally heteroscedastic time series $\{\epsilon_t\}$:

$$G = \{\text{time series models of the form 1 satisfying (C1)}\}. \quad (3)$$



Asymmetric GARCH processes featuring both threshold effect and bilinear structure

M.S. Choi, J.A. Park, S.Y. Hwang*

Department of Statistics, Sookmyung Women's University, Seoul, 140-742, Republic of Korea

ARTICLE INFO

Article history:

Received 11 October 2011
Received in revised form 22 November 2011
Accepted 22 November 2011
Available online 8 December 2011

MSC:

primary 62M10

Keywords:

Asymmetric volatility
Bilinear GARCH
Threshold GARCH

ABSTRACT

A class of asymmetric GARCH models is proposed by combining threshold effect and bilinear structure. The class is referred to as threshold-bilinear GARCH processes. A simulation study demonstrates that the class exhibits diverse asymmetries in volatilities, accommodating existing asymmetric models. Stationarity and existence of moments are discussed. Applications to Korean stock prices are illustrated.

© 2011 Elsevier B.V. All rights reserved.

1. Motivation

It is now widely recognized that conditionally heteroscedastic models have proved useful in analyzing volatilities arising mainly from financial time series context. When a zero mean time series $\{\epsilon_t\}$ is generated by the equation

$$\epsilon_t = \sqrt{h_t} e_t \quad \text{and} \quad h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}, \quad (1.1)$$

the time series $\{\epsilon_t\}$ is referred to as a first order GARCH, so called GARCH(1,1) (refer to [Bollerslev \(1986\)](#)). In (1.1) and throughout the paper, h_t denotes a volatility, that is, h_t is a conditional variance of ϵ_t given the past values $\epsilon_{t-1}, \epsilon_{t-2}, \dots$. Throughout, it will be assumed that $\{e_t\}$ is a sequence of iid innovations (not necessarily Gaussian) with mean zero and variance unity. Higher order GARCH models, i.e., GARCH(p, q) are obtained by adding additional lags of $\epsilon_{t-1}^2, \dots, \epsilon_{t-p}^2$ and $h_{t-1}, \dots, h_{t-q}$ in the RHS of (1.1). Standard GARCH(p, q) models are said to be symmetric (in the volatility h_t) in the sense that h_t is a symmetric function of past values of $\epsilon_{t-1}, \epsilon_{t-2}, \dots$. It is noted that symmetric GARCH models provide the identical volatility h_t both from ϵ_{t-i} and $-\epsilon_{t-i}$, $i = 1, 2, \dots$.

It is however often the case, in particular with financial time series, that we observe asymmetric patterns in that h_t varies depending on the signs of past values (cf. e.g., [Gourieroux \(1997, p. 90\)](#)), which in turn leads us to consider a threshold effect in modeling volatility. A first order threshold model (T-GARCH(1,1)) is formulated by

$$h_t = \alpha_0 + \alpha_{11}(\epsilon_{t-1}^+)^2 + \alpha_{12}(\epsilon_{t-1}^-)^2 + \beta_1 h_{t-1}, \quad (1.2)$$

where $\alpha_{11}, \alpha_{12} \geq 0$, and the notation $\epsilon^+ = \max(\epsilon, 0)$ and $\epsilon^- = \max(-\epsilon, 0)$ will be used throughout. Note that $\epsilon = \epsilon^+ - \epsilon^-$ and $|\epsilon| = \epsilon^+ + \epsilon^-$. Refer to, among others, [Glosten et al. \(1993\)](#), [Hwang and Basawa \(2004\)](#), [Liu \(2006\)](#), [Pan et al. \(2008\)](#)

* Corresponding author.

E-mail address: shwang@sookmyung.ac.kr (S.Y. Hwang).

and Hwang and Baek (2009) for a comprehensive discussion of probabilistic structures, inferential problems and data applications of T-GARCH(1,1). Note that $\alpha_{11} = \alpha_{12}$ in T-GARCH(1,1) reduces to the standard GARCH(1,1) model in (1.1). The case of $\alpha_{11} < \alpha_{12}$ corresponds to a “leverage effect” frequently observed in financial time series literature. A leverage effect in (financial) time series tends to provide a larger volatility h_t from the negative value of ϵ_{t-1} than from the positive value of the same magnitude (c.f. e.g., Tsay (2010, p. 111)). A useful tool for identifying involvement of the volatility h_t along with the immediate past value ϵ_{t-1} is called the news impact curve (NIC, hereafter) of h_t versus ϵ_{t-1} . The NIC was originally developed by Engle and Ng (1993). A typical NIC of T-GARCH(1,1) is shown in (d) of Fig. 4.1 for which a leverage effect is observed. It is noted that T-GARCH(1,1) models allow asymmetries in h_t by having “different slopes” α_{11} and α_{12} respectively for positive and negative values of ϵ_{t-1} .

Alternatively, one may introduce an asymmetry in h_t in (1.1) by adding the product of “first order” terms ϵ_{t-1} and $\sqrt{h_{t-1}}$. A first order bilinear GARCH process, denoted by BL-GARCH(1,1) is defined by

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} + \gamma_1 \epsilon_{t-1} \sqrt{h_{t-1}}, \quad (1.3)$$

where $-\infty < \gamma_1 < \infty$. Note that $\gamma_1 = 0$ reduces to a standard GARCH(1,1) and a leverage effect corresponds to $\gamma_1 < 0$ since then a positive quantity is added to h_t for $\epsilon_{t-1} < 0$. An asymmetry from BL-GARCH(1,1) essentially comes in h_t via a “shift” (from the origin) of the vertex in the NIC. Refer to (e) in Fig. 4.1 for a shift of the vertex to the right hand side from the origin. Bilinear GARCH models are known to be capable of accommodating shift-features in h_t . Refer to Storti and Vitale (2003), and Diongue et al. (2010). Similar to BL-GARCH(1,1) in quadratic nature, a Generalized Quadratic ARCH (GQARCH) model of first order is defined by (cf., Sentana (1995))

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} + \delta_1 \epsilon_{t-1}. \quad (1.4)$$

A comparative study between the aforementioned asymmetric models was carried out by Storti and Vitale (2003) where BL-GARCH(1,1) is shown to have a slight edge over the other models in analyzing NASDAQ data.

This article proposes to combine the “threshold effect” in (1.2) with the “bilinear structure” of (1.3) in order to generate a broader class of asymmetric GARCH models. The resulting model will be referred to as a threshold-bilinear GARCH (TBL-GARCH, for short). A TBL-GARCH(1,1) is characterized by two sources of asymmetry in volatility, viz., threshold-asymmetry and shift-asymmetry. A simulation study in Section 3 shows that the suggested TBL-GARCH(1,1) class is able to capture surprisingly diverse asymmetric features of volatilities h_t which would have been unaccountable without proposing the TBL-GARCH(1,1) class. In Section 2, basic probabilistic properties of the class are investigated including conditions for stationarity and existence of stationary moments up to certain order. Applications to Korean financial time series are illustrated in Section 4 with a view to model comparisons. An extension to higher order models, viz., TBL-GARCH(p, q), is briefly addressed as concluding remarks.

2. The model description and basic probabilistic properties

Incorporating a threshold effect into the BL-GARCH(1,1) model defined in (1.3), we propose the following threshold bilinear GARCH(1,1) time series $\{\epsilon_t\}$ which will be referred to as a TBL-GARCH(1,1), defined by $\epsilon_t = \sqrt{h_t} e_t$ and

$$h_t = \alpha_0 + \alpha_{11} \epsilon_{t-1}^{+2} + \alpha_{12} \epsilon_{t-1}^{-2} + \gamma_{11} \epsilon_{t-1}^+ \sqrt{h_{t-1}} + \gamma_{12} \epsilon_{t-1}^- \sqrt{h_{t-1}} + \beta_1 h_{t-1}, \quad (2.1)$$

where all parameters involved in h_t are collected as $\theta = (\alpha_0, \alpha_{11}, \alpha_{12}, \gamma_{11}, \gamma_{12}, \beta_1)^T$ with T denoting “transpose”. For notational convenience, we use ϵ_{t-1}^{+2} and ϵ_{t-1}^{-2} for denoting $(\epsilon_{t-1}^+)^2$ and $(\epsilon_{t-1}^-)^2$. Here γ_{11} and γ_{12} are permitted to have values in $(-\infty, \infty)$ while $\alpha_0 > 0$ and the other parameters α_{11}, α_{12} and β_1 are, as usual, non-negative. To avoid triviality, we rule out the case when both α_{11} and α_{12} are zero. In particular when $\alpha_{11} = \alpha_{12}$ and $\gamma_{11} = -\gamma_{12}$, the TBL-GARCH(1,1) reduces to BL-GARCH(1,1) which does not exhibit threshold effects. If $\gamma_{11} = 0$ and $\gamma_{12} = 0$ then the TBL-GARCH(1,1) becomes a threshold model (T-GARCH(1,1)) defined in (1.2). Since h_t is a conditional variance, h_t must be strictly positive and thus the following “positivity condition” will be assumed throughout the paper.

(P). For any fixed t , $P(h_t \text{ is strictly positive}) = 1$.

Remarks. The condition (P) requires additional restrictions on θ . It is not easy to derive a closed form of the necessary and sufficient condition for (P) in terms of the parameter $\theta = (\alpha_0, \alpha_{11}, \alpha_{12}, \gamma_{11}, \gamma_{12}, \beta_1)^T$. One can rewrite h_t in the defining Eq. (2.1) as a quadratic form $h_t = a_{t-1}^T A(\theta) a_{t-1}$ where $a_{t-1} = (1, \epsilon_{t-1}^+, \epsilon_{t-1}^-, \sqrt{h_{t-1}})^T$ and $A(\theta)$ is a suitable 4×4 symmetric matrix as a function of θ . A sufficient condition for (P) is that the matrix $A(\theta)$ is positive definite. Similarly to Storti and Vitale (2003), provided $\beta_1 > 0$, it can be shown that a set of sufficient conditions ensuring $A(\theta)$ is positive definite (and thus (P)) is given by

$$(i) 4(\alpha_{11}\alpha_{12} + \alpha_{11}\beta_1 + \alpha_{12}\beta_1) > \gamma_{11}^2 + \gamma_{12}^2 \quad \text{and} \quad (ii) 4\alpha_{11}\alpha_{12}\beta_1 > \alpha_{12}\gamma_{11}^2 + \alpha_{11}\gamma_{12}^2. \quad (2.2)$$

In particular when $\alpha_{11} = \alpha_{12}$ and $\gamma_{11} = -\gamma_{12}$, that is, for the BL-GARCH(1,1) model, $A(\theta)$ is shown to be positive definite and hence (P) is fulfilled if $\gamma_1^2 < 4\alpha_1\beta_1$ (cf. Diongue et al. (2010)) where $\alpha_1 = \alpha_{11} = \alpha_{12}$ and $\gamma_1 = \gamma_{11} = -\gamma_{12}$ in (2.1).



Asymptotics for a class of generalized multicast autoregressive processes

S.Y. Hwang^{a,*}, Kee-Hoon Kang^b

^a Department of Statistics, Sookmyung Women's University, Seoul 140-742, Republic of Korea

^b Department of Statistics, Hankuk University of Foreign Studies, Yongin 449-791, Republic of Korea

ARTICLE INFO

Article history:

Received 18 February 2012

Accepted 17 April 2012

Available online 7 May 2012

AMS 2000 subject classifications:

primary 62M10

secondary 62P10

Keywords:

Branching correlation

Generalized multicast autoregressive model

Martingale estimating functions

Multi-casting tree

Pathwise stationarity

ABSTRACT

We propose a new class of generalized multicast autoregressive (GMCAR, for short, hereafter) models indexed by a multi-casting tree where each individual produces exactly the same number of offspring. This class includes standard bifurcating autoregressive processes (BAR, cf. Cowan and Staudte (1986)) and multicast autoregressive (MCAR, cf. Hwang and Choi (2009)) models as special cases. Accommodating non-Gaussian, non-negative and count data, the class includes various models such as nonlinear autoregression, conditionally heteroscedastic process and conditional exponential family. The pathwise stationarity of the GMCAR model is discussed. A law of large numbers and a central limit theorem are established which are in turn used to derive asymptotic distributions associated with martingale estimating functions.

© 2012 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.

1. Introduction

Bifurcating autoregressive (BAR) processes indexed by a binary-splitting tree have been studied by several authors including Basawa and Zhou (2004), Cowan and Staudte (1986), Huggins and Basawa (1999, 2000), and Huggins and Staudte (1994). In a BAR model, observations are indexed by a bifurcating tree where each individual (mother) in one generation produces two individuals (sisters) in the next generation. For each individual, an observation X is recorded. A first order BAR(1) process $\{X_t, t = 1, 2, \dots\}$ is defined by

$$\begin{aligned} X_{2t} &= \theta X_t + \varepsilon_{2t} \\ X_{2t+1} &= \theta X_t + \varepsilon_{2t+1} \end{aligned} \quad (1.1)$$

where $|\theta| < 1$ and $\{(\varepsilon_{2t}, \varepsilon_{2t+1}), t = 1, 2, \dots\}$ is a sequence of independent and identically distributed (i.i.d.) normal random errors with mean zero and covariance matrix given by

$$\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \sigma^2, \quad -1 < \rho < 1. \quad (1.2)$$

The BAR(1) model has proved useful in modeling, in particular, cell lineage data. Various extensions have been addressed in the literature. Refer to, for instance, Huggins and Basawa (1999) and references therein. In the BAR(1) formulation, it is noted that ρ represents a correlation between sisters, conditionally on their mother. That is,

$$\rho = \text{Corr}(X_{2t}, X_{2t+1} | X_t). \quad (1.3)$$

* Corresponding author.

E-mail address: shwang@sookmyung.ac.kr (S.Y. Hwang).

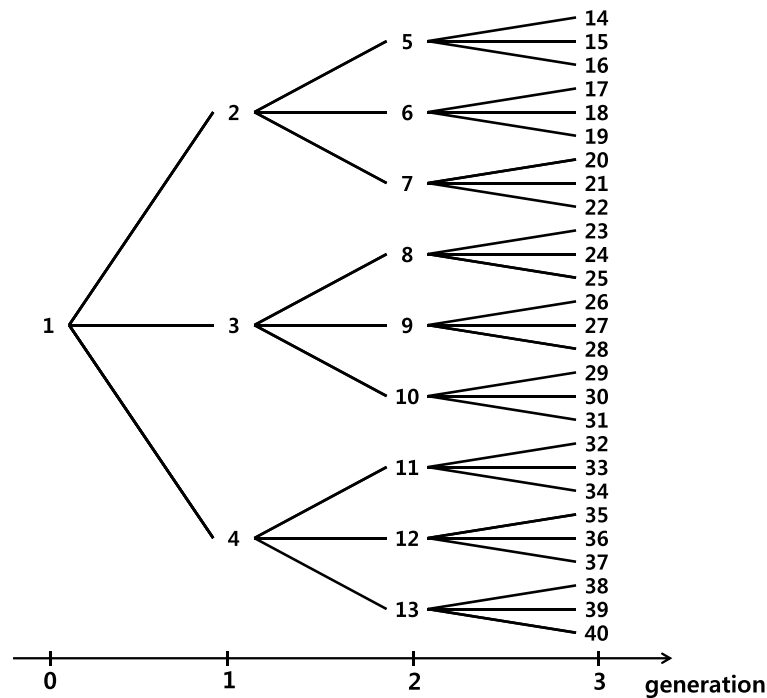


Fig. 1. An example of tri-casting tree ($m = 3$).

In this article, we shall refer to conditional correlation ρ between sisters as a branching correlation. Guyon (2007) extended the BAR(1) structure allowing different distributions on sisters. Hwang and Basawa (2009, 2011) investigated a class of Markov processes indexed by branching processes in which sisters are assumed to be conditionally independent (i.e., branching correlation $\rho = 0$).

Many applications require a multi-casting tree in which each node may have m branches where m may be larger than 2. For instance, in an internet traffic data, a package (signal) may be sent along m links and delay times X over each link may be of interest. As a natural extension of the BAR(1), Hwang and Choi (2009) have recently introduced an AR(1) model defined on a multi-casting tree where each individual gives rise to exactly m -offspring in the next generation. They referred to such an AR(1) as the multicast autoregressive model of order 1 (MCAR(1), for short). Fig. 1 illustrates a tri-casting ($m = 3$) tree consisting of 40 observations and three generations. The MCAR(1) process $\{X_t, t = 1, 2, \dots\}$ is defined by the m -equations ($m \geq 2$)

$$\begin{aligned} X_{mt-(m-2)} &= \theta X_t + \varepsilon_{mt-(m-2)} \\ &\vdots \\ X_{mt} &= \theta X_t + \varepsilon_{mt} \\ X_{mt+1} &= \theta X_t + \varepsilon_{mt+1}, \end{aligned} \quad (1.4)$$

where $\{(\varepsilon_{mt-(m-2)}, \dots, \varepsilon_{mt}, \varepsilon_{mt+1}), t = 1, 2, \dots\}$ is a sequence of i.i.d. m -variate normal random vectors with mean zero vector and covariance matrix V_1 given by

$$V_1 = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \sigma^2. \quad (1.5)$$

Refer to Hwang and Choi (2009) for more details on MCAR(1). Note that the MCAR(1) with $m = 2$ reduces to the BAR(1) model. It is noticed that the MCAR(1) model assumes an equi-branching correlation structure in the sense that correlations between any of the two sisters, conditionally on their mother, are the same as given by ρ . An obvious drawback, however, in V_1 of equi-branching correlation structure is to have to impose a certain restriction on ρ for V_1 to be a positive definite matrix. The required restriction is shown to be $-(m-1)^{-1} < \rho < 1$ (see proposition in Section 2). The BAR(1) model ($m = 2$) identifies the restriction trivially as $-1 < \rho < 1$ while tri-casting case ($m = 3$) requires $-1/2 < \rho < 1$ for a proper covariance matrix V_1 . As m increases, one may essentially exclude negative values of ρ in V_1 . In the next section, we will propose another useful covariance matrices (other than V_1) which do not require any restrictions on ρ .



Arbitrary initial values and random norm for explosive AR(1) processes generated by stationary errors

S.Y. Hwang

Department of Statistics, Sookmyung Women's University, Seoul, Republic of Korea

ARTICLE INFO

Article history:

Received 20 June 2012

Received in revised form 26 August 2012

Accepted 26 August 2012

Available online 7 September 2012

MSC:

primary 62M10

Keywords:

Explosive AR(1)

Initial value

Random norm

Dependent errors

ABSTRACT

This article is concerned with a broad class of explosive AR(1) models. Allowing stationary dependence on the error process, we do not restrict ourselves to independent and identically distributed errors. The model accommodates, as special cases, GARCH errors, AR(1) errors and Gaussian ARMA errors. The error distribution is permitted to be non-normal. To circumvent the effect of initial values, the limit distribution of the least squares estimate using a random norm (rather than a constant norm) is derived. It is shown that the limit distribution using a random norm is free from the initial value provided the error is symmetrically distributed about zero.

© 2012 Elsevier B.V. All rights reserved.

1. Motivation of the study

Consider the following one-sided AR(1) process $\{X_t, t \geq 0\}$ defined recursively by

$$X_t = \theta X_{t-1} + \epsilon_t, \quad t = 1, 2, \dots, \quad (1.1)$$

for which the initial value is fixed as a constant x_0 , that is, $X_0 = x_0$, and the error process $\{\epsilon_t, t \geq 1\}$ is independent and identically distributed (iid) with mean zero and variance σ_ϵ^2 . The marginal distribution of ϵ_t is not necessarily normal. The defining equation (1.1) is understood to mean that X_t is made from θX_{t-1} with an additional error ϵ_t . Based on a sample $X_1, X_2, \dots, X_n$ of size n , the ordinary least squares estimator $\hat{\theta}_n$ of θ is given by

$$\hat{\theta}_n = \frac{\sum_{t=1}^n X_t X_{t-1}}{\sum_{t=1}^n X_{t-1}^2}. \quad (1.2)$$

The limit distribution of $\hat{\theta}_n$ for $|\theta| < 1$ (asymptotically stationary) is well documented as normal, and the unit root case of $\theta = 1$ gives an asymmetric limit distribution as a functional on standard Brownian motion (see e.g., Fuller (1996, Theorem 10.1.1)). The limit distribution of $\hat{\theta}_n$ for $\theta = -1$ is known as a mirror image of the limit distribution for $\theta = 1$ (see Corollary 10.1.1.1 of Fuller (1996)). It is noted that the limit distribution of $\hat{\theta}_n$ for $|\theta| \leq 1$ is free from the initial value x_0 . However, for $|\theta| > 1$, this is not the case any more. Although the initial value x_0 is typically set to be zero in the literature of the explosive case (cf. e.g., Basawa and Brockwell (1984), Monsour and Mikulski (1998), Hwang and Basawa (2005), Hwang et al. (2007)), this article will treat x_0 as an arbitrarily fixed constant.

E-mail address: shwang@sookmyung.ac.kr.

The case of $|\theta| > 1$ is referred to as “explosive AR(1)”. One may rewrite (1.1) as

$$X_t = \epsilon_t + \theta\epsilon_{t-1} + \cdots + \theta^{t-1}\epsilon_1 + \theta^t x_0, \quad (1.3)$$

and thus $E(X_t) = \theta^t x_0$ and $\text{Var}(X_t) = \sigma_\epsilon^2(1 + \theta^2 + \cdots + \theta^{2(t-1)})$. For $|\theta| > 1$, notice that the variance of X_t grows explosively to infinity. Unless $x_0 = 0$, the mean of X_t is also “explosive”. Statistical analysis for the explosive AR(1) model dates back to remarkable early works of White (1958), Anderson (1959) and Rao (1961). It is well documented (cf. e.g., Fuller (1996, Theorem 10.2.1)) that $(\theta^2 - 1)^{-1}\theta^n(\hat{\theta}_n - \theta)$ converges in distribution to a certain random variable whose distribution depends on the initial value x_0 . In particular, when $x_0 = 0$ and ϵ_t is normal, the limit distribution of $(\theta^2 - 1)^{-1}\theta^n(\hat{\theta}_n - \theta)$ reduces to a standard Cauchy distribution. For non-zero initial values x_0 and normally distributed ϵ_t , White (1958, Eq. 4.25) identified the limit distribution of $(\theta^2 - 1)^{-1}\theta^n(\hat{\theta}_n - \theta)$ which clearly depends on the initial value x_0 . Admitting that the information is contained only in the sample $X_1, X_2, \dots, X_n$, it is reasonable to say that we do not know the initial value x_0 of the one-sided AR(1) process (1.1). Suppose that we are willing to use $\hat{\theta}_n$ as a pivotal value for conducting inference concerning θ . For $|\theta| \leq 1$, the initial value does not affect the limit distribution of $\hat{\theta}_n$, as discussed above, and thus inference based on $\hat{\theta}_n$ continues to hold regardless of x_0 . For the explosive case of $|\theta| > 1$, however, asymptotic inference based on $\hat{\theta}_n$ suffers from unknown x_0 as long as we use a non-random norm $(\theta^2 - 1)^{-1}\theta^n$ for $\hat{\theta}_n$. This article proposes using a random norm in order to circumvent the problem of involving unknown x_0 in asymptotic distributions of $\hat{\theta}_n$, facilitating statistical inference based on $\hat{\theta}_n$.

The novelty of the paper is twofold. First, allowing stationary dependence structures among the error process $\{\epsilon_t\}$, we will not confine ourselves to the case of iid errors. Anderson (1959) investigated uncorrelated sequences of $\{\epsilon_t\}$. Hwang et al. (2007) discussed the limit distribution of $(\theta^2 - 1)^{-1}\theta^n(\hat{\theta}_n - \theta)$ when $x_0 = 0$ and $\{\epsilon_t\}$ is a sequence of autoregressive conditionally heteroscedastic (ARCH) errors. This article aims to extend the literature to explosive AR(1) processes generated by stationary (dependent) errors. Second, treating the initial value x_0 as an arbitrarily fixed constant, the limit distribution of $\hat{\theta}_n$ using a random norm is derived, which is shown to be free from the initial value x_0 provided the error ϵ_t is symmetrically distributed about zero. The limit distribution is verified to be a mixture of two random variables for which a mixing variable is binary, taking values -1 and 1 according to the signs of a certain random variable. See Theorem 1. Some statistical applications are briefly illustrated.

2. The model description and preliminary results

This article is concerned with the following explosive one-sided AR(1) process $\{X_t, t \geq 0\}$ generated by a stationary error process. Specifically, $\{X_t\}$ is defined by

$$X_t = \theta X_{t-1} + \epsilon_t, \quad t = 1, 2, \dots, |\theta| > 1, \quad (2.1)$$

where the initial value $X_0 = x_0$ and the error process $\{\epsilon_t, t \geq 1\}$ is a zero-mean strictly stationary and ergodic time series having the auto-covariance function (ACVF) $\gamma_\epsilon(k) = \text{Cov}(\epsilon_t, \epsilon_{t+k})$, $k = 0, 1, 2, \dots$. Note that $\gamma_\epsilon(-k) = \gamma_\epsilon(k)$. It will be assumed throughout that $\gamma_\epsilon(k)$ is of a short memory (absolutely summable), that is, $\sum_{k=0}^{\infty} |\gamma_\epsilon(k)| < \infty$. Some interesting examples for $\{\epsilon_t\}$ are given below.

Example 1: iid errors. Consider the iid sequence $\{\epsilon_t\}$ with mean zero and variance σ_ϵ^2 , denoted by “iid $(0, \sigma_\epsilon^2)$ ”. In particular, when the distribution of ϵ_t is normal, that is, $\{\epsilon_t\}$ is iid following $N(0, \sigma_\epsilon^2)$, we shall use the notation “nid $(0, \sigma_\epsilon^2)$ ” emphasizing “normal”.

Example 2: martingale differences. A sigma field generated by $X_t, X_{t-1}, \dots, X_1$ is denoted by F_t . Due to one to one correspondence between $\{X_t\}$ and $\{\epsilon_t\}$ in the defining equation (2.1), F_t denotes also the sigma field generated by $\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_1$. By convention, F_0 denotes a trivial sigma field. Consider a sequence of martingale differences $\{\epsilon_t\}$ so that $E(\epsilon_t | F_{t-1}) = 0$. Note that $\{\epsilon_t\}$ constitutes an uncorrelated sequence and thus $\gamma_\epsilon(k) = 0$, $k \neq 0$. The GARCH class is a typical example of martingale differences. A GARCH(1, 1) process $\{\epsilon_t\}$ is defined by

$$\epsilon_t = \sqrt{h_t} e_t$$

and

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1},$$

where $\{e_t\}$ is iid $(0, 1)$ and the conditional variance $h_t = \text{Var}(\epsilon_t | F_{t-1})$ is referred to as a volatility at time point t . For various GARCH-type models, see for instance Hwang and Baek (2009).

Example 3: AR(1) errors. Consider the following stationary AR(1) errors $\{\epsilon_t\}$ defined by

$$\epsilon_t = \phi \epsilon_{t-1} + \eta_t, \quad |\phi| < 1, \quad (2.2)$$

where $\{\eta_t\}$ is iid $(0, \sigma_\eta^2)$. It is noted that $\gamma_\epsilon(k) = \phi^k \cdot \sigma_\eta^2 / (1 - \phi^2)$, $k \geq 0$.

Example 4: MA(1) errors. Let $\{\epsilon_t\}$ be obtained recursively by

$$\epsilon_t = \eta_t - \delta \eta_{t-1}, \quad |\delta| < 1, \quad (2.3)$$

where $\{\eta_t\}$ is iid $(0, \sigma_\eta^2)$. It is clear that $\gamma_\epsilon(0) = \sigma_\eta^2(1 + \delta^2)$, $\gamma_\epsilon(1) = -\delta \gamma_\epsilon(0)$ and $\gamma_\epsilon(k) = 0$ for all $k \geq 2$. Omitting details, one may extend (2.2) and (2.3) to strictly stationary ARMA errors (cf. e.g., Fuller (1996, p. 70)).

Non-ergodic martingale estimating functions and related asymptotics

S.Y. Hwang^{a*}, I.V. Basawa^b, M.S. Choi^a and S.D. Lee^c

^aDepartment of Statistics, Sookmyung Women's University, Seoul, South Korea; ^bDepartment of Statistics, University of Georgia, Athens, GA, USA; ^cDepartment of Statistics, Chungbuk National University, Cheongju, South Korea

(Received 17 September 2011; final version received 24 September 2012)

Well-known estimation methods such as conditional least squares, quasilielihood and maximum likelihood (ML) can be unified via a single framework of martingale estimating functions (MEFs). Asymptotic distributions of estimates for ergodic processes use constant norm (e.g. square root of the sample size) for asymptotic normality. For certain non-ergodic-type applications, however, such as explosive autoregression and super-critical branching processes, one needs a *random norm* in order to get normal limit distributions. In this paper, we are concerned with non-ergodic processes and investigate limit distributions for a broad class of MEFs. Asymptotic optimality (within a certain class of non-ergodic MEFs) of the ML estimate is deduced via establishing a convolution theorem using a random norm. Applications to non-ergodic autoregressive processes, generalized autoregressive conditional heteroscedastic-type processes, and super-critical branching processes are discussed. Asymptotic optimality in terms of the maximum *random limiting power* regarding large sample tests is briefly discussed.

Keywords: non-ergodic process; martingale estimating functions; branching process; asymptotic optimality; explosive autoregressive model

2000 MSC Classification: Primary 62M10

1. Introduction

Estimating functions corresponding to standard estimation methods such as conditional least squares (CLS), quasilielihood (QL), and maximum likelihood (ML) are special cases of martingale estimating functions (MEF, hereafter) (refer to Section 2 to follow; see, for instance, [1]). Consistent and asymptotically normal estimators are usually obtained by equating MEFs to zero, that is, by solving the martingale estimating equations (MEEs, for short). Most of the literature on MEF (or MEE) asymptotics has been directed to ergodic processes for which a *non-random norm* is used to get asymptotic normal distributions of the estimates. For ergodic stationary processes, a constant norm $\sqrt{n}$ (throughout the paper, n denotes the sample size) is often used to get asymptotic normal distributions. For a comprehensive discussion of estimating functions from ergodic processes, see among others [1–5]. In particular, refer to [6] for a rigorous treatment of MEFs from a general ergodic Markov processes. On the other hand, for non-ergodic-type processes (refer to [7, Preface] for a definition of non-ergodic process), limit distributions of standard

*Corresponding author. Email: shwang@sm.ac.kr

estimators (including CLS, QL, and ML) are mixed normal when a non-random norm is used. Instead, a *random norm* is required to get normal limit distributions for non-ergodic processes. Asymptotics of various statistics normalized by random norms in a broad context have recently been discussed by Pena *et al.* [8]. The non-ergodic process is distinguished from ergodic process by the random norm (other than non-random norm) required to get normal limits. Refer to [7] for various non-ergodic processes including normal mixture models, explosive autoregressive processes, birth processes, branching processes, and non-stationary diffusion processes. Some specific examples of non-ergodic processes are discussed in detail in Section 4.

As commented in [3], rigorous treatments of asymptotics on MEFs in the context of non-ergodic processes have not yet been adequately explored in the literature. In this paper, we consider large sample inference based on MEFs for a class of non-ergodic processes. Under a mild set of regularity conditions, asymptotic distributions of estimators obtained by solving MEEs are derived. Furthermore, we show that the maximum-likelihood estimator (MLE) is asymptotically optimum in a sense of ‘minimum variance’ within a class of estimators obtained from MEEs. We use a natural random norm in our asymptotics. A convolution theorem (using a random norm) is established which yields the desired optimality (refer to Theorems 3 and 4). These results are illustrated with specific applications to non-ergodic processes. Section 2 introduces a general class of MEFs and derives the limit distributions of the estimators therefrom. The convolution theorem for non-ergodic MEFs is established in Section 3. Three specific applications (one ergodic and two non-ergodic models) are discussed in Section 4. Section 5 addresses large sample tests based on MEFs and asymptotic optimality of tests is discussed in terms of the ‘maximum’ *random limiting power*. One of the main goals is to provide a simplified but a broad treatment of asymptotic theory on MEFs from non-ergodic processes.

2. MEFs and asymptotic distributions

Let $\{X_t, t = 0, 1, 2, \dots\}$ denote a possibly non-ergodic stochastic process with initial value $X_0 = x_0$. Suppose that the probability measure associated with $\{X_t\}$ is indexed by a $(k \times 1)$ vector parameter θ taking values in Θ which is an open subset of the k -dimensional Euclidian space R^k . Based on a sample of observations $X_1, X_2, \dots, X_n$, we are interested in estimating the parameter vector θ . Let F_t denote the σ -field generated by $X_t, X_{t-1}, \dots, X_1$. Use $p_t(\theta)$ to denote the conditional density of X_t given F_{t-1} and consider the following MEF $U_n(\theta)$ given by

$$U_n(\theta) = \sum_{t=1}^n u_t(\theta) : (k \times 1) \text{ vector}, \quad (1)$$

where $\{u_t(\theta)\}$ is a sequence of martingale differences satisfying $E(u_t(\theta)|F_{t-1}) = 0$. It will be assumed throughout that the $(k \times 1)$ vector $u_t(\theta)$ is differentiable (with respect to θ) and differentiability under integral sign is valid. Denote by ξ_n the sum of conditional covariance matrices corresponding to $u_t(\theta)$, that is,

$$\xi_n = \sum_{t=1}^n \text{Var}(u_t(\theta)|F_{t-1}) = \sum_{t=1}^n E(u_t(\theta)u_t^T(\theta)|F_{t-1}) : (k \times k) \text{ matrix}. \quad (2)$$

We suppress the dependence of ξ_n on θ whenever necessary, that is, $\xi_n = \xi_n(\theta)$, for notational simplicity. In addition, ξ_n is assumed to be positive definite almost surely. Throughout, the matrix (or vector) norm will be denoted by $\|\cdot\|$, viz., for any matrix A , $\|A\|^2 = \text{tr}(A^T A) = \text{tr}(A A^T)$. Here, A^T is the transpose of A . For simplicity, for any non-singular matrix A , A^{-T} is used for denoting $A^{-T} = (A^{-1})^T = (A^T)^{-1}$. The identity matrix of order k is denoted by I_k . In addition, for a positive

재충전이 있는 연속시간 리스크 모형에서 파산확률 연구

고한나¹ · 최승경² · 이의용³

¹숙명여자대학교 통계학과, ²숙명여자대학교 통계학과, ³숙명여자대학교 통계학과

(2011년 9월 30일 접수, 2011년 12월 23일 수정, 2011년 12월 26일 채택)

요 약

재충전이 있는 연속시간 리스크 모형이 고려된다. 프리미엄은 일정한 율로 들어오고, 보험금 청구는 복합 포아송 과정을 따라 이루어진다. 초기 잉여금 $u > 0$ 로 시작하여 잉여금은 프리미엄에 의해 증가하고 보험금 청구에 의해 감소한다. 잉여금의 수준이 τ ($0 < \tau < u$) 아래로 떨어지면 초기 잉여금 수준까지 재충전이 이루어진다고 가정한다. 재충전이 고려된 리스크 모형에서 잉여금이 없어지는 파산확률을 적미분 방정식을 통해 유도하고, 보험 청구액이 독립적으로 지수분포를 따르는 경우는 파산확률의 명확한 공식이 유도됨을 보인다.

주요용어: 리스크 모형, 잉여금 과정, 파산확률, 재충전, 적미분 방정식.

1. 서론

보험 수리에서 다루는 일반적인 리스크 모형은 다음과 같다. 보험 상품의 잉여금은 초기치 $u > 0$ 에서 시작하여 단위 시간당 $c > 0$ 로 들어오는 보험료에 의해 증가하며 고객의 보험금 청구가 발생할 때마다 감소한다. 여기서, 보험금 청구는 발생률 $\lambda > 0$ 인 포아송 과정 $\{N(t), t \geq 0\}$ 을 따라 발생하며, 지급되는 보험금의 크기는 서로 독립이고 평균 $\mu > 0$ 인 일반적인 분포 F 를 따른다. 시간 $t > 0$ 에서 잉여금을 수식으로 표현하면 다음과 같다.

$$U(t) = u + ct - \sum_{i=1}^{N(t)} X_i, \quad t \geq 0.$$

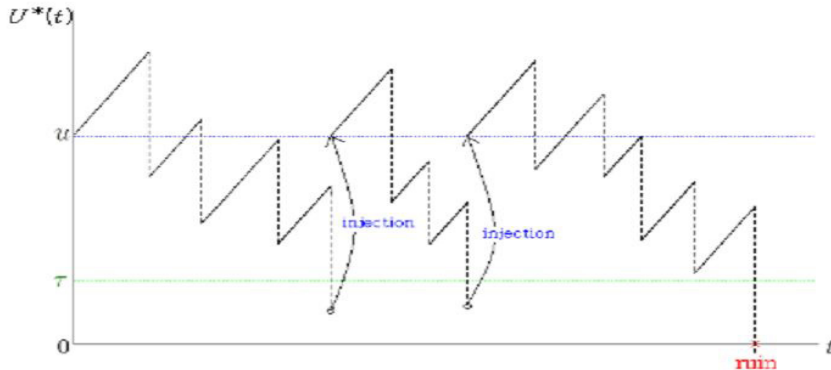
여기서, X_i 는 i 번째 보험 청구액이고, $c = (1 + \theta)\lambda\mu$ 이다. $\theta > 0$ 는 상대적인 보완 부가금(relative security loading, premium loading factor)으로 이로 인해 보험료를 c 가 단위 시간당 기대되는 보험 청구액 $\lambda\mu$ 보다 커진다. 이와 같은 보험 상품의 잉여금 과정에서 $U(t)$ 의 값이 0보다 작아지면 보험 상품이 파산한다고 말한다.

잉여금 과정의 모형화와 파산확률을 구하는 기존의 연구는 Klugman 등 (2004)에 잘 요약되어 있다. Dufresne와 Gerber (1991)는 확산(diffusion) 과정이 더해진 리스크 모형에서 잉여금의 파산확률을 연구하였다. 파산할 때까지의 시간에 대한 연구는 Gerber (1990)와 Gerber와 Shiu (1997) 등에 의해 연구되었고, Dickson과 Willmot (2005)은 라플라스 역변환을 통해 파산할 때까지 시간의 확률밀도함수를 명확하게 구하였다. Oh 등 (2007)은 마팅게일 이론을 이용하여 잉여금이 주어진 범위를 벗어나는데 걸

본 연구는 숙명여자대학교 2010학년도 교내연구비 지원에 의해 수행되었음.

³교신저자: (140-742) 서울특별시 용산구 청파동 2가, 숙명여자대학교 통계학과, 교수.

E-mail: eylee@sookmyung.ac.kr

그림 1.1. $U^*(t)$ 의 표본경로

리는 시간과 이 시간 동안의 총 잉여금의 기대값을 구하였고, Jeong 등 (2009)은 기존의 리스크 모형에 잉여금이 충분할 때 투자(investment)와 잉여금이 부족할 때 재충전(injection)의 개념을 도입하여 잉여금의 최적 투자 정책을 연구하였다.

재충전은 기존의 리스크 모형에서 잉여금의 파산확률을 줄이는 한 방법이다. 본 논문에서는 재충전이 고려된 리스크 모형에서 잉여금의 파산확률을 연구한다. 즉, 적정수준 τ ($0 < \tau < u$)를 설정하여 τ 아래로 잉여금이 떨어지면 즉시 u 까지 재충전하는 잉여금 과정을 고려한다. 이러한 리스크 모형에서 잉여금 과정을 $\{U^*(t), t \geq 0\}$ 라 놓으면 표본 경로는 그림 1.1과 같다. 2장에서는 기존의 리스크 모형에서 연구된 파산확률을 요약하고, 재충전이 고려된 리스크 모형에서 잉여금의 파산확률을 적미분 방정식(integro-differential equation)을 세워 유도한다. 3장에서는 보험 청구액이 독립적으로 지수분포를 따르는 경우에 파산확률의 명확한 공식을 구하고, 수치적인 예를 통해 파산확률의 변화를 관찰한다.

2. 재충전이 있는 리스크 모형에서 파산확률

우선, Klugman 등 (2004)을 참조하여 기존의 리스크 모형에서 잉여금의 파산확률을 요약하면 다음과 같다. 초기치 u 에서 시작한 잉여금의 파산확률 $\psi(u)$ 는

$$\psi(u) = \Pr\{U(t) < 0, \text{ for some } t > 0 | U(0) = u\}$$

이다. Klugman 등 (2004, pp. 239–242)에서와 같이 최대 총 손실(maximum aggregate loss)을 이용하여 파산확률 $\psi(u)$ 를 구하면

$$\psi(u) = \sum_{k=1}^{\infty} \frac{\theta}{1+\theta} \left(\frac{1}{1+\theta} \right)^k S_e^{(k)}(u), \quad u \geq 0$$

이 된다. 여기서 $S_e^{(k)}(u) = 1 - F_e^{(k)}(u)$ 이고, $F_e^{(k)}$ 는 F_e 의 k 차 공률(k -fold recursive convolution)이고, F_e 는 F 의 평형분포함수(equilibrium distribution function)로

$$F_e(u) = \frac{1}{\mu} \int_0^u [1 - F(x)] dx$$

이다. 보험금 크기의 분포 F 가 평균이 μ 인 지수분포인 경우, 즉, $F(x) = 1 - e^{-x/\mu}$, $x \geq 0$, 잉여금의

Ruin Probabilities in a Risk Model with Two Types of Claims

Ji-Yeon Han¹ · Seung Kyoung Choi² · Eui Yong Lee³

¹Department of Statistics, Sookmyung Women's University

²Department of Statistics, Sookmyung Women's University

³Department of Statistics, Sookmyung Women's University

(Received June 12, 2012; Revised July 11, 2012; Accepted August 28, 2012)

Abstract

A surplus process with two types of claims is considered, where Type I claims occur more frequently, however, their sizes are smaller stochastically than Type II claims. The ruin probabilities of the surplus caused by each type of claim are obtained by establishing integro-differential equations for the ruin probabilities. The formulas of the ruin probabilities contain an infinite sum and convolutions that make the formulas hard to be applicable in practice; subsequently, we obtain explicit formulas for the ruin probabilities when the sizes of both types of claims are exponentially distributed. Finally, we show through a numerical example, that Type II claims have more impact on the ruin probability of the surplus than Type I claims.

Keywords: Continuous time risk model, surplus process with two types of claims, ruin probability, integro-differential equation.

1. Introduction

In this paper, we consider a variation of the classical continuous time risk model. The surplus in the model is initially at $u > 0$, thereafter it increases linearly at a rate $c > 0$ due to incoming premiums and decreases jumpwise due to arriving claims. There are two types of claims in the model. Type I claims arrive according to a Poisson process of rate $\lambda_1 > 0$ and sizes of claims are *i.i.d.* with distribution G , meanwhile, Type II claims arrive according to another Poisson process of rate $\lambda_2 > 0$ and sizes of claims are *i.i.d.* with distribution H . We assume that $\lambda_1 \geq \lambda_2$ and distribution H is stochastically larger than distribution G . That is, Type I claims occur more frequently, however, their sizes are smaller stochastically than Type II claims.

The surplus process of the model can be written as

$$U(t) = u + ct - \sum_{i=1}^{N_1(t)} Y_i - \sum_{j=1}^{N_2(t)} Z_j,$$

This research was supported by Sookmyung Women's University Research Grants 2011.

³Corresponding author: Professor, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea. E-mail: eylee@sookmyung.ac.kr

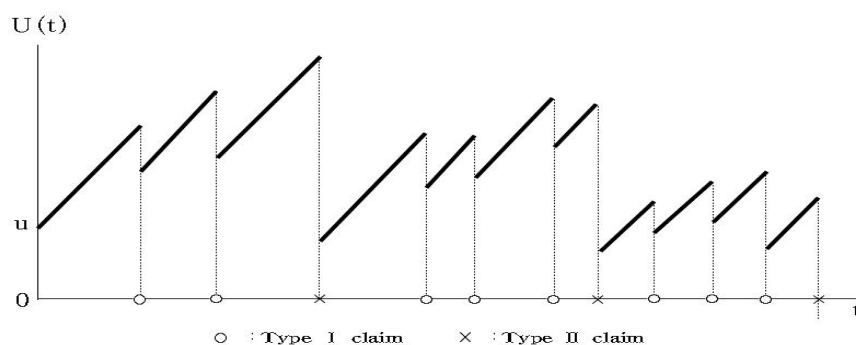


Figure 1.1. A sample path of $\{U(t), t \geq 0\}$

where $U(t)$ is the surplus at time $t > 0$, $N_1(t)$ and $N_2(t)$ are the numbers of Type I and Type II claims up to time t , Y_i and Z_j denote the sizes of the i^{th} Type I and the j^{th} Type II claims with means μ_1 and μ_2 , where $0 < \mu_1 \leq \mu_2$, and $c = (1 + \theta)(\lambda_1\mu_1 + \lambda_2\mu_2)$ is the premium rate, where $\theta > 0$ is the relative security loading. A sample path of the surplus process in the model is illustrated in Figure 1.1. A ruin is said to occur if the surplus becomes zero or negative.

The ruin probabilities of the continuous time surplus process and its variations have been studied by many authors. The studies on the ruin probability of the surplus in the classical risk model is well summarized in Klugman *et al.* (2004). However, the ruin probability of the surplus even in the classical risk model contains an infinite sum and convolutions, hence, many approximations have been driven and well-known formulas were summarized and compared in Grandell (2000). Dufresne and Gerber (1991) considered a continuous time risk model where the surplus process was perturbed by diffusion and obtained the ruin probabilities. Tsai (2009) studied the ordering of the ruin probabilities in a similar model and obtained bounds on the ruin probabilities.

Recently, the risk model where there exist different types of claims was studied by several researchers. Chan *et al.* (2003) and Guo *et al.* (2007) studied the ruin probabilities in two-dimensional risk models. Chan *et al.* (2003) considered a bivariate risk model and derived a partial integro-differential equation satisfied by the two-dimensional ruin probabilities; however, the exact solution of the equation had not been obtained. Guo *et al.* (2007) considered two separate surplus processes and obtained a bound on the ruin probability of the sum of two surplus processes and bounds on the two-dimensional ruin probabilities of two separate surplus processes.

Li and Garrido (2005) and Lv *et al.* (2010) studied the ruin probabilities of the surplus in the risk model where there are two types of claims; claims of one type arrive according to a Poisson process and claims of the other type arrive according to a renewal process. Li and Garrido (2005) derived a Laplace transform of the non-ruin probability and obtained explicit results when the initial surplus is zero. Lv *et al.* (2010) obtained an exponential bound on the ruin probability. The formulas or the bounds of the ruin probability of the whole surplus process in the risk models have been studied by many authors; however, the exact ruin probabilities caused by each type of claim have not been derived yet. We, in this paper, obtain the formulas of the ruin probabilities caused by each type of claim by establishing integro-differential equations for the ruin probabilities.

In Section 2, we obtain the Laplace transforms of the ruin probabilities from the integro-differential equations and find the formulas of the ruin probabilities by inverting the Laplace transforms. These



Development of differential criteria on tongue coating thickness in tongue diagnosis[☆]

Jinsung Kim^{a,*}, Ga-Jin Han^a, Bin-Hye Choi^a, Jae-Woo Park^a, Kyungmo Park^b, In-Kwon Yeo^c, Bong-Ha Ryu^a

^a Department of Gastroenterology, College of Oriental Medicine, Kyung Hee University, Seoul, Republic of Korea

^b Department of Biomedical Engineering, School of Electronics and Information, Kyung Hee University, Yongin, Kyunggi-do, Republic of Korea

^c Department of Statistics, Sookmyung Women's University, Seoul, Republic of Korea

Available online 21 April 2012

KEYWORDS

Tongue diagnosis;
Tongue coating;
Diagnostic criteria;
Digital tongue
imaging system;
Proportional odds
model

Summary

Objective: To develop differential criteria on the tongue coating thickness (TCT), and especially propose the standard for judgment on thin and thick coating in tongue diagnosis.

Methods: Sixty oriental medical doctors evaluated the TCT (none, thin, and thick coating) in 50 realistic tongue photographs revealing from tip-to-root. The photographs were obtained with a digital tongue imaging system (DTIS) which measured the percentages of tongue coating on the tongue surface. We calculated the match rate which is the ratio in which the assessor's judgment is consistent with the decision standard on the tongue coating, and then 24 assessors ($\geq 80\%$ match rate) were selected to improve the reliability of the decision. The agreement level among 24 assessors was examined to assess the inter-rater reliability. The correlation between TCT judgments and DTIS-measured values was examined to ascertain the reliability of DTIS measurements. Finally, the assumption probability for the analysis of quantified characteristics of the tongue coating was calculated with a proportional odds model.

Results: The inter-rater reliability was assessed as moderate ($\kappa = 0.56$) among 24 assessors, the level of correlation between TCT judgments and DTIS measurements in 24 assessors was relatively high ($0.76, p < 0.01$). As the analysis of the proportional odds model, 29.06% was a cut-off point to separate no coating and thin coating, 63.51% was a cut-off point to separate thin and thick coating.

Conclusions: The differential criteria for TCT in tongue diagnosis were suggested, and particularly thick coating is defined as one that tongue coating which tongue body is invisible, occupy approximately more than two third areas on the tongue surface.

© 2012 Elsevier Ltd. All rights reserved.

[☆] This research was supported by a grant KHU-20091418 from the Kyung Hee University Program for the Young Researcher in Medical Science.

* Corresponding author at: Department of Gastroenterology, Kyung Hee University Oriental Medicine Hospital, 1 Hoegi-dong, Dongdaemun-gu, Seoul 130-702, Republic of Korea. Tel.: +82 2 958 8895; fax: +82 2 958 8895.

E-mail address: oridoc@khu.ac.kr (J. Kim).

Introduction

Tongue diagnosis is a useful method to examine the physiologic function and pathologic changes of the human body in Oriental Medicine (OM).¹ It consists of visually inspecting the tongue body for color, shape, moisture, and movement and assessing tongue coating for color, thickness, distribution, and characteristics at the root.² The tongue coating reflects the nature and location of the pathogenic factors, and often indicates the state of some organs, particularly stomach function.³ Therefore, fastidious observation on the alteration of the tongue coating is particularly required in the course of the diagnosis and the progress of gastrointestinal diseases.

The conventional way for tongue diagnosis, including the evaluation of the tongue coating, is to look at the tongue (from tip-to-root) directly by observers. The tongue coating thickness (TCT) is expressed as no coating, thin coating, and thick coating. No coating is literally the absence of tongue coating, and the dividing standard between thin and thick coating is whether or not the bottom is visible (thin coating, through the tongue coating, one can vaguely see the tongue body; thick coating: through the tongue coating, one cannot see the tongue body).³ However, these judging criteria is insufficient and vague. Moreover, there is no distinct standard with respect to the distribution of the coating on the tongue surface. Therefore, in order to improve the low levels of inter- and intra-practitioner agreements, it is required that adequate operational definitions of the tongue characteristics including tongue coating.^{4,5} The inter-practitioner consistency for the TCT was partially investigated in the previous studies,^{4,6} but there is no precise operational definitions concerning the judging criteria for the TCT.

A digital tongue imaging system (DTIS) was designed to obtain a digital image of the tongue surface, segment this into regions and extract the areas corresponding to the tongue coating.⁷ We showed the new possibility that the tongue coating can be quantitatively evaluated using the DTIS.⁸ By starting from the percentage of the tongue coating area among the tongue surface analyzed by the DTIS, we attempted to suggest the differential criteria on the TCT compared with three diagnostic endpoints (no coating, thin coating, and thick coating). First, we verified the 60 assessors as to how well they evaluated the TCT using the conventional method of tongue diagnosis, we selected reliable assessors and investigated inter-assessor reliability, then confirmed the correlation between the TCT judgments and the DTIS-measured values (percentages of tongue coating). Finally, a proportional odds model was applied to calculate the assumption probability for the analysis of quantified characteristics of the tongue coating.⁹ The aim of this study was to develop differential criteria for the TCT, and especially propose the standard for judgment on thin coating and thick coating with the amount of tongue coating in tongue diagnosis.

Materials and methods

Tongue photographs and DTIS

We prepared 50 tongue photographs revealing the entire tongue surface from tip-to-root. The photographs had high

resolution which selected based on the clarity and reality of the image, depiction of different diagnostic characteristics, and diverse aspects of the tongue coating distribution. We used the DTIS to acquire the tongue photographs under the same conditions and to measure the tongue coating quantitatively.⁸

The DTIS consists of the exterior part with an interface for the patient's face, image acquisition part with a high-resolution digital camera, illumination part, and software part (Fig. 1). In particular, the interface of the exterior part was designed to block out outside light, and the DTIS software controls image acquisition, locating an appropriate region of the tongue, isolating this from the background using a snake algorithm, and extracting the areas of the tongue coating from the tongue images.^{7,10,11} The tongue surface on the DTIS tongue images were divided into six sections, as the method of Winkel et al.,¹² and calculated the percentage of the tongue coating area for each section, and for the tongue as a whole (Fig. 2). Informed consent was obtained from each patient before taking their tongue photographs, as approved by the Institutional Review Board of the Kyung Hee University Oriental Medicine Hospital.

Assessors and procedures

Sixty assessors were recruited to evaluate the tongue coating in the study. They graduated from 6-year programs of Oriental medical colleges in Korea, and were working for the Kyung Hee University Oriental Medicine Hospital as clinicians. They were reminded about the evaluation method such as how to differentiate the TCT and educated on how to divide the tongue surface into 6 areas before participating in the study. The procedures were performed as two sessions. In the first session, the assessors rated the tongue coating as no coating = 0, thin coating = 1, and thick coating = 2, and rated it again in each of 6 areas with the same method in the second session (the results in the 2nd session will be used in the subsequent study on the relationship between the decision for TCT and the regions of the tongue coating). All data collection was conducted at the same location under identical conditions over 50 min. These photographs were presented with a picture viewer software on the LCD monitor (47" HD, Samsung, Korea). All assessments were independent, and no assessors were color blind.

Data processing and assessors selection

The features of the TCT judgments on the 50 tongue photographs were computed as frequency and mode according to the classification of tongue coating and the DTIS-measured values. We regarded the highest number of the assessors' choice (mode) as the decision standard. In order to select reliable assessors, we calculated the match rate, which is the ratio that the assessor's judgment is consistent with the decision standard on the TCT (number of TCT judgments/50 photographs [%]). The reliable assessors were considered to be $\geq 80\%$ match rates. The agreement level of the assessors' judgment for each tongue photograph was calculated as an agreement rate (number of TCT judgments/reliable assessors [%]), and investigated the inter-rater reliability among the selected reliable

Statistical Interpretation of Economic Bubbles

In-Kwon Yeo¹

¹Department of Statistics, Sookmyung Women's University

(Received August 27, 2012; Revised October 11, 2012; Accepted November 6, 2012)

Abstract

In this paper, we propose a statistic to measure investor sentiment. It is a usual phenomenon that an asymmetric volatility (referred to as the leverage effect) is observed in financial time series and is more sensitive to bad news rather than good news. In a bubble state, investors tend to continuously speculate on financial instruments because of optimism about the future; subsequently, prices tend to abnormally increase for a long time. Estimators of the transformation parameter and the skewness based on Yeo-Johnson transformed GARCH models are employed to check whether a bubble or abnormality exist. We verify the appropriacy of the proposed interpretation through analyses of KOSPI and NIKKEI.

Keywords: GARCH model, leverage effect, skewness, Yeo-Johnson transformation.

1. 서론

어떤 자산이나 물건이 내재가치보다 현저히 높은 가격으로 거래되는 경우 일반적으로 가격에 거품 또는 버블(bubble)이 끼었다고 한다. Garber (1990)의 연구에 의하면 17세기 네덜란드에서 벌어진 ‘튤립마니아(Tulip mania)’ 파동이 최초의 대규모 거품현상으로 기록되고 있다. 당시 네덜란드에 수입된 튤립의 구근에 대한 사재기와 선물거래에 의해 가격이 급격히 상승하면서 1630년대 튤립 뿌리 하나가 숙련된 장인의 연간 소득보다 10배 높은 가격으로 팔리기도 했다고 한다. 금융에서의 거품라는 용어는 1711년 설립된 남해회사(South Sea Company)의 주식에 대한 투기로 형성된 ‘남해거품(South Sea Bubble)’에서 어원을 찾고 있으며 프랑스의 미시시피 버블(Mississippi bubble)이 비슷한 시기의 경제 거품으로 기록되고 있다. 이후 크고 작은 거품들이 간헐적으로 발생하였으며 일본에서는 1950년대부터 시작된 장기 호황과 1980년대 저축 억제정책으로 인한 저금리 정책으로 자금의 유동성이 커지고 미국과의 플라자 합의 후 금융완화법을 제정하여 기업들이 돈을 더 쉽게 빌릴 수 있도록 만들면서 1980년대 말부터 부동산가격과 주가가 급격히 상승하였다. 또한 인터넷의 인프라가 구축되기 시작하면서 관련 분야의 급성장과 관련 산업 주가의 급속한 상승으로 1995년부터 2000년까지 IT 버블이라고도 불리는 닷컴 버블(dot-com bubble)이 발생하였다. 미국에서는 신용이 가장 낮은 사람들을 상대로 고금리로 대출해주는 주택담보대출 서브프라임 모기지론(subprime mortgage loan)이 주택가격의 상승으로 높은 수익을 얻자 세계적 금융업체와 헤지펀드 등이 부동산에 대규모로 투자하면서 부동산 거품을 형성하기 시작하였다.

This research was supported by the Sookmyung Women's University Research Grants 2011.

¹Professor, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea.

E-mail: inkwon@sm.ac.kr

문제는 거품이 꺼지면 거품이 형성되는 동안 누적된 비정상적인 투자자산이 급격히 붕괴되면서 원금 손실 및 금융파산, 경기침체 등의 경제적 위기가 장기간 지속될 수 있다는 것이다. 튜립파동 때 가격이 하락세로 반전하면서 파산자가 속출하였고 이 파동은 그 당시 동인도 무역을 통해 막대한 부를 축적하며 실물경제와 금융경제에서 세계 최고 위치에 있었던 네덜란드가 영국에게 경제대국의 자리를 넘겨 준 결정적 요인 중 하나가 되기도 하였다. 또한 1720년 남해거품이 꺼지면서 대규모 금융파산을 초래하였는데 이 회사의 거품 붕괴 때 물리학자 아이작 뉴턴도 막대한 손해를 보았는데 Spence (1820, p.368)에 의하면 뉴턴은 “나는 천체의 움직임은 계산할 수 있지만 인간의 광기는 짐작조차 할 수 없었다.”라는 말을 남겼다고 한다. 일본에서 1990년 주식 가격과 부동산의 거품이 꺼지면서 수많은 기업과 은행이 도산하였고 1991년부터 2002년까지 ‘잃어버린 10년’이라고 불리는 장기침체가 진행되었으며 현재까지도 그 영향을 받고 있다. 미국의 집값이 하락하면서 서브프라임 모기지 대출자들이 대출금을 상환하지 못하게 되었고, 여기에 투자했던 미국의 서브프라임 모기지 회사를 비롯한 헤지펀드, 은행, 보험사 등이 전세계적으로 연쇄 부도가 났으며 이 금융위기는 북부 유럽의 부자국가였던 아이슬란드의 몰락과 이후 그리스, 이탈리아, 스페인, 포르투갈 등 지중해 인근 지역 국가의 국가부도 위기를 초래하는 주요 원인이 되었으며 이로 인해 유로화를 사용하는 EU 경제 전체의 부실을 초래하고 있다.

거품은 재정자금과 금융 완화 조치 등에 따른 유동성의 확대와 인플레이션이와 같은 금융 및 경제적 요인뿐만 아니라 경제 상황을 인식하는 투자자의 사회 심리적 요인 등이 복합적으로 결합되어 발생된다고 한다. 대부분의 거품은 거품이 붕괴되어 사회경제적으로 심각한 문제가 발생한 후에 인지되고 있으며 붕괴 이전에 거품인지 또는 정상적인 성장인지에 대한 납득할만한 이론적 근거를 제시하는 경우가 드물다. 이 논문에서는 거품형성에서 있어 투자자의 심리적 요인을 바탕으로 거품에 대한 통계적 정의를 유도하고자 한다.

2. 거품의 통계적 정의

금융시계열분석에서는 변동성군집(volatility clustering) 현상이 흔히 관측된다. 변동성군집 현상은 전체 시계열의 변동성에는 차이가 있으나 지협적으로 비슷한 변동성을 가지는 성질을 의미하며 이러한 성질을 가지는 시계열을 조건부이분산(conditional heteroscedastic) 시계열이라고 한다. 조건부이분산 시계열분석에서는 기본적으로 시계열 X_t 가 다음과 같은 모형을 따른다고 가정한다.

$$X_t = \delta + \sum_{i=1}^p \phi_i X_{t-i} - \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_t} a_t,$$

여기서 h_t 는 $t-1$ 시점까지의 시계열 정보 I_{t-1} 가 주어졌을 때 조건부 분산 $h_t = \text{Var}(\varepsilon_t | I_{t-1})$ 로 정의되고 a_t 는 분산이 1인 백색잡음을 나타낸다. h_t 를 어떻게 정의하는가에 따라 SV(stochastic volatility) 모형과 GARCH-type 모형으로 나눌 수 있다. 선형모형을 랜덤효과모형과 고정효과모형으로 나누는 것처럼, SV 모형은 조건부분산식에 랜덤효과가 포함하고 있으며 GARCH-type 모형은 결정적 구조를 가진다. 여기서는 모수추정과 설명이 상대적으로 쉬운 GARCH-type 모형을 중심으로 설명하고자 한다.

주식이나 채권 등 금융상품의 내재가치는 경제 상황 등에 의해 수시로 변하고 측정이 어렵기 때문에 상품의 현재 가격이 실제 내재가치에 비해 높고 낮음을 판단하기 어렵다. 이러한 이유 때문에 가격에 거품이 끼었는지는 거품이 꺼지면서 문제가 발생해야 인지되는 경우가 많다. 거품이 경제나 금융시장에 주는 영향이나 충격이 매우 크기 때문에 거품에 대한 다양한 연구가 진행되어 왔는데 Camerer (1989), Diba와 Grossman (1988), Flood와 Hodrick (1990), Wu (1997), Ma와 Kanas (2004) 등은 거품에 대한 수량적 분석방법을 제시하였다. 이 논문에서는 경제거품 또는 금융거품에 대한 투자자나 자산운역자의 심리적 측면을 표시할 수 있는 통계적 측도를 제안하려고 한다.

Analysis of Food Poisoning via Zero Inflation Models

Hwan Sik Jung¹ · Byung Jip Kim² · Sinsup Cho³ · In-Kwon Yeo⁴

¹Department of Statistics, Seoul National University

²Department of Statistics, Seoul National University

³Department of Statistics, Seoul National University

⁴Department of Statistics, Sookmyung Women's University

(Received May 30, 2012; Revised August 14, 2012; Accepted September 18, 2012)

Abstract

Poisson regression and negative binomial regression are usually used to analyze counting data; however, these models are unsuitable for fit zero-inflated data that contain unexpected zero-valued observations. In this paper, we review the zero-inflated regression in which Bernoulli process and the counting process are hierarchically mixed. It is known that zero-inflated regression can efficiently model the over-dispersion problem. Vuong statistic is employed to compare performances of the zero-inflated models with other standard models.

Keywords: Negative binomial regression, Poisson regression, Vuong statistic.

1. 서론

임의의 기간 동안 발생한 사건의 건수가 설명변수에 영향을 받는 경우 그 관계식은 흔히 포아송회귀모형(Poisson regression)에 의해 설명된다. 하지만 평균과 분산이 같아야 하는 포아송분포의 특징은 실증 분석에서 발생하는 과산포(over-dispersion), 즉 표본분산이 평균에 비해 큰 현상을 설명하지 못하는 단점이 있다. 과산포가 심각한 경우 대체 모형으로 음의 이항회귀모형(negative binomial regression)이 사용되기도 한다. 사건 발생건수에 대한 실증자료 분석에서 과산포와 더불어 발생하는 문제는 모형에서 기대되는 0의 관측값 빈도보다 많은 0이 관측되는 것이다. 관측값에 0이 많으면 과산포 문제가 더욱 심화되고 모형의 적합도가 떨어진다. 이러한 문제를 해결하는 방안으로 Lambert (1992)는 영과잉모형(zero-inflated model)을 제안했다.

식품의약품안전청의 통계에 따르면 우리나라에서 발생하는 60% 이상의 식중독은 세균이나 바이러스 때문에 발생하고 있다고 한다. 주요 원인균으로는 노로바이러스, 병원성 대장균, 살모넬라, 장염비브리오균, 황색포도상구균이 있는데 이들 원인균은 대부분 기온이나 습도와 같은 기후에 영향을 받는 것으로 알려져 있다. 이 논문에서는 기후요인이 식중독 발생에 어떻게 영향을 주는지를 주요 원인균별로 분석하고자 한다. 분석에 사용된 자료는 2005년 1월부터 2010년 5월까지 원인균별로 신고된 전국의 주별 식중독 발생건수와 해당 주의 전국 60개 기상관측소에서 관측된 평균기온과 평균습도, 일사량이다. 원

⁴Corresponding author: Professor, Department of Statistics, Sookmyung Women's University, Chongpa-dong 2ga, Yongsan-gu, Seoul 140-742, Korea. E-mail: inkwon@sm.ac.kr

인균별로 나누어 보는 이유는 기온, 습도, 일사량에 대한 원인균 활성화 정도가 다르고 이에 따른 식중독 평균 발생건수에 차이가 있기 때문이다. 원인균별로 식중독 발생건수를 나누어 분석할 때의 문제는 발생건수가 0인 관측값이 많이 있다는 것이다. 이 논문에서는 관측값에 0이 많이 발생하는 경우 영과잉모형을 이용하여 문제를 해결해 보고 Vuong 통계량을 이용하여 표준 포아송회귀모형이나 음의 이항회귀모형에 근거한 분석결과와 비교하는 방안에 대해 알아보려고 한다.

2. 영과잉 회귀모형

임의의 사건이 t 시점에서 발생한 횟수를 Y_t 라고 하고 Y_t 의 기대값 $E(Y_t) = \mu_t$ 는 설명변수 $\mathbf{x}_t = (x_{1t}, \dots, x_{pt})^T$ 에 영향을 받는다고 하자. 포아송회귀모형이나 음의 이항회귀모형에서는 이 관계식을 연결함수 $g(\cdot)$ 를 이용하여 다음과 같이 설명한다.

$$g(\mu_t) = \beta_0 + \beta_1 x_{1t} + \dots + \beta_p x_{pt} = \boldsymbol{\beta}^T \mathbf{x}_t,$$

여기서 $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ 는 추정해야 할 모수이며 Y_t 들은 독립이고 포아송분포 또는 음의 이항분포를 따른다는 가정 하에서 최대가능도 추정법을 이용하여 추정한다.

어떤 자료에서는 포아송분포나 음의 이항분포에서 기대할 수 있는 0의 빈도보다 훨씬 많은 수의 0이 관측되어 기존 포아송회귀분석이나 음의 이항회귀분석으로는 적합도가 떨어지는 경우가 있다. 이러한 자료를 분석하기 위해 Lambert (1992)는 영과잉모형을 제안하였다. Agarwal 등 (2002)에 의하면, 기본 자료생성분포가 포아송분포라고 할 때, 영과잉모형에서는 t 시점에서의 자료 관측이 다음과 같은 자료생성과정으로 이루어진다는 가정에서 모형설계가 이루어진다.

- 단계 1 (베르누이 난수생성): 0일 확률을 ϕ_t 로 하고 0이면 자료로 사용하고 1이면 단계 2로 감.
- 단계 2 (포아송 난수생성): 평균이 μ_t 인 포아송 난수를 발생하여 자료로 사용함.

이러한 자료생성과정을 거쳐 얻어진 자료의 확률질량함수는 다음과 같이 쓸 수 있다.

$$P(Y_t = y) = f(y; \phi_t, \mu_t) = \begin{cases} \phi_t + (1 - \phi_t)f^*(y; \mu_t), & y = 0, \\ (1 - \phi_t)f^*(y; \mu_t), & y = 1, 2, \dots, \end{cases}$$

여기서 $f^*(y; \mu)$ 는 평균이 μ 인 포아송확률질량함수를 의미한다.

만약 베르누이 확률 ϕ_t 도 설명변수 $\mathbf{x}_t$ 에 영향을 받는다고 하면, 이진자료에 대한 일반화선형모형에서 가정했던 것처럼 임의의 연결함수 $h(\cdot)$ 에 대해 관계식을 다음과 같이 가정할 수 있다.

$$h(\phi_t) = \alpha_0 + \alpha_1 x_{1t} + \dots + \alpha_p x_{pt} = \boldsymbol{\alpha}^T \mathbf{x}_t,$$

여기서 $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_p)^T$ 또한 추정해야 할 모수이다. 일반적으로 이진자료에 대한 연결함수 $h(\cdot)$ 는 분포함수의 역수인 위수(quantile) 함수가 많이 사용되고 있는데 로지스틱분포인 경우 로짓, 표준정규분포인 경우 프로빗, 극한값 분포인 경우 역로그로그 연결함수가 된다.

위와 같은 가정을 만족할 때, 모수 $\boldsymbol{\alpha}$ 와 $\boldsymbol{\beta}$ 의 로그가능도함수는 다음과 같다.

$$\begin{aligned} l(\boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{y}) &= \sum_{t=1}^n I(y_t = 0) \log(\phi_t(\boldsymbol{\alpha}, \mathbf{x}_t) + (1 - \phi_t(\boldsymbol{\alpha}, \mathbf{x}_t))f^*(y_t; \mu_t(\boldsymbol{\beta}, \mathbf{x}_t))) \\ &\quad + \sum_{t=1}^n I(y_t > 0) \{\log(1 - \phi_t(\boldsymbol{\alpha}, \mathbf{x}_t)) + \log(f^*(y_t; \mu_t(\boldsymbol{\beta}, \mathbf{x}_t)))\}, \end{aligned}$$

식중독 발생 예측모형[†]

여인권¹

¹숙명여자대학교 통계학과

접수 2012년 9월 26일, 수정 2012년 10월 29일, 게재확정 2012년 11월 6일

요약

식중독 발생에 대한 기존 연구에서는 기온과 습도와 같은 기후변수가 주된 설명변수로 취급되어 왔다. 이 논문에서는 주별 식중독 발생건수와 기후변수 간에 관계를 고찰하고 식중독 발생건수를 예측하기 위한 모형으로 포아송 회귀모형과 자기회귀이동평균모형을 비교한다. 비교결과 우리나라 식중독 발생은 시차를 두고 기후 변수에 영향을 많이 받고 있으나 식중독 발생 예측은 이들 변수보다 이전 시점의 식중독 발생 건수에 더 많이 영향을 받는 것으로 나타났으며 포아송 회귀모형은 예측의 관점에서 문제가 있음을 보였다.

주요용어: 교차상관관계, 자기회귀이동평균모형, 포아송 회귀모형.

1. 머리말

기후변화에 따른 자연재해와 생태계 변화에 대한 경고가 현실화되면서 국제적으로 기후변화와 관련된 문제에 대한 고찰과 이를 위한 자료 축적이 활발하게 진행되어 오고 있다. 식중독과 같은 질병은 기후변수와 밀접한 관계가 있는 것으로 알려져 있으며 영국, 미국, 캐나다와 같은 선진국을 중심으로 기후변화에 따른 식중독 발생에 대한 연구가 활발하게 진행되고 있다. 이들 연구에서는 온도가 식중독 발생에 가장 중요한 기상학적 요인으로 나타났으며 일부 기상자료는 시차를 두고 식중독 발생에 영향을 미치는 것으로 나타났다.

기후와 식중독 발생에 대한 기존 연구에서는 분석모형으로 일반선형모형, 포아송 회귀분석, 일반화 가법모형, 계절형 자기회귀이동평균 모형 등을 이용하였다. Bentham과 Langford (1995)는 일반화 가법모형을 이용하여 영국 웨일즈 지역의 식중독 발생률과 기온간의 관계를 분석하였으며 Patrick 등 (2004)은 덴마크에서 캄필로박터에 의한 식중독과 기후와의 관련성 분석을 통해 3, 4주 전의 기온에 영향을 많이 받는 것으로 나타났다. 이와 유사한 연구들이 미국, 캐나다, 호주 등에서 수행되었는데 대표적인 연구로 Magny 등 (2008)과 Fleury 등 (2006)이 있다.

우리나라의 경우 기상자료를 이용한 기후변화 연구는 Kim 등 (2009)과 Jang과 Joo (2009)에 의해 실시되었으며 기후와 식중독 발생 관계는 2000년대 중반부터 국가기관을 중심으로 연구가 진행되고 있다. 2006년 보건복지부와 2008년 식품의약품안전청의 연구결과에 의하면 기온이 1℃ 상승하면 세균성 이질이 5~9% 증가하고 식중독 발생건수는 5.3%, 환자 수는 6.2% 정도 증가할 것으로 예측한 결과를 발표하였다. 2009년부터 기후변화에 따른 식중독 발생영향의 다양한 주제에 대한 종합적 연구가 진행되고 있는데 자세한 내용은 Jeong과 Oh (2009)을 참고하기 바란다. Choi 등 (2008)은 2004년부터 2006년까지의 수도권 자료를 이용하여 최고기온, 습도, 월 효과가 일별 식중독 발생에 어떻게 영향을 주는지를 로그선형모형으로 설명하였다. 우리나라에서 발생하는 식중독의 60% 이상은

[†] 이 논문은 2011년도 숙명여자대학교의 교내연구비에 의하여 수행되었음.

¹ (140-742) 서울특별시 용산구 청파로47길, 숙명여자대학교, 통계학과, 교수. E-mail: inkwon@sm.ac.kr

세균이나 바이러스 때문에 발생하고 있으며 주요 원인균은 노로바이러스, 병원성 대장균, 살모넬라, 장염비브리오균, 황색포도상구균이 있다. 이들 원인균은 기온, 습도, 일사량과 같은 기후변수에 영향을 받는데 Jung 등 (2012)에 따르면 원인균에 따라 기후에 대한 활성도가 다른 것으로 나타난다. Yeo (2011)는 지역별로 나누어 식중독 분석을 하기 위해 우리나라 기후에 대한 군집분석을 했으며 Park (2010)은 온도와 저장 기간에 따른 냉장조리 식품인 콩나물 무침에 오염 초기 군수에 대한 확률분포를 예측하였다.

식중독은 기후의 영향뿐만 아니라 식자재의 종류와 요리법 등의 문화적인 요인에 영향을 받을 수 있기 때문에 식중독 발생 모형을 만들 때 해외에서 개발된 모형을 그대로 사용하는 것을 지양해야 하며 우리나라에 맞는 모형개발이 필요하다. 이 논문에서는 2002년 1월부터 2010년 5월까지 신고된 전국의 주별 식중독 발생건수와 해당 주의 전국 60개 기상관측소에서 관측된 평균기온, 평균습도, 일조량 자료를 이용하여 예측력이 좋은 모형을 선택하고자 한다.

2. 자료의 특성

식중독 발생 예측은 국민보건관리 차원에서 신속하게 대처해야 하기 때문에 월간 분석보다는 주간 분석이, 주간 분석보다는 일간 분석이 더 중요할 수 있다. 일간 분석의 경우, 문제는 식중독 발생건수가 0인 경우가 상대적으로 많아 영과잉모형 (zero-inflated model)과 같은 모형을 사용해야 하며 기후변수가 어느 정도의 시차를 가지고 영향을 주는지를 알아보고자 할 때 상당 기간의 항이 모형에 포함시켜야 하며 그 변동성이 크기 때문에 안정적인 모형을 개발하기 어렵다. 이를 고려해 이 논문에서는 주간 분석을 하고자 한다.

식중독 발생에 대한 예측모형을 선택하기 위해 먼저 우리나라에서 발생한 주별 식중독 발생건수와 이에 영향을 줄 것으로 예상되는 평균기온, 평균습도, 일조량의 특성을 알아보고자 한다. Figure 2.1은 2002년 1월부터 2010년 5월까지 총 439주 동안의 전국에서 보고된 식중독 발생건수와 전국 주요 기상관측소 60곳에서 측정된 평균기온, 습도, 일조량의 평균에 대한 시계열그림이다. 그림에서 수직 점선은 매년 1월 첫 번째 주를 의미한다.

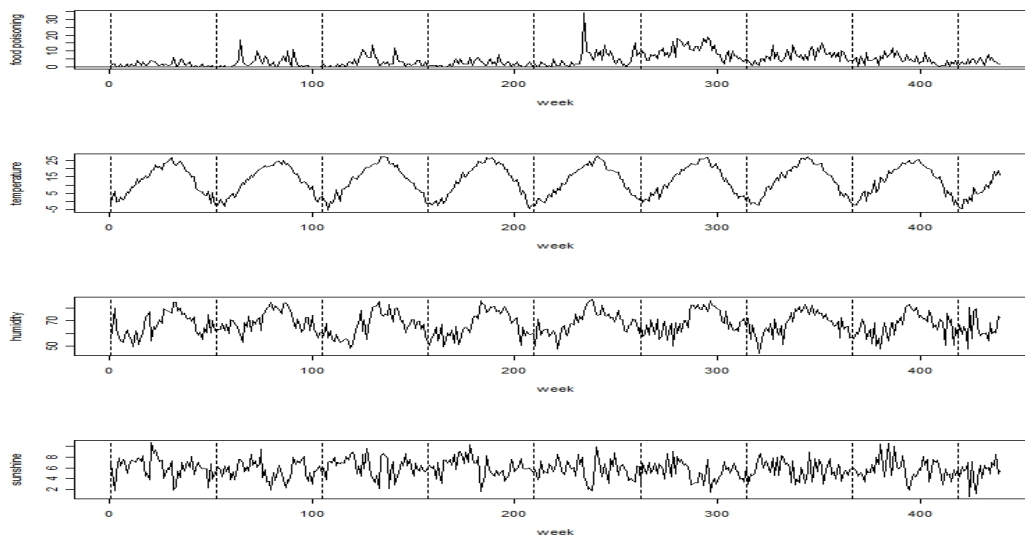


Figure 2.1 Time series plots of the number of food poisoning occurrences, average temperature, humidity and amount of sunshine

Statistical Analysis of Bivariate Current Status Data with Informative Censoring Using Frailty Effects

Yang-Jin Kim¹

¹Department of Statistics, Sookmyung Women's University

(Received November 2011; Revised January 2012; Accepted January 2012)

Abstract

In animal tumorigenicity data, tumor onsets occur at several sites and onset times cannot be exactly observed. Instead, the existence of tumors is examined only at death time or sacrifice time of the animal. Such an incomplete data structure makes it difficult to investigate the effect of treatment on tumor onset times; in addition, such dependence should be considered when censoring due to death is related with tumor onset. A bivariate frailty effect is incorporated to model bivariate tumor onsets and to connect death with tumor. For the inference of parameters, EM algorithm is applied and a real NTP(National Toxicology Program) dataset is analyzed as an illustrative example.

Keywords: Bivariate current status data, bivariate frailty effect, informative censoring, tumorigenicity.

1. Introduction

Current status data occur when subject is not under continuous examinations, instead, the occurrence of event is checked at only one examination time. This kind of data is also known as case I interval-censored data and was reviewed in detail by Sun (2006). In this paper, our interest is to consider bivariate current status data with informative censoring. For example, in a carcinogenicity study, tumors can occur at two different sites. However, instead of exact tumor onset times, the available data for the i^{th} subject are composed of $(C_i, D_i, \Delta_{i1}, \Delta_{i2})$, where C_i denotes a censoring time at the death or sacrifice time and D_i has a value one for a natural death and zero for sacrifice. Δ_{i1} and Δ_{i2} are indicator variables showing that tumor 1 and tumor 2 are found at C_i , respectively. Most studies assume that the censoring time is independent of the event occurrence time. For example, when the censoring time is predetermined, this assumption is valid. In the

This research was supported by Sookmyung Women's University Research Grants 2011.

¹Assistant Professor, Department of Statistics, Sookmyung Women's University, Seoul 122-807, Korea.

E-mail: yjin@sookmyung.ac.kr

tumorigenicity study, however, the censoring can happen at death and this requires investigation of the relation between death occurrence and tumor onset. As related studies, a three-state model with no tumor, tumor and death state has been applied to univariate current status data (Lindsey and Ryan, 1994; French and Ibrahim, 2002; Kim *et al.*, 2011). They constructed a likelihood for four possible outcomes depending on tumor onset and death. Dunson and Dinse (2002) considered informative censoring for multivariate current status data and developed a MCMC method under conditional independence using frailty effect. Chen *et al.* (2009) and Hens *et al.* (2009) applied a univariate frailty effect for multivariate current status data without death. In this paper, we extend the three state model to analyze bivariate current status data with informative censoring and assume the occurrence of two events are conditionally independent given a bivariate frailty.

We introduce notations and related likelihood in Section 2 and an estimation procedure is developed in detail in Section 3. The suggested method is applied to a real dataset in Section 4 and related comments are remarked in Section 5.

2. Data and Likelihood

We extend a three state model into a multi-state model in order to model bivariate current status data with informative censoring. For the $i (= 1, \dots, n)^{th}$ subject, let X_{ij} denote the $j (= 1, 2)^{th}$ unobservable tumor occurrence time and $\Delta_{ij} = I(X_{ij} \leq C_i)$ indicates whether the j^{th} tumor occurs before censoring time. Hereafter, lowercase letters are applied for observed data, $(c_i, d_i, \delta_{i1}, \delta_{i2}, z_i)$. There are two features to consider for modeling in our study. The first one is the association between tumor onsets. Even though their exact onset times are unobservable, their occurrences are considered in the context of multivariate survival times. Therefore, a bivariate frailty effect is adopted to model such association. That is, given a bivariate frailty $u_i = (u_{i1}, u_{i2})$, the conditional independence between two tumor occurrences is assumed. With covariate vector $z_i(r \times 1)$, the hazard functions of tumor 1 and tumor 2 are defined, respectively,

$$\alpha_{i1}(s|u_{i1}) = \alpha_0(s)\exp(z'_i\beta_1 + u_{i1}), \quad \alpha_{i2}(s|u_{i2}) = \alpha_0(s)\exp(z'_i\beta_2 + u_{i2}),$$

where $u_i = (u_{i1}, u_{i2})$ is assumed to follow a bivariate normal distribution with zero means and the following variance-covariance,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

The second feature of this data is the relation between tumor onset and death occurrence. In the presence of tumors, we assume that tumor 1 and tumor 2 have different lethality and the hazards of death is denoted by

$$\lambda_{i1}(t|u_{i1}) = \lambda_0(t)\exp(z'_i\gamma_1 + u_{i1}), \quad \lambda_{i2}(t|u_{i2}) = \lambda_0(t)\exp(z'_i\gamma_2 + u_{i2}).$$

Also, it is still possible for death to occur without tumors and corresponding hazards are given as

$$\lambda_{i3}(t) = \lambda_0(t)\exp(z'_i\gamma_3).$$

Statistical Analysis of K-League Data using Poisson Model

Yang-Jin Kim¹

¹Department of Statistics, Sookmyung Women's University

(Received June 25, 2012; Revised August 28, 2012; Accepted September 24, 2012)

Abstract

Several statistical models for bivariate poisson data are suggested and used to analyze 2011 K-league data. Our interest is composed of two purposes: The first purpose is to exploit potential attacking and defensive abilities of each team. Particular, a bivariate poisson model with diagonal inflation is incorporated for the estimation of draws. A joint model is applied to estimate an association between poisson distribution and probability of draw. The second one is to investigate causes on scoring time of goals and a regression technique of recurrent event data is applied. Some related future works are suggested.

Keywords: Bivariate poisson data, diagonal inflation model, K-league, random effect, recurrent event data.

1. 서론

이변량 포아송 자료는 쌍으로 표현되는 서로 연관된 계수형 자료로 여러 분야에서 예를 찾아볼 수 있다. 의학학 분야에서 치료 전후의 종양 개수, 역학 분야의 예는 각 농장에서 구제역으로 죽은 돼지와 소의 수, 경제학에서는 각 회사에서 발생한 자발적 이직수와 비자발적 이직수 등 한 개체로부터 측정된 두 사건의 발생 건수 또는 서로 연관된 두 관측 개체로부터 측정된 계수형 자료 등이 그 예가 될 것이다. 이변량 포아송 자료에서 공변량의 효과를 추정하기 위해 여러가지 방법들이 적용되어 왔다. Kocherlakota와 Kocherlakota (2001)은 EM 알고리즘을 Ho와 Singer (2001)은 일반화 최소 제곱법을 사용하였다. 이러한 이변량 포아송 분포의 확장으로 Li 등 (1999)은 다변량 포아송 자료에 대한 통계적 모형을 제시하였다.

본 연구에서는 이변량 포아송 모형을 K-리그 축구 경기 결과 분석에 적용하고자 한다. 특히 각 팀의 득점은 상대방 팀의 공격성을 자극하고 수비력을 강화함으로써 두 팀의 득점은 상관관계가 존재할 것으로 예상된다. 이러한 연관관계를 표현하기 위해 랜덤 효과를 포함한 이변량 포아송 분포를 적용한다. 이와 관련된 연구로 Karlis와 Ntzoufras (2003)는 포아송 확률 변수가 같이 공유하는 확률변수를 고려한 다음의 이변량 포아송 분포를 적용하였다.

$$\Pr(y_1, y_2) = \Pr(Y_1 = y_1, Y_2 = y_2) = \exp\{-(\lambda_1 + \lambda_2 + \lambda_3)\} \frac{\lambda_1^{y_1}}{y_1!} \frac{\lambda_2^{y_2}}{y_2!} \sum_{k=0}^{\min(y_1, y_2)} \binom{y_1}{k} \binom{y_2}{k} k! \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right)^k,$$

This work was supported by the National Research Fund(2011-0013221)

¹Assistant Professor, Department of Statistics, Sookmyung Women's University, 52 Hyochangwon-gil Yongsan-gu, Seoul 140-742, Korea. E-mail: yjin@sookmyung.ac.kr

Table 2.1. Frequency of goals by two teams

$Y_1 Y_2$	0	1	2	3	4	5	6
0	22	17	10	5			
1	25	25	16	11			
2	21	22	14	6	1		1
3	5	9	11	3	1		
4	1	3	1	2	1		
5		2	1				
6		1		1			
7	1	1					

여기서 두 팀의 득점은 확률 변수 Y_1 과 Y_2 로 표현되며 $E(Y_1) = \lambda_1 + \lambda_3$, $E(Y_2) = \lambda_2 + \lambda_3$, $\text{Cov}(Y_1, Y_2) = \lambda_3$ 로 λ_3 는 두 팀간의 연관 정도를 보여준다. 따라서 $\lambda_3 = 0$ 은 두 팀의 골점수가 독립임을 의미하며 이러한 경우 위 분포는 이중 포아송 분포(double poisson distribution)로 축소된다. 본 연구에서는 두 팀의 연관성을 모형화하기 위해 랜덤 효과를 고려한다. 본 논문의 2장에서는 K-리그 결과를 간략하게 설명하고 자료의 특징을 기술할 것이며 3장에서는 새로운 이변량 모형을 제안하고 그 추론 과정을 서술하고자 한다. 또한 4장에서는 득점 소요시간을 분석하기 위해 포아송 확률 과정에 근거한 재발 사건 자료 분석 기법을 적용할 것이다. 5장에서는 제안된 모형의 확장을 제시하고자 한다.

2. 2011 K-리그 득점 현황

K-리그는 2012년 현재 16개의 프로팀으로 구성되어 있으며 지난 2011년 리그에서는 전북 FC가 우승을 하였다. 전체 240번의 경기가 치루어졌으면 각 팀당 30번의 경기가 있었다. 경기는 홈경기와 원정경기로 두 팀이 번갈아 각자의 홈경기장에서 시합을 하게 된다. Table 2.1은 두 팀간의 골 점수의 빈도를 보여준다. 자료에 의하면 홈팀 승:무승부:원정팀 승의 비율은 107:65:68로 홈팀이 승리하는 빈도가 훨씬 더 높았다. 그리고 165시합(약 68.8%)이 골 득점차가 0 또는 1로($|Y_1 - Y_2| \leq 1$) 두 팀의 전력의 비슷함을 보였다. 본 연구에서는 각 경기당 홈팀과 원정팀간의 골점수의 연관 관계를 모형화하기 위해 랜덤 효과를 고려한다. 또한 무승부의 높은 발생 빈도를 위해, 동점 발생 확률을 포함한 결합 모형(joint model)이 적용될 것이다. 두 팀 간의 골 득점에 대한 상관 계수는 0.142 ($p\text{-value} = 0.026$)로 유의한 양의 상관관계가 존재함을 보여준다. 즉 상대팀의 득점은 서로에게 자극을 주어 득점을 이루어지게 하였음이 예상된다. 이에 대한 분석을 위해 재발 사건 방법론이 적용된다. 즉, 각 팀의 골 득점 소요시간에 대한 분포와 공변량의 효과를 분석한다. 재득점의 평균 소요 시간은 20.153분이고 가장 빠른 재득점 시간은 1분이었다.

3. 이변량 포아송 자료

본 연구에서는 한 경기당 두 팀의 골 득점에 대해 이변량 포아송 변수를 가정한다. 특히 랜덤 효과(u_{ht_i}, u_{at_i})가 주어져 있을 때, 홈팀과 원정팀의 골점수, Y_{i1} 와 Y_{i2} 는 조건부적으로 서로 독립이며 그들의 평균 λ_{i1} 와 λ_{i2} 에 대해 다음의 모형이 각각 적용된다.

$$\log(\lambda_{i1}|x_{i1}, u_{ht_i}) = \beta'_1 x_{i1} + u_{ht_i}, \quad (3.1)$$

$$\log(\lambda_{i2}|x_{i2}, u_{at_i}) = \beta'_2 x_{i2} + u_{at_i}, \quad (3.2)$$

Interactions of single nucleotide polymorphisms with dietary calcium intake on the risk of metabolic syndrome¹⁻³

Kirang Kim, Yoon Jung Yang, Kyunga Kim, and Mi Kyung Kim

ABSTRACT

Background: Gene-nutrient interactions may be important in modulating susceptibility to metabolic disorders.

Objectives: The objectives of this study were to assess the association of dietary calcium intake with the risk of metabolic syndrome and to investigate the interaction effects between dietary calcium intake and candidate gene polymorphisms.

Design: Subjects were participants in the Korea Association Resource project, which was initiated in 2007 as a large-scale, genomewide association analysis. A total of 8031 subjects were included in the study. Associations were assessed by using multi-variable-adjusted logistic regression analyses.

Results: High calcium intake appeared to be associated with a low risk of metabolic syndrome after covariates in both men (P -trend = 0.03) and women (P -trend = 0.0002) were controlled for. Among 27 single nucleotide polymorphisms (SNPs) selected as possible candidate gene polymorphisms affecting the risk of metabolic syndrome, 3 SNPs [rs6445834 in Rho guanine nucleotide exchange factor 3 (*ARHGEF3*), rs10850335 in T-box 5 (*TBX5*), rs180349 in *BUD13* homolog (*Saccaromyces cerevisiae*) (*BUD13*)] showed significant interaction effects with calcium intake tertiles or sufficiency in both men and women. Subjects with major allele homozygotes of these gene polymorphisms and high calcium intakes generally had a lower risk of metabolic syndrome than did those with minor allele homozygotes and low calcium intakes.

Conclusion: Dietary calcium intake appears to be inversely associated with the risk of metabolic syndrome and may modulate susceptibility to the syndrome in subjects who are minor allele carriers of rs6445834 in *ARHGEF3*, rs10850335 in *TBX5*, or rs180349 in *BUD13*. *Am J Clin Nutr* 2012;95:231–40.

INTRODUCTION

Metabolic syndrome, a clustering phenomenon of metabolic phenotypes such as abdominal obesity, dyslipidemia, hypertension, and insulin resistance, is an important precursor of cardiovascular disease and type 2 diabetes (1). Metabolic syndrome is a major public health challenge in both developed and developing countries, and the prevalence of diseases related to metabolic syndrome shows an increasing trend (2, 3). The prevalence of metabolic syndrome in Korea has also been steadily increasing over recent years, from 24.9% in 1998 to 31.3% in 2007. This prevalence is relatively high compared with those of other Asian countries (4, 5).

The progression of metabolic syndrome is influenced by genetic susceptibility and environmental factors, including diet,

and the interactions between them (6–9). Recently, published genomewide association studies have described candidate genes that contribute to metabolic syndrome and its phenotypes around the world (10–19). Some of these studies have also tried to investigate the interaction effects between the identified gene polymorphisms and dietary intake. Among dietary factors, interactions between fat intake and genes on the risk of metabolic syndrome have been studied and been found to be potential modulators of genetic susceptibility to metabolic syndrome (7, 20–22).

A growing body of research shows an inverse association between dairy consumption and the prevalence of metabolic syndrome (23). The effects of dairy consumption on the risk of metabolic syndrome can be partially explained by calcium intake in dairy foods (24), and several studies have found that people with higher calcium intakes have a lower prevalence of metabolic syndrome and its components (25–27). Dietary calcium is known to play a key role in blood pressure control and adipocyte metabolism, and thus its effects on metabolic syndrome could be partially mediated by its effects on body fat, blood pressure, and insulin sensitivity (28).

Given the substantial influence of gene-diet interactions on metabolic disorders, calcium intake could modify genetic susceptibility to metabolic syndrome similar to the way in which fat intake has an effect on the risk of metabolic syndrome conferred by gene polymorphisms (11–14). To our knowledge, there is no study investigating an interaction effect between calcium intake and gene polymorphisms on the risk of metabolic syndrome. A finding of gene-calcium interactions could improve our understanding of the mechanisms of metabolic alterations and could eventually lead to effective, personalized prevention of metabolic syndrome. Thus, the objectives of this study were to investigate the association of dietary calcium intake with the risk of metabolic syndrome and its phenotypes and to study the interaction

¹ From the Department of Preventive Medicine, Hanyang University, College of Medicine, Seoul, Korea (Kirang K and MKK); the Department of Food and Nutrition, Dongduk Women's University, Seoul, Korea (YJY); and the Department of Statistics, Sookmyung Women's University, Seoul, Korea (Kyunga K).

² Supported by grants from Korea Centers for Disease Control and Prevention, Republic of Korea (4845-301, 4851-302, 4851-307).

³ Address correspondence to MK Kim, Department of Preventive Medicine, Hanyang University, College of Medicine, Seoul, 133-791, Korea. E-mail: kmkkim@hanyang.ac.kr.

Received June 30, 2011. Accepted for publication October 11, 2011.

First published online December 14, 2011; doi: 10.3945/ajcn.111.022749.

effects between dietary calcium intake and candidate gene polymorphisms.

SUBJECTS AND METHODS

Subjects

Our study subjects were participants in the KARE⁴ project, which was initiated in 2007 as a large-scale, genomewide association analysis. The KARE population was recruited from 2 community-based cohorts that had been established as part of the KoGES in 2001. The cohorts were taken from the Ansung and Ansan areas in the Gyeonggi Province of South Korea and were a source of extensive phenotypic data for >260 traits. Data were collected through epidemiologic surveys, physical examinations and laboratory tests. The initial cohort included 10,038 participants (5018 for Ansung and 5020 for Ansan) between 39 and 70 y of age. After performing genotype calling and quality-control processes, a total of 8842 individuals were included in the KARE study. Further description of the study population may be found in a previous publication (10). Among those 8842 subjects, we excluded subjects who did not complete the food-frequency questionnaire ($n = 674$) or who reported implausibly low or high energy intakes (<500 kcal/d or >4000 kcal/d) ($n = 137$). Thus, the total number of study subjects used in our final analyses was 8031 (3846 men and 4185 women). The institutional review board of Hanyang University Medical Center approved this study.

General characteristics, anthropometric measurements, and biochemical variables

Information on demographic characteristics, education (less than high school or high school or greater), smoking status (current smoker or non-current smoker), alcohol intake (g/d), exercise (metabolic equivalent task hours/d), and family history of disease (hypertension, type 2 diabetes, or myocardial infarction) was collected by using an interview-administered questionnaire. Height was measured with a stadiometer to the nearest 0.1 cm, and weight was measured with a metric scale to the nearest 0.01 kg while the subjects were wearing light clothing with no shoes. BMI was calculated as weight (kg)/height (m²). WC was measured half-way between the lowest rib margin and the iliac crest. BP was measured by using a standard protocol after each subject had been sitting for ≥ 5 min. Systolic BP and diastolic BP measurements were recorded at least twice at 30-s intervals, and then an average value was used. Blood samples were collected after ≥ 8 h of fasting, and plasma concentrations of total cholesterol, HDL cholesterol, triacylglycerol, and glucose were quantified by using biochemical assays performed by a central laboratory (Seoul Clinical Laboratories, Seoul, Republic of Korea). For subjects with triacylglycerol concentrations <400 mg/dL, LDL cholesterol was calculated as described by Friedewald et al (29).

Metabolic syndrome

Metabolic syndrome was defined by using modified criteria proposed by the Third Report of the National Cholesterol

Education Program Adult Treatment Panel. A metabolic syndrome diagnosis was made when a subject fulfilled 3 of the following 5 criteria: WC ≥ 90 cm in men and ≥ 85 cm in women (30), triacylglycerol ≥ 150 mg/dL or treatment of dyslipidemia, HDL cholesterol <40 mg/dL in men and <50 mg/dL in women or treatment of dyslipidemia, systolic and diastolic BP ≥ 130 and 85 mm Hg or antihypertensive treatment, and fasting blood glucose ≥ 100 mg/dL or treatment of type 2 diabetes (31).

Dietary measurements

Well-trained interviewers collected dietary data by using a semiquantitative food-frequency questionnaire that asked each participant to provide his or her usual intake of 103 food items over a period of 12 mo. A semiquantitative food-frequency questionnaire was developed and validated by KoGES (Korean Genome Epidemiologic Information Management System) (32, 33). It consisted of 9 frequency categories ranging from "never" to ≥ 3 servings/d and 3 portion-size categories ["less" (0.5 serving), a standard amount (1 serving), or "more" (1.5 servings)] for each food item. All frequencies were standardized into "times per day" by using a conversion factor of 30.4 d/mo or 4.3 wk/mo. Food and nutrient intakes per day were calculated by using a weighted frequency per day, a portion size per unit, and a recipe and nutrient database that was provided by KoGES (32).

Genotyping and imputation

Genomic DNA samples were isolated from peripheral blood drawn from the participants and were genotyped on the Affymetrix Genome-Wide Human SNP Array 5.0 (Affymetrix). Bayesian robust linear modeling with the Mahalanobis distance (BRLMM) Genotyping Algorithm (Affymetrix) was used for the genotype calling of 500,568 SNPs. The genotype calling and quality-control processes are described in more detail in a previous study (10). The missing genotypes were imputed by using PLINK software and the JPT/CHB reference panel in HAPMAP (<http://hapmap.ncbi.nlm.nih.gov>). From the sample and SNP quality controls, a total of 8842 individuals and 327,872 SNPs were ultimately included in the analyses. Note that there was no evidence for possible population stratification in these data.

Statistical analysis

Before examining the interaction effects of calcium intake and gene polymorphism on metabolic syndrome, we conducted a genomewide association analysis of 327,872 SNPs to select candidate SNPs to be included in the interaction analysis. For each SNP, association with metabolic syndrome was tested on the basis of logistic regression. We adjusted for age, sex, recruitment area, and BMI, and an additive model was assumed for the genetic mode. First, we screened out SNPs with a weak association to metabolic syndrome ($P > 0.001$). Then, we screened out SNPs with minor allele frequency (<0.05) and SNPs that are not mapped to an exon/intron or that were within the 5-kbp upstream/0.5-kbp downstream regions of known genes. Finally, among the SNPs that passed the above screening, we selected 27 SNPs that showed strong statistical evidence from our data (ie, $P < 0.0001$) and/or have been reported to have significant

⁴ Abbreviations used: BP, blood pressure; KARE, Korea Association Resource; kbp, kilobase pair; KoGES, Korean Genome Epidemiology Study; SNP, single nucleotide polymorphism; WC, waist circumference.

Detection of Differentially Expressed Gene Sets in a Partially Paired Microarray Data Set

Johan Lim, Jayoun Kim, Sang-cheol Kim, Donghyeon Yu, Kyunga Kim, and
Byung Soo Kim

Abstract

Partially paired data sets often occur in microarray experiments (Kim et al., 2005; Liu, Liang and Jang, 2006). Discussions of testing with partially paired data are found in the literature (Lin and Stivers 1974; Ekbohm, 1976; Bhoj, 1978). Bhoj (1978) initially proposed a test statistic that uses a convex combination of paired and unpaired t statistics. Kim et al. (2005) later proposed the t_3 statistic, which is a linear combination of paired and unpaired t statistics, and then used it to detect differentially expressed (DE) genes in colorectal cancer (CRC) cDNA microarray data. In this paper, we extend Kim et al.'s t_3 statistic to the Hotelling's T^2 type statistic T_p for detecting DE gene sets of size p . We employ Efron's empirical null principle to incorporate inter-gene correlation in the estimation of the false discovery rate. Then, the proposed T_p statistic is applied to Kim et al.'s CRC data to detect the DE gene sets of sizes $p=2$ and $p=3$. Our results show that for small p , particularly for $p=2$ and marginally for $p=3$, the proposed T_p statistic compliments the univariate procedure by detecting additional DE genes that were undetected in the univariate test procedure. We also conduct a simulation study to demonstrate that Efron's empirical null principle is robust to the departure from the normal assumption.

KEYWORDS: cDNA microarray, empirical null distribution, false discovery rate, Hotelling's T^2 statistic, multiple testing, partially paired data

Author Notes: Byung Soo Kim is the corresponding author. Johan Lim's research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by Ministry of Education, Science and Technology(No. 2010-0011448). Byung Soo Kim's work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2010-0022549).

1. INTRODUCTION

In all organisms biological functions and epigenetic interactions alter the expression patterns of multiple genes simultaneously, but these signals are not always strong enough to be detected across all genes involved in function. As Subramanian *et al.* (2005) noted, in some cases, an increase of 20% in the expressions of all genes encoding members of a metabolic pathway might be more important than a 20-fold increase in the expression of a single gene. Recently, there have been efforts to detect differentially expressed (DE) gene sets rather than to detect individual DE genes in microarray studies. These efforts are supported by gene set analysis (GSA) approaches. GSA utilizes existing biological knowledge, such as biological pathways or cytobands; constructs gene sets; evaluates their differential expressions according to the phenotype of interest. Ever since Subramanian *et al.* (2005) and Mootha *et al.* (2003) proposed Gene Set Enrichment Analysis, some 30 statistical methods have been proposed. Overviews of these statistical methods are provided by Ackerman and Strimmer (2009) and Tsai and Chen (2009).

It is well established that inter-gene correlations among gene expression levels in a microarray experiment are strong and long-ranged, which means a gene may have strong pair-wise correlations with a large number of (partner) genes. Ignoring these inter-gene correlations in the analysis of the microarray data may produce highly unstable outcomes in estimation or testing (Owen, 2005; Klebanov and Yakovlev, 2007; Efron 2007a).

The strong and long-ranged inter-gene correlations and the recent move to GSA encourages the use of a multivariate test statistic for detecting differentially expressed gene sets in microarray data sets. In this context, Tsai and Chen (2009) proposed using a multivariate analysis of variance (MANOVA) approach for the GSA in studies with two or more experimental conditions. The “small n , large p ” setting is common in microarray analysis and causes a singular sample covariance matrix in the MANOVA approach. Tsai and Chen (2009) proposed the use of a linear shrinkage covariance estimator in the MANOVA approach, following the work of Schäfer and Strimmer (2005) and Warton (2008), to avoid inverting the singular sample covariance matrix. The linear shrinkage covariance estimator is a convex combination of a standard sample covariance matrix and a suitably chosen target that is structured and low dimensional, e.g., an identity or a diagonal matrix. The shrinkage covariance matrix estimator was utilized by Schäfer and Strimmer (2005), Opgen-Rhein and Strimmer (2007) and Ahdesmäki and Strimmer (2010) in various aspects of microarray data analyses.

The statistical properties of shrinkage estimators have been extensively studied. However, previous studies have primarily focused on the estimation of the covariance matrix and not on procedures using the covariance matrix estimator. Warton (2008) and Tsai and Chen (2009) showed the superiority of the



Permutation test for incomplete paired data with application to cDNA microarray data

Donghyeon Yu^a, Johan Lim^a, Feng Liang^b, Kyunga Kim^c, Byung Soo Kim^d, Woncheol Jang^{e,*}

^a Department of Statistics, Seoul National University, Seoul, Republic of Korea

^b Department of Statistics, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, USA

^c Department of Statistics, Sookmyung Women's University, Seoul, Republic of Korea

^d Department of Applied Statistics, Yonsei University, Seoul, Republic of Korea

^e Department of Epidemiology and Biostatistics, University of Georgia, 30602 Athens, GA, USA

ARTICLE INFO

Article history:

Received 15 December 2010

Received in revised form 19 August 2011

Accepted 21 August 2011

Available online 7 September 2011

Keywords:

Colorectal cancer

Incomplete paired data

Microarray data

Permutation test

ABSTRACT

A paired data set is common in microarray experiments, where the data are often incompletely observed for some pairs due to various technical reasons. In microarray paired data sets, it is of main interest to detect differentially expressed genes, which are usually identified by testing the equality of means of expressions within a pair. While much attention has been paid to testing mean equality with incomplete paired data in previous literature, the existing methods commonly assume the normality of data or rely on the large sample theory. In this paper, we propose a new test based on permutations, which is free from the normality assumption and large sample theory. We consider permutation statistics with linear mixtures of paired and unpaired samples as test statistics, and propose a procedure to find the optimal mixture that minimizes the conditional variances of the test statistics, given the observations. Simulations are conducted for numerical power comparisons between the proposed permutation tests and other existing methods. We apply the proposed method to find differentially expressed genes for a colorectal cancer study.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Testing the equality of means in a paired data set is a problem which we often encounter in practice. In particular, when one wants to identify differentially expressed genes between normal and tumor tissues in microarray experiments, it is common to extract RNA samples from both normal and tumor tissues of the same individual. This indirect design is usually adopted for the experiment that uses a common reference RNA hybridized both to normal and tumor tissues. Ideally, this experiment produces n pairs of microarray data. However, for some individuals, either the normal tissue or the tumor tissue is not big enough for the experimenter to extract enough RNA for conducting the experiment. For this reason, one of the paired expression levels often fails to be assessed for these individuals.

A simple approach for the unpaired missing data is to remove those unpaired incomplete observations from the analysis, but this would lead to a power loss in testing because tests based on all the available data are supposed to have better performance. Another approach to deal with missing data is imputation, but this may work only when incomplete pairs

* Corresponding author.

E-mail addresses: bunguji@snu.ac.kr (D. Yu), johanlim@snu.ac.kr (J. Lim), liangf@uiuc.edu (F. Liang), kyunga@sookmyung.ac.kr (K. Kim), bskim@yonsei.ac.kr (B.S. Kim), jang@uga.edu (W. Jang).

occur for a small number of genes. We may have, however, a sizeable number of missing arrays in either the tumor or the normal tissues.

Much attention has been paid to tests for incomplete paired data in previous literature. Many of the proposed tests are based on estimates of the mean difference using all the available data (Lin and Stivers, 1974; Ekbohm, 1974; Bhoj, 1974). These existing methods heavily rely on the assumption that the data are generated from a bivariate normal, but this assumption might not be true for gene expression data.

In this study, we develop a statistical method for incomplete paired data, which does not require any parametric assumption and utilizes the information from both paired and unpaired data. We consider the class of convex combinations of two linear statistics T_p and T_{up} :

$$T_\lambda = \lambda T_{up} + (1 - \lambda) T_p, \quad 0 \leq \lambda \leq 1,$$

where T_p and T_{up} are statistics for paired and unpaired samples, respectively. The optimal combination is further chosen to minimize the conditional variance of permuted T_λ , given the observations.

The rest of this paper is organized as follows. In Section 2, we introduce a statistical model for paired gene expression data and review the existing methods for incomplete paired data. In Section 3, we propose a permutation method for testing the difference of means. We numerically investigate the performance of our method based on simulations in Section 4. In Section 5, we apply the proposed method to identify differentially expressed genes from the colorectal cancer microarray data which previously studied by Kim et al. (2005). Finally, we discuss possible extensions of our method in Section 6.

2. Existing methods

We represent incomplete paired data as a combination of paired and unpaired data sets as follows. Let (X_i, Y_i) ($i = 1, 2, \dots, n_3$) be paired observations, for the i th individuals under two different conditions. Denote X_i^U ($i = 1, 2, \dots, n_1$) and Y_j^U ($j = 1, 2, \dots, n_2$) as unpaired observations on X and Y , respectively. Then, the entire data set can be arranged as follows:

$$\begin{array}{lll} X_1^U, \dots, X_{n_1}^U &; & X_1, \dots, X_{n_3} \\ Y_1^U, \dots, Y_{n_2}^U &; & Y_1, \dots, Y_{n_3}. \end{array}$$

Suppose that $(X_i, Y_i)^T$ is a bivariate normal random vector with mean $\mu = (\mu_X, \mu_Y)^T$ and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{XX} & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_{YY} \end{pmatrix}.$$

Then the marginal distributions of unpaired data are normal with mean μ_X and variance σ_{XX} for X^U , and normal with mean μ_Y and variance σ_{YY} for Y^U , respectively. We define $\bar{X}$, $\bar{Y}$, S_{XX} , S_{YY} , S_{XY} , and $r = S_{XY}/\sqrt{S_{XX}S_{YY}}$ as the sample means, the sample variances, the sample covariance, and the sample correlation for the paired data, respectively. We further define $\bar{X}^U$, $\bar{Y}^U$, S_{XX}^U , and S_{YY}^U as the sample means and the sample variances for the unpaired data. The primary interest here is testing the mean difference $\delta = \mu_X - \mu_Y$.

When missing data occur only on one side (i.e., $n_1 = 0$ or $n_2 = 0$), one can obtain a closed form of the maximum likelihood estimate (MLE) of μ and Σ (Anderson, 1957). However, there is no closed form of the MLE for the incomplete data, when missing data occur on both sides (i.e., $n_1 \neq 0$ and $n_2 \neq 0$), though one can find the MLE of μ and Σ with an iterative procedure (Orchard and Woodbury, 1970). To get a closed form of an estimator, Lin and Stivers (1974) proposed a modified MLE with non-iterative procedures and provided several test statistics based on their estimator. Ekbohm (1974) and Bhoj (1974) also suggested similar, but simpler, test statistics. In particular, Bhoj (1974) proposed a test statistic that was a linear combination of the paired and the unpaired t -statistics:

$$T_{\text{Bhoj}} = \lambda \frac{\bar{X}^U - \bar{Y}^U}{S_p^U \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} + (1 - \lambda) \frac{\bar{X} - \bar{Y}}{S_p / \sqrt{n_3}},$$

where

$$(S_p^U)^2 = \frac{(n_1 - 1)S_{XX}^U + (n_2 - 1)S_{YY}^U}{n_1 + n_2 - 2} \quad \text{and} \quad S_p^2 = \frac{S_{XX} + S_{YY} - 2S_{XY}}{n_3}.$$

The test statistic is a linear combination of two independent t -distributed random variables. Thus, it can be adequately approximated by a t -distribution with f_b degrees of freedom, where f_b is calculated by equating the second and fourth moments of the test statistic and the t -distribution (Patil, 1965).

Recently, Kim et al. (2005) considered a nonlinear combination of paired and unpaired statistics, which is referred to as t_3 :

$$t_3 = \frac{n_3(\bar{X} - \bar{Y}) + n_H(\bar{X}^U - \bar{Y}^U)}{\sqrt{n_3 S_p^2 + n_H^2 \left(\frac{S_{XX}^U}{n_1} + \frac{S_{YY}^U}{n_2} \right)}},$$

Novel Loci for Adiponectin Levels and Their Influence on Type 2 Diabetes and Metabolic Traits: A Multi-Ethnic Meta-Analysis of 45,891 Individuals

Zari Dastani^{1,9}, Marie-France Hivert^{2,3,9}, Nicholas Timpson^{4,9}, John R. B. Perry^{5,6,9}, Xin Yuan^{7,9}, Robert A. Scott^{8,9}, Peter Henneman^{9,9}, Iris M. Heid^{10,9}, Jorge R. Kizer^{11,9}, Leo-Pekka Lyytikäinen^{12,9}, Christian Fuchsberger^{13,9}, Toshiko Tanaka¹⁴, Andrew P. Morris⁵, Kerrin Small^{15,16}, Aaron Isaacs^{17,18}, Marian Beekman¹⁹, Stefan Coassin²⁰, Kurt Lohman²¹, Lu Qi²², Stavroula Kanoni¹⁶, James S. Pankow²³, Hae-Won Uh²⁴, Ying Wu²⁵, Aurelian Bidulescu²⁶, Laura J. Rasmussen-Torvik²⁷, Celia M. T. Greenwood²⁸, Martin Ladouceur²⁹, Jonna Grimsby^{3,30}, Alisa K. Manning³¹, Ching-Ti Liu³¹, Jaspal Kooner³², Vincent E. Moser⁷, Peter Vollenweider³³, Karen A. Kapur³⁴, John Chambers³⁵, Nicholas J. Wareham⁸, Claudia Langenberg⁸, Rune Frants⁹, Ko Willems-vanDijk⁹, Ben A. Oostra^{18,36}, Sara M. Willems¹⁷, Claudia Lamina²⁰, Thomas W. Winkler¹⁰, Bruce M. Psaty^{37,38}, Russell P. Tracy³⁹, Jennifer Brody⁴⁰, Ida Chen⁴¹, Jorma Viikari⁴², Mika Kähönen⁴³, Peter P. Pramstaller^{44,45,46}, David M. Evans⁴, Beate St. Pourcain⁴⁷, Naveed Sattar⁴⁸, Andrew R. Wood⁶, Stefania Bandinelli⁴⁹, Olga D. Carlson⁵⁰, Josephine M. Egan⁵⁰, Stefan Böhringer²⁴, Diana van Heemst⁵¹, Lyudmyla Kedenko⁵², Kati Kristiansson⁵³, Marja-Liisa Nuotio⁵³, Britt-Marie Loo⁵⁴, Tamara Harris⁵⁵, Melissa Garcia⁵⁵, Alka Kanaya⁵⁶, Margot Haun²⁰, Norman Klopp⁵⁷, H.-Erich Wichmann^{57,58,59}, Panos Deloukas¹⁶, Efi Katsareli⁶⁰, David J. Couper⁶¹, Bruce B. Duncan^{62,63}, Margreet Kloppenburg⁶⁴, Linda S. Adair⁶⁵, Judith B. Borja⁶⁶, DIAGRAM+ Consortium[‡], MAGIC Consortium[‡], GLGC Investigators[‡], MuTHER Consortium, James G. Wilson⁶⁷, Solomon Musani⁶⁸, Xiuqing Guo⁶⁹, Toby Johnson^{34,70,71}, Robert Semple⁷², Tanya M. Teslovich¹³, Matthew A. Allison⁷³, Susan Redline⁷⁴, Sarah G. Buxbaum⁷⁵, Karen L. Mohlke²⁵, Ingrid Meulenbelt¹⁹, Christie M. Ballantyne⁷⁶, George V. Dedoussis⁶⁰, Frank B. Hu²², Yongmei Liu²¹, Bernhard Paulweber⁵², Timothy D. Spector¹⁵, P. Eline Slagboom⁷⁷, Luigi Ferrucci¹⁴, Antti Jula⁵⁴, Markus Perola⁵³, Olli Raitakari⁷⁸, Jose C. Florez^{30,79,80,81}, Veikko Salomaa⁸², Johan G. Eriksson^{83,84,85,86,87}, Timothy M. Frayling⁶, Andrew A. Hicks⁴⁴, Terho Lehtimäki¹², George Davey Smith⁴, David S. Siscovick⁸⁸, Florian Kronenberg²⁰, Cornelia van Duijn^{17,18}, Ruth J. F. Loos⁸, Dawn M. Waterworth⁷, James B. Meigs^{3,30}, Josee Dupuis^{31,89}, J. Brent Richards^{15,90}*

1 Department of Epidemiology, Biostatistics, and Occupational Health, Jewish General Hospital, Lady Davis Institute, McGill University, Montreal, Canada, **2** Department of Medicine, Université de Sherbrooke, Sherbrooke, Canada, **3** General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **4** MRC CAITE Centre and School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom, **5** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **6** Genetics of Complex Traits, Peninsula Medical School, University of Exeter, Exeter, United Kingdom, **7** Genetics, GlaxoSmithKline, King of Prussia, Pennsylvania, United States of America, **8** MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, United Kingdom, **9** Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, **10** Department of Epidemiology and Preventive Medicine, Regensburg University Medical Center, Regensburg, Germany, **11** Departments of Medicine and Public Health, Weill Cornell Medical College, New York, New York, United States of America, **12** Department of Clinical Chemistry, University of Tampere and Tampere University Hospital, Tampere, Finland, **13** Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, United States of America, **14** Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, United States of America, **15** Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom, **16** Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom, **17** Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands, **18** Centre for Medical Systems Biology, Leiden, The Netherlands, **19** Section of Molecular Epidemiology, Leiden University Medical Center and The Netherlands Genomics Initiative, The Netherlands Consortium for Healthy Aging, Leiden, The Netherlands, **20** Division of Genetic Epidemiology, Innsbruck Medical University, Innsbruck, Austria, **21** Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, **22** Harvard School of Public Health, Boston, Massachusetts, United States of America, **23** Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota, United States of America, **24** Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands, **25** Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, United States of America, **26** Cardiovascular Research Institute, Morehouse School of Medicine, Atlanta, Georgia, United States of America, **27** Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, United States of America, **28** Lady Davis Institute for Medical Research, Department of Oncology, McGill University, Montreal, Canada, **29** Department of Human Genetics McGill University, Montreal, Canada, **30** Department of Medicine, Harvard Medical School, Boston, Massachusetts, United States of America, **31** Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, United States of America, **32** Cardiology, Ealing Hospital National Health Service (NHS) Trust, London, United Kingdom, **33** Department of Internal Medicine, University of Lausanne, Lausanne, Switzerland, **34** Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland, **35** Epidemiology and Biostatistics, Imperial College London, London, United Kingdom, **36** Department of Clinical Genetics and Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands, **37** Cardiovascular Health Research Unit, Departments of Medicine and Epidemiology, University of

Washington, Seattle, Washington, United States of America, **38** Group Health Research Institute, Group Health Cooperative, Seattle, Washington, United States of America, **39** Departments of Pathology and Biochemistry, University of Vermont, Burlington, Vermont, United States of America, **40** Cardiovascular Health Research Unit, University of Washington, Seattle, Washington, United States of America, **41** Medical Genetics Research Institute, Cedars Sinai Medical Center, Los Angeles, California, United States of America, **42** Department of Medicine, University of Turku and Turku University Hospital, Turku, Finland, **43** Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland, **44** Center for Biomedicine, European Academy Bozen/Bolzano (EURAC) (Affiliated Institute of the University of Lübeck, Lübeck, Germany), Bolzano, Italy, **45** Department of Neurology, General Central Hospital, Bolzano, Italy, **46** Department of Neurology, University of Lübeck, Lübeck, Germany, **47** School of Social and community medicine, University of Bristol, Bristol, United Kingdom, **48** British Heart Foundation Glasgow Cardiovascular Research Centre, University of Glasgow, Glasgow, United Kingdom, **49** Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence, Italy, **50** Laboratory of Clinical Investigation, National Institute of Aging, Baltimore, Maryland, United States of America, **51** Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands, **52** First Department of Internal Medicine, St. Johann Spital, Paracelsus Private Medical University Salzburg, Salzburg, Austria, **53** Public Health Genomics Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, and Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland, **54** Population Studies Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, Turku, Finland, **55** Intramural Research Program, Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, **56** Division of General Internal Medicine, Women's Health Clinical Research Center, University of California San Francisco, San Francisco, California, United States of America, **57** Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany, **58** Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany, **59** Klinikum Großhadern, Munich, Germany, **60** Harokopio University, Athens, Greece, **61** Collaborative Studies Coordinating Center, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **62** School of Medicine, Federal University of Rio Grande do Sul, Porto Alegre, Brazil, **63** Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **64** Department of Rheumatology and Department of Clinical Epidemiology, Leiden, The Netherlands, **65** Department of Nutrition, University of North Carolina, Chapel Hill, North Carolina, United States of America, **66** Office of Population Studies Foundation, University of San Carlos, Cebu City, Philippines, **67** Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi, United States of America, **68** Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, United States of America, **69** Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, United States of America, **70** University Institute of Social and Preventative Medicine, Centre Hospitalier Universitaire Vaudois (CHUV) and University of Lausanne, Lausanne, Switzerland, **71** Swiss Institute of Bioinformatics, Lausanne, Switzerland, **72** Metabolic Research Laboratories, Institute of Metabolic Science, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom, **73** Department of Family and Preventive Medicine, University of California San Diego, La Jolla, California, United States of America, **74** Brigham and Women's Hospital, Boston, Massachusetts, United States of America, **75** Jackson Heart Study Coordinating Center, Jackson State University, Jackson, Mississippi, United States of America, **76** Baylor College of Medicine and Methodist DeBakey Heart and Vascular Center, Houston, Texas, United States of America, **77** Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands, **78** Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku and the Department of Clinical Physiology, Turku University Hospital, Turku, Finland, **79** Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, United States of America, **80** Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **81** Diabetes Research Center, Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **82** Chronic Disease Epidemiology and Prevention Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland, **83** Diabetes Prevention Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland, **84** Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland, **85** Folkhalsan Research Centre, Helsinki, Finland, **86** Vaasa Central Hospital, Vaasa, Finland, **87** Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland, **88** University of Washington, Seattle, Washington, United States of America, **89** National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, United States of America, **90** Departments of Medicine, Human Genetics, Epidemiology, and Biostatistics, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Canada

Abstract

Circulating levels of adiponectin, a hormone produced predominantly by adipocytes, are highly heritable and are inversely associated with type 2 diabetes mellitus (T2D) and other metabolic traits. We conducted a meta-analysis of genome-wide association studies in 39,883 individuals of European ancestry to identify genes associated with metabolic disease. We identified 8 novel loci associated with adiponectin levels and confirmed 2 previously reported loci ($P=4.5\times 10^{-8}$ – 1.2×10^{-43}). Using a novel method to combine data across ethnicities ($N=4,232$ African Americans, $N=1,776$ Asians, and $N=29,347$ Europeans), we identified two additional novel loci. Expression analyses of 436 human adipocyte samples revealed that mRNA levels of 18 genes at candidate regions were associated with adiponectin concentrations after accounting for multiple testing ($p<3\times 10^{-4}$). We next developed a multi-SNP genotypic risk score to test the association of adiponectin decreasing risk alleles on metabolic traits and diseases using consortia-level meta-analytic data. This risk score was associated with increased risk of T2D ($p=4.3\times 10^{-3}$, $n=22,044$), increased triglycerides ($p=2.6\times 10^{-14}$, $n=93,440$), increased waist-to-hip ratio ($p=1.8\times 10^{-5}$, $n=77,167$), increased glucose two hours post oral glucose tolerance testing ($p=4.4\times 10^{-3}$, $n=15,234$), increased fasting insulin ($p=0.015$, $n=48,238$), but with lower in HDL-cholesterol concentrations ($p=4.5\times 10^{-13}$, $n=96,748$) and decreased BMI ($p=1.4\times 10^{-4}$, $n=121,335$). These findings identify novel genetic determinants of adiponectin levels, which, taken together, influence risk of T2D and markers of insulin resistance.

Citation: Dastani Z, Hivert M-F, Timpson N, Perry JRB, Yuan X, et al. (2012) Novel Loci for Adiponectin Levels and Their Influence on Type 2 Diabetes and Metabolic Traits: A Multi-Ethnic Meta-Analysis of 45,891 Individuals. *PLoS Genet* 8(3): e1002607. doi:10.1371/journal.pgen.1002607

Editor: Peter M. Visscher, The University of Queensland, Australia

Received: September 30, 2011; **Accepted:** February 3, 2012; **Published:** March 29, 2012

Copyright: © 2012 Dastani et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Distribution-free tests of mean vectors and covariance matrices for multivariate paired data

Erning Li · Johan Lim ·
Kyunga Kim · Shin-Jae Lee

Received: 26 October 2009
© Springer-Verlag 2011

Abstract We study a permutation procedure to test the equality of mean vectors, homogeneity of covariance matrices, or simultaneous equality of both mean vectors and covariance matrices in multivariate paired data. We propose to use two test statistics for the equality of mean vectors and the homogeneity of covariance matrices, respectively, and combine them to test the simultaneous equality of both mean vectors and covariance matrices. Since the combined test has composite null hypothesis, we control its type I error probability and theoretically prove the asymptotic unbiasedness and consistency of the combined test. The new procedure requires no structural assumption on the covariances. No distributional assumption is imposed on the data, except that the permutation test for mean vector equality assumes symmetric joint distribution of the paired data. We illustrate the good performance of the proposed approach with comparison to competing methods via simulations. We apply the proposed method to testing the symmetry of tooth size in a dental study and to finding differentially expressed gene sets with dependent structures in a microarray study of prostate cancer.

E. Li (✉)
Department of Statistics, Texas A&M University,
College Station, TX 77843, USA
e-mail: erningli.work@gmail.com

J. Lim
Department of Statistics, Seoul National University,
Seoul 151-747, Korea

K. Kim
Department of Statistics, Sookmyung Women's University,
Seoul, Korea

S.-J. Lee
School of Dentistry, Seoul National University, Seoul, Korea

Published online: 30 March 2011

 Springer

Keywords Multivariate paired data · Permutation · Equality of mean vectors · Homogeneity of covariance matrices

1 Introduction

Multivariate paired data (paired multivariate observations from the same group of experimental units) are common in many studies, and it is often of interest to test qualitative properties of population (or treatment) mean vectors and/or covariance matrices. Obviously, multivariate responses within each specific experimental unit are likely to be dependent. The two samples are correlated due to the fact of having paired observations. Moreover, in some cases, data may demonstrate distributions other than multivariate normal. The main contribution of the presented paper is the development of distribution-free tests that do not require stringent assumptions, their supporting theory, and a permutation procedure for implementation that take these considerations into account when testing the equality of mean vectors, equality of covariance matrices, and simultaneous equality of both mean vectors and covariance matrices between multivariate paired samples.

This work is motivated by two studies. One is a study on human tooth size. The data set contains 179 male adults who had natural normal occlusion with age range of 17–24 through a community dental health survey between 1999 and 2002 in Seoul, Korea, and is part of a standard occlusion study that has been undergoing since 1997 (Kim et al. 2005; Lee et al. 2007). A human adult normally has 14 permanent teeth in either maxilla (upper jaw) or mandible (lower jaw) with 7 teeth (central incisor to second molar) on either left or right side. The tooth sizes in terms of mesiodistal diameter of teeth of the patients were measured using digital Vernier calipers with sharpened points. Investigating whether tooth size profile is the same for the left and right sides around central incisors in terms of both mean values and variation would be helpful to understand the developmental biology of teeth and gain important insight into normative data of human tooth size, diagnostic criteria for malocclusion, and dental treatment planning (Garn et al. 2002; Uysan et al. 2005). Notice that the tooth sizes on the left and right sides from the same patient are naturally paired. In statistical aspects, however, previous analysis mainly focused on the asymmetry in mean vectors between left and right sides, while little has been done to justify the asymmetry between the covariance parameters, letting alone the possible difference between the underlying distributions. Among the subjects in the data set, we found several teeth exhibited excessive sample skewness and/or kurtosis, implying some violations of the usual multivariate normality assumption. We thus study a test procedure to address all these issues.

Another example is to test the asymmetry in either mean vectors or covariance matrices in multivariate paired observations in a microarray study. In cancer microarray experiments, it is a common practice to collect the tumor tissue from the same individual from whom the normal tissue is taken. Such experiments produce multivariate paired observations. The microarray data analysis focuses on finding differentially expressed genes between the control and the treatment groups. In recent years, much efforts have been given to identifying sets of genes that are significantly differentially

Evaluating adherence to recommended diets among cancer patients

Hye In Kim · Kyunga Kim · Jung Eun Lee

Received: 1 July 2011 / Accepted: 1 November 2011
© Springer-Verlag 2011

Abstract

Purpose We examined dietary and health-related behaviors among Korean cancer patients in the fourth Korean National Health and Nutrition Examination Survey.

Methods To compare the dietary and health-related behaviors among cancer patients, patients who had recovered from cancer, and those who had never had cancer, we obtained odds ratio (OR)s and 95% confidence interval (CI)s for the adherence to American Institute for Cancer Research (AICR) recommendations or a Dietary Approaches to Stop Hypertension (DASH)-style diet using the multivariate polytomous logistic regression.

Results A total of 103 cancer patients, 139 patients who had recovered from cancer, and 7,963 participants who had never had cancer were included. Patients who had ever been diagnosed with cancer were more likely to follow AICR recommendations or a DASH-style diet compared to those who had never had cancer. Compared to bottom tertile of the AICR adherence score, ORs (95% CIs) in the top tertile were 3.19 (1.86–5.46) for patients who recovered from cancer and 3.34 (1.81–6.17) for cancer patients. For a DASH-style diet, we found an OR of 2.26 (95% CI 1.28–3.99) for patients who recovered from cancer and an OR of 1.60 (95% CI 0.89–2.89) for cancer patients, comparing top to bottom tertiles. We also found that the degree of adherence among cancer patients was stronger among ever smokers than never smokers.

Conclusions Korean patients who had ever been diagnosed with cancer showed higher adherence to diets designed to prevent cancer or high blood pressure than those who had never had cancer. Our observations warrant further prospective studies to evaluate the association of adherence to a healthy diet with survival and quality of life among Asian cancer patients.

Keywords AICR recommendation · DASH-style diet · Cancer patients

Introduction

Cancer is the leading cause of death in Korea. Age-standardized incidence rates (per 100,000) were 281.2 for men and 160.3 for women in 1999 and 305.9 for men and 224.9 for women in 2007 [1, 2]. The annual percent increases in incidence of cancers prevalent in Korea were 15.0% for prostate cancer, 6.6% for colorectal cancer, and 6.6% for breast cancer from 1997 to 2007 [2]. Following improvement of treatment strategies and the implementation of a nationwide early detection program, the 5-year survival rate has been steadily increasing, reaching 48.3% for men and 67.1% for women in Korea in 2007 [2]. Given that the common cancers that are increasing in Korea are related to diet and to sedentary lifestyles and that survival rates are as high as in Western countries, it is important to examine diet among cancer survivors and the effect of diet on their survival and quality of life. Several epidemiologic studies have found high levels of transition to healthy lifestyles, including regular exercise, smoking cessation, and increased healthy food intake among cancer patients, after cancer diagnosis [3–6], but investigation in Asian populations is lacking.

H. I. Kim · J. E. Lee (✉)
Department of Food and Nutrition,
Sookmyung Women's University,
52 Hyochangwon-gil,
Yongsan-gu, Seoul 140-742, Republic of Korea
e-mail: junglee@sm.ac.kr

K. Kim
Department of Statistics, Sookmyung Women's University,
Seoul, Republic of Korea

The American Institute for Cancer Research (AICR) released its second set of recommendations about diet and other factors in 2007, with the goal of decreasing cancer incidence and mortality in the USA [7]. The Dietary Approaches to Stop Hypertension (DASH) Trials emphasized a diet that is high in vegetables, fruits, and low-fat dairy products to decrease systolic and diastolic blood pressure in normotensive and hypertensive adults [8, 9]. Participants who had a high AICR adherence score had a lower risk of cancer incidence and lower cancer mortality [10]. Adherence to DASH-style diet was associated with lower risks of hypertension [8], cancer [11], and diabetes [12].

Cancer patients during cancer treatment are advised to prevent or reverse nutrient deficiencies, and dietary guidelines for cancer prevention are suggested to reduce the risk of second cancers among cancer survivors; however, the scientific evidence is not sufficient to support the hypothesis that healthy diet and physical activity are linked to lower risk of cancer recurrence or survival [13]. Given that data on food choices among cancer patients in Asia are limited, we examined health and dietary behaviors by analyzing the degree of adherence to AICR-recommended or DASH-style diets among cancer patients.

Materials and methods

Study participants

The Korea National Health and Nutrition Examination Survey (KNHANES) is a nationwide study designed by the Ministry for Health and Welfare and the Korea Centers for Disease Control and Prevention to evaluate the health and nutrition status of the Korean population [14]. KNHANES was conducted in four phases (phases 1–4 in 1988, 2001, 2005, and 2007–2009, respectively), using a stratified, multistage probability sampling design to find a representative sample of the civilian, non-institutionalized Korean population. KNHANES was constructed in three parts: a health interview survey, a health behavior survey, and a nutrition survey. Trained interviewers using constructed questionnaires asked participants about sociodemographic characteristics, health status, nutritional status, and lifestyle. A medical team performed physical examinations at mobile health clinics, recording the following: anthropometric measurements, vital signs, blood and urine tests, a dental checkup, and a chest X-ray. A detailed description of the KHANES is reported elsewhere [14]. We analyzed data from phases 1 ($n=4,594$) and 2 ($n=9,744$) of KNHANES 4, which was performed between July 2007 and December 2008.

Cancer patients, patients who recovered from cancer, and those who had never had cancer were defined as follows: participants were requested to answer the following ques-

tions in relation to cancers of the stomach, liver, colorectum, breast, cervix, lung, and other organs: “Have you ever been diagnosed with cancer by a physician?” (yes/no); “Is the cancer cured?” (yes/no); “Do you currently suffer from cancer?” (yes/no); and “Are you currently being treated for cancer?” (yes/no). We divided the participants into three groups according to the questions: cancer patients, patients who recovered from cancer, and those who had never had cancer. “Cancer patients” were defined as those who have been diagnosed with cancer by a physician but not cured. “Patients who recovered from cancer” were defined as those who have been diagnosed cancer by a physician but cured and no longer had cancers in the present. The “never had cancer” group was defined as those who had never been diagnosed with cancer.

We excluded participants who did not have information on intake of food and nutrition ($n=94$), those who reported unreasonable total energy intake (total caloric intake less than 400 or greater than 5,000 kcal/day) ($n=1,367$), and those for whom information on body mass index (BMI) ($n=673$), smoking status ($n=58$), or physical activity ($n=15$) was not available. After these exclusions, 103 cancer patients (among these patients, 17 had been diagnosed with two or more cancers), 139 patients who recovered from cancer, and 7,963 participants who had never had cancer were included in the analysis.

Assessment of dietary and non-dietary factors

Food intake was assessed with a food frequency questionnaire containing 63 food items and with a 24-h recall [14]. Questionnaire responses were as follows: “3 times a day”, “2 times a day”, “once a day”, “4 to 6 times a week”, “2 or 3 times a week”, “once a week”, “2 or 3 times a month”, “once a month”, “6 to 11 times a year”, and “rarely”. All categories were converted to times per day using the median value. We used the recommendations relevant to modifiable factors among the AICR recommendations [15] as follows: (1) be as lean as possible without becoming underweight; (2) be physically active for at least 30 min everyday; (3) avoid sugary drinks and limit consumption of energy-dense foods; (4) eat more of a variety of vegetables, fruit, whole grains, and legumes such as beans; (5) limit consumption of red meats (such as beef, pork, and lamb) and avoid processed meats; (6) If consumed at all, limit alcoholic drinks to two for men and one for women a day; (7) limit consumption of salty foods and foods processed with salt (sodium); and (8) quit smoking.

We have classified food intake into the food groups named in the AICR recommendations and in the DASH-style diet: whole grain, red meat and processed meat, vegetable, fruit, legume, and sweetened beverage from food frequency questionnaires. Cabbage and radish were excluded

cuGWAM: Genome-wide Association Multifactor Dimensionality reduction using CUDA-enabled high-performance graphics processing unit

Min-Seok Kwon

Interdisciplinary Program in Bioinformatics,
Seoul National University,
Gwanak-gu, Seoul 151-742, Korea
E-mail: intellims@gmail.com

Kyunga Kim

Department of Statistics,
Sookmyung Women's University,
Yongsan-gu, Seoul 140-742, Korea
E-mail: kyunga.j.kim@gmail.com

Sungyoung Lee

Interdisciplinary program in Bioinformatics,
Seoul National University,
Gwanak-gu, Seoul 151-742, Korea
E-mail: biznok@gmail.com

Taesung Park*

Department of Statistics and Interdisciplinary program,
Seoul National University,
Gwanak-gu, Seoul 151-742, Korea
E-mail: tspark@stats.snu.ac.kr

*Corresponding author

Abstract: Multifactor dimensionality reduction (MDR) method has been widely applied to detect gene-gene interactions that are well recognized as playing an important role in understanding complex traits. However, because of an exhaustive analysis of MDR, the current MDR software has some limitations to be extended to the genome-wide association studies (GWAS) with a large number of genetic markers up to ~1 million. To overcome this computational problem, we developed CUDA (Compute Unified Device Architecture) based genome-wide association MDR (cuGWAM) software using efficient hardware accelerators. cuGWAM has better performance than CPU-based MDR methods and other GPU-based methods.

Keywords: MDR; GWAS; GP-GPU; gene-gene interaction; parallel computing.

Reference to this paper should be made as follows: Kwon, M-S., Kim, K., Lee, S. and Park, T. (2012) ‘cuGWAM: Genome-wide Association Multifactor Dimensionality reduction using CUDA-enabled high-performance graphics processing unit’, *Int. J. Data Mining and Bioinformatics*, Vol. 6, No. 5, pp.471–481.

Biographical notes: Min-Seok Kwon is currently a PhD Candidate in Interdisciplinary program in Bioinformatics at the Seoul National University. His research interests include statistical methods, data mining, database and high performance computing system in genomic and proteomic informatics.

Kyunga Kim received her PhD Degree in Statistics from Purdue University, West Lafayette, in 2007. Currently, she is an Assistant Professor in the Department of Statistics at Sookmyung Women’s University in South Korea. Her current research interests include statistical methodologies and data mining for high-dimensional data with applications to genetics, genomics and bioinformatics.

Sungyoung Lee is a MS Student in Interdisciplinary program in Bioinformatics at the Seoul National University. His research interests include data mining, database and high performance computing system in bioinformatics.

Taesung Park received his BS and MS in Statistics from Seoul National University, Korea, in 1984 and 1986, respectively, and PhD in Biostatistics from the University of Michigan in 1990. He served as chair of the Bioinformatics Program, Seoul National University from April 2005 to March 2008. He is a Professor at the Department of Statistics, Seoul National University, Korea. He is the Director for the National Research Laboratory of Bioinformatics and Biostatistics. His research areas include microarray data analysis, genome-wide association studies, statistical genetics, missing data analysis and longitudinal data analysis.

1 Introduction

One of the main objectives in Genome-Wide Association Studies (GWAS) is to identify and characterise susceptibility variants involved in common and complex disorders. With the rapid development of high-throughput genotyping technologies, GWAS have been facilitated to discover multiple genetic factors of common complex diseases. Since complex diseases are influenced by multiple genes with variant alleles, identifying multiple causal genes and their interactions is critical for the success of GWAS. The Multifactor Dimensionality Reduction (MDR) method addresses the high-dimensionality problem (e.g., a large number of genetic variables and relatively small sample sizes) via constructive induction, and was proposed to identify the gene-gene interactions that affect disease susceptibility (Ritchie et al., 2001; Hahn et al., 2003; Moore, 2003; Williams et al., 2004; Neuman et al., 2010).

The original MDR software was implemented by JAVA programming language and had graphical user interface (Hahn et al., 2003). Because this MDR JAVA version is designed to run on only single workstation, it is suitable for candidate gene studies with a small number (<500) of genetic variables, such as Single Nucleotide Polymorphisms (SNPs). In order to handle large dataset, Bush et al. proposed Parallel MDR (pMDR)

컴퓨터과학과

A State Preserving Approach to Recognizing Human Behavior using Wireless Infrared and Vibration Sensors

Seung Ho Cho

Division of Computer Eng.
Kangnam University
Yongin, Republic of Korea
shcho@kangnam.ac.kr

William D. Phillips^{*†}

^{*}Network Centric Systems,
Raytheon
Largo, FL, USA
bill.phillips@raytheon.com

Ravi Sankar[†]

[†]Dept of Electrical Eng.
Univ. of South Florida
Tampa, USA
sankar@usf.edu

Bonghee Moon

Dept of Computer Science
Sookmyung Womens' University
Seoul, Republic of Korea
moon@sookmyung.ac.kr

Abstract—Some home monitoring research has tried to deal with a broad range of human behaviors such Activity of Daily Living (ADL) or Instrumented ADL (IADL) in an entire home, but most research results does not yet seem satisfactory. So we focused on modeling human behaviors in a bedroom, not an entire home. To do this, we define a behavior state and behaviors. We constructed the experimental system by installing a wireless infrared sensor on the ceiling above vibration sensors located on a bed. Based on the behavior definition, we formed a feature vector for each behavior in a bedroom. The total 110 experiments on five behaviors were tested by 5 subjects. We achieved a recognition ratio of 75.4%. After analyzing causes of false negatives, we revised some features. Then, the total 115 experiments were tested by four additional subjects. The recognition ratio improved to 94.8%. The high recognition ratio indicates that the proposed behavior model is effective in recognizing human behaviors in a bedroom. This research will contribute to forming a foundation to track human behavioral log in a bedroom.

Keywords- home monitoring; human behavior recognition; feature vector; state preserving approach

I. INTRODUCTION

The growing aging population has resulted in an increasing interest in ubiquitous computing and wireless sensor technology. An estimated 9% of adults age 65+ and 50% of adults age 85+ need care in their daily life. It has become and continues to be burdensome on governments and families. In addition, many elderly individuals prefer to live at home longer, rather than move to an assisted living facility. They seek an alternative way to combine their need for care and their wish to live at home for as long as possible.

The technological advance in various wireless sensors, communication and artificial intelligence allows a kind of home monitoring that can provide an opportunity to solve this aging problem. However, a satisfactory solution requires that device installation should not cause the user any inconvenience nor pose a privacy problem. We offer a non-invasive and unobtrusive approach through the use of passive sensors. People at home or assisted living facility do not need to be aware of these sensors.

Many researchers have been trying to recognize human behaviors at home or assisted living facilities for the past decade [1]. Their interest has focused on recognizing a part or all of Activity of Daily Living (ADL) (e.g., dressing, eating, ambulating, transferring, hygiene) or IADL (Instrumented ADL) [2] with various sensors such as motion sensors, RFID, and accelerometers. However, individual behaviors tend to show a great deal of variation. There is still much difficulty in recognizing many kinds of behaviors, such as ADL or IADL, even though much research has been performed on this topic. Therefore, we take the research direction as follows:

- 1) From a practical point of view, an approach to recognizing behaviors in a confined space seems more effective.
- 2) Furthermore, we keep in mind that this research will be used for an elderly living independently or assisted living facilities.

According to the directions above, the examined movement space of an individual in this research is confined to a bedroom, not an entire home. So we focus on recognizing kinds of behavior that occur in a bedroom, such as sitting up/down in bed, lying up/down on the bed, and tossing and turning in bed.

The rest of the paper is organized as follows. In Section 2, we introduce related work. In Section 3, we define a behavior model. Section 4 explains the process of extracting a feature vector, and Section 5 explains a state preserving recognition approach. In Section 6, we validate the behavior. In Section 7, we summarize our results and future work.

II. RELATED WORK

Several studies have examined using motion sensors to monitor the activities of individuals at home [3]–[6]. These researches focused on finding out any abnormality in an elderly individual's activity pattern. The results of this research have provided a way to alert relatives, caregivers, or healthcare personnel to any abnormality in an activity pattern. Atypical days were observed in the histograms of the infrared (IR) motion sensor output counts for each hour and each room [3]. The behavioral patterns of older adults in assisted living

were analyzed by using statistical estimation based on the average amount of time spent in each room [7]. Circadian rhythmic analysis was also used as an automatic telemedicine system to analyze the daily activities of the elderly [8].

The circadian rhythmic approach has been studied in the daily activity analysis with ADL graphs [3]. Gaussian probability density function [8], and Support Vector Data Description (SVDD) method [9]. However, reference [9] reported 71 abnormal events. Out of the total abnormal events, there were 58 false positives where the system misclassified normal behavior as abnormal behavior, and 13 true negatives where the system correctly identified the subject's abnormal condition. This report implies that it is difficult to recognize ADL generally.

C. Franco et al. [10] tried to measure the dissimilarity between sequences of activity using a variant of the Hamming distance and interpreted their findings according to the circular Gumbel distribution. This research is not focused on recognizing activities, but capturing similarity between sequences of activity, instead.

Chan et al. [11] developed a system, also based on passive infrared (PIR) and magnetic switches that not only detected a subject's absence/presence in a particular room, but also measured his or her mobility in kilometers. An artificial neural network (ANN) monitored the subject's mobility data for deviations from his or her usual pattern. This system was based on the assumption that the monitored subject lived alone and had repetitive and identifiable habits. LeBellego et al. [12] tried to propose criteria to estimate the activity status of a patient from PIR presence detectors.

III. DEFINITION OF BEHAVIOR

A. Behavioral state

We define a behavioral log as a way to track individual behavior over time in daily life. It consists of a static behavioral state and a behavior as follows;

Behavioral log = {behavioral state, behavior},

where a behavioral state is defined as follows;

Behavioral state = {A, S, L}

where a state 'A' (Approaching) means that a person is standing near a bed, a state 'S' (Sitting) means that a person is sitting on a bed, and a state 'L' (Lying) means that a person is lying on a bed.

We define a person's five behaviors near or on a bed as in Table I [13].

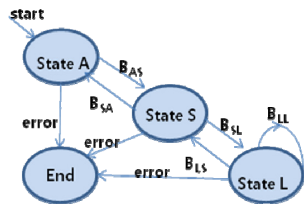


Figure 1. A state machine for human behaviors.

TABLE I. DEFINITIONS OF BEHAVIORS

Behaviors	State transition	Description
B _{AS}	A-->S	Behavior for a person standing near a bed to go and sit down on the bed
B _{SL}	S-->L	Behavior for a person sitting on a bed to lie down on the bed
B _{LL}	L-->L	Behavior for a person lying on a bed to toss and turn
B _{LS}	L-->S	Behavior for a person lying on a bed to get up and sit on the bed
B _{SA}	S-->A	Behavior for a person sitting on a bed to stand up and go out of the bed

The state machine in Fig. 1 represents the relationship between behavioral states and behaviors. For example, if a person stands nearby a bed, i.e. in state A and does B_{AS}, this means that he then sits down on the bed, and transits in the next state 'S'.

B. The Proposed System

Fig. 2 shows the configuration of the proposed system. The system consists of wireless sensor nodes, a base station, a monitoring application and a database on a PC.

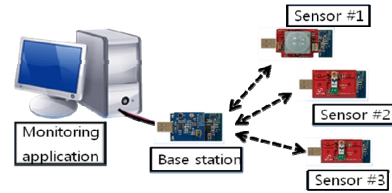


Figure 2. System configuration

We use three wireless sensor nodes in order to detect human behaviors on or near a bed. One sensor node consists of an infrared sensor and a wireless communication module. It is installed on the ceiling above the bed. This node measures movement of a human body around or on the bed. The other two sensor nodes consist of a MiniSense vibration sensor [14] and a wireless communication module. The MiniSense vibration sensor is a kind of cantilever-type piezo sensor which is loaded by a mass to offer high sensitivity at low frequencies. The cantilever-type piezo sensor detects vibration with no need for battery power.

One vibration sensor node is placed inside a pillow. By measuring pillow vibration caused by head movement, it detects behaviors performed by an individual who is lying down. The other vibration sensor node is installed beside the bed. The piezo sensor beside the bed measures bed vibration caused by human movement on or beside the bed and is used to detect whether a person sits up/down or is lying on the bed.

The sensor node, UBee430 [15] is composed of a sensor, TI MSP430 processor and Chipcon's CC2420 RF designed for low-powered wireless communication. The method of wireless communication is Zigbee. TinyOS 2.0.2 is ported on the



Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Tridex: A lightweight triple index for relational database-based Semantic Web data management

Seungseok Kang^{a,*}, Junho Shim^b, Sang-goo Lee^a^a Department of Computer Science and Engineering, Seoul National University, Seoul, Republic of Korea^b Department of Computer Science, Sookmyung Women's University, Seoul, Republic of Korea

ARTICLE INFO

Keywords:

Database

Triple

Semantic Web

Integrity constraint

Index

ABSTRACT

As more data are provided in Semantic Web, processing large amounts of data with triple-format triple-data, and interlinking the applications with triples, have become important for a variety of applications. A challenge is to index a triple database efficiently. Most existing triple indexing techniques suffer from data duplication and the oversized index tree problem. We analyzed the drawbacks of existing triple indexing methods from the viewpoint of the consistency and effectiveness of a triple database. We also considered the issues that need to be addressed to build a triple index for the management of relational database (RDB)-based triple data. Tridex is a lightweight B+-tree index structure designed to facilitate efficient processing of RDB-based triple databases. Tridex is beneficial in reduced index tree size and less data redundancy. We conducted an experimental evaluation to demonstrate the superiority of Tridex compared to existing triple index structures.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Triple (Berners-Lee, Hendler, & Lassila, 2001) is a well-known flexible data representation format in Semantic Web. It provides expressivity and interoperability for exchanging data between applications. In Semantic Web, information and relationships are represented by a collection of triples, each consisting of a subject, a predicate, and an object. World Wide Web Consortium (W3C) developed the triple data representation scheme as a standard format, such as RDF (<http://www.w3.org/RDF/>) and Manola & Miller, 2004). However, most data are structured and stored in non-triple data formats, such as relational databases. In the present study, we analyzed the essential features of a triple database from the viewpoint of a relational database. In addition, we investigated the fundamental aspects of a triple-based data model, which can maintain the consistent state of a triple database.

The increasing amount of triple data calls for efficient storage management of triple databases. Much research has focused on storing and querying triple data. However, existing triple storing and indexing techniques often do not consider the maintenance of triple data. Consider, for example, a bank maintenance application where each customer's information is stored in a triple database: customer account information at a bank is very sensitive, so it should be handled carefully during the execution of transition

queries, with no loss of information. If we consider triple as the representation scheme for bank information, the entire triple database should always be maintained in the consistent state. That is, for reliability, the database should maintain equivalent conditions, such as the integrity constraints of conventional databases.

In particular, the violation of integrity constraints is examined using the index structure of a set of attributes in a relational database. Similarly, there should be a tree index structure for a triple database to derive the equivalent state of the database. Several works on index systems for the triple data format have been proposed: MAP (Harth & Decker, 2005), HexTree (Weiss, Karras, & Bernstein, 2008), and TripleT (Fletcher & Beck, 2009). However, these approaches have inadequate structures for maintaining efficient triple index querying as well as the validation of database consistency.

In this paper, we argue that there are two main requirements for an efficient triple index from the viewpoint of the practical use of triple data: a data retrieval process (e.g., transactional query execution) and a data management process (e.g., validation of database consistency). We developed a novel technique for building a lightweight triple index structure to consider enforcement of the integrity constraint of a triple database. Our work can be applied to an RDB-based triple database, which requires reliability and trustworthiness of data: for example, bank account information. The remainder of this paper is organized as follows. Section 2 presents the background and related work on triple databases and the integrity checking process. In Sections 3 and 4, we present an analysis of the problems of conventional solutions using a triple

* Corresponding author. Tel.: +82 1036339935.

E-mail addresses: pyxis81@europa.snu.ac.kr (S. Kang), jshim@sookmyung.ac.kr (J. Shim), sglee@europa.snu.ac.kr (S.-g. Lee).

대수적 특성을 고려한 벡터 유사도 측정 함수의 고찰

Survey on Vector Similarity Measures : Focusing on Algebraic Characteristics

이동주(Dongjoo Lee)*, 심준호(Junho Shim)**

초 록

전자 상거래 시스템 환경에서 상품, 상품평, 사용자 특성 등은 주요한 정보 객체이다. 벡터는 객체의 표현기법으로 널리 사용되고 있다. 전자 상거래 데이터 객체들은 벡터로서 모델되어 각 특질에 해당하는 차원의 숫자 값으로 표현될 수 있다. 전자 상거래의 특성상 이러한 객체들은 방대한 분량이 되고 있고, 이중 여러 객체들은 실제로 같거나 유사한 객체일 수 있다. 따라서 객체간 유사도 측정은 전자상거래 시스템에서 중요한 역할을 한다. 본 논문에서는 벡터 객체에서 사용되는 대표적인 유사도 측정 함수들을 고찰한다. 유사 함수들은 각각의 대수적 특성을 가지고 있고 서로 연결된 특성을 보인다. 이러한 특성을 분석하고 또한 유사 함수들을 분류해 본다. 이러한 과정은 표준 벡터 유사도 함수가 가져야 할 대수적 특성을 제시해준다.

ABSTRACT

Objects such as products, product reviews, and user profiles are important in e-commerce domain. Vector is one of the most widely used object representation scheme. Information of e-commerce objects may be modeled by vectors in which the featured values are assigned to various dimensions. E-commerce objects are in general quantitatively large while some are similar or even same in reality. It plays, therefore, an important role to measure the similarity between objects. In this paper, we survey the state-of-the-art vector similarity measures. Similarity measures are analyzed to feature the algebraic characteristics and relationship of those, and upon which we classify the related measures accordingly. We then present such features that standard vector similarity measures should convey.

키워드 : 유사도, 벡터, 근접성, 거리
Similarity, Vector, Proximity, Distance

본 연구는 숙명여자대학교 2011년 교비연구로 수행되었음.

* DMC R&D Center, Samsung Electronics Co

** Corresponding Author, Division of Computer Science, Sookmyung Women's University
(E-mail : jshim@sookmyung.ac.kr)

2012년 10월 31일 접수, 2012년 11월 19일 심사완료 후 2012년 11월 20일 게재확정.



The dynamic competitive recommendation algorithm in social network services

Seok Jong Yu^{*}

Dept. of Computer Science, Sookmyung Women's University, Seoul 140-742, Republic of Korea

article info

Article history:

Received 9 March 2011
Received in revised form 15 September 2011
Accepted 31 October 2011
Available online 6 November 2011

Keywords:

Social network service
Recommender system
Recommendation algorithm
Twitter
PageRank

abstract

As the number of Twitter users exceeds 175 million and the scale of social network increases, it is facing with a challenge to how to help people find right people and information conveniently. For this purpose, current social network services are adopting personalized recommender systems. Existing recommendation algorithms largely depend on one of content-based algorithm, collaborative filtering, or influential ranking analysis. However, these algorithms tend to suffer from the performance fluctuation phenomenon in common whenever an active user changes, and it is due to the diversities of personal characteristics such as the local social graph size, the number of followers, or sparsity of profile content. To overcome this limitation and to provide consistent and stable recommendation in social networks, this study proposes the dynamic competitive recommendation algorithm based on the competition of multiple component algorithms. This study shows that it outperforms previous approaches through performance evaluation on actual Twitter dataset.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

For recent years, there have been significant progresses in social network services (SNS) qualitatively as well as quantitatively. As of November 2010, it was reported that more than 175 million people were using Twitter and nearly one billion tweets were being created a week. In addition, it is said that 240 million photos are being uploaded a day on Facebook and 12,000 photos were being browsed a second on Flickr. Furthermore, major e-commerce sites like Amazon and Netflix¹ are successfully exploiting SNS-based personalized recommendation services for marketing products [1]. Unlike Facebook and MySpace,² Twitter is a directed social network that allows people to follow others without prior reciprocal agreement. This unrestricted following mechanism makes the Twitter network grow fast and enables the large scale information propagation. For this reason, Twitter is currently evaluated as a new type of media other than TV or newspaper [8,16,28]. However, as the scale of Twitter increases enormously, it is facing with a significant problem with “finding right people or information to subscribe. When a new user signs up in Twitter, the “first thing to do is to search existing Twitter friends by email accounts. However, this function has a limitation that can “find only the people who are already known to the user.

Alternatively, recommender systems have been adopted for the purpose of supplementing the capability of traditional search systems. Personalized recommendation is particularly effective for application domains such as movie, music, book, and app stores that the number of items keeps increasing over time [9,16,20]. The performance of recommender system is

^{*} Tel.: +82 2 710 9831; fax: +82 2 710 9296.

E-mail address: yusjong@gmail.com

¹ www.amazon.com, www.netflix.com

² www.myspace.com

클러스터 비디오 서버에서 버퍼 공유를 최대화하는 동적 요청 분산 기법

A Dynamic Request Distribution Method for Maximizing
Buffer Sharing in Clustered Video Servers

조대현 *, 최승락 **, 이기용 1) ***

Dae Hyun Cho, Seunglak Choi, Ki Yong Lee

*(443-742) 경기도 수원시 영통구 매탄3동 416 삼성전자
dae.hyun.cho@samsung.com

** (463-867) 경기도 성남시 분당구 정자동 178-1 NHN
seunglak@gmail.com

*** (140-742) 서울시 용산구 청파로47길 숙명여자대학교 컴퓨터과학부
kiyonglee@sookmyung.ac.kr

요 약

본 논문은 각 서버가 모든 비디오를 저장할 수 있는 클러스터 비디오 서버 환경에서, 사용자의 요청들을 서버들에 분배하는 방법을 논의한다. 제안 방법인 DynaBS는 각 비디오에 대한 인기도가 동적으로 변하는 환경에서 클러스터 비디오 서버의 성능을 최대화하는 요청 분배 방법이다. DynaBS는 사용자의 요청들을 관찰하고 이를 기반으로 주기적으로 새로운 목표 서비스 상태(Target Service State, TSS)를 세운다. TSS는 클러스터 비디오 서버의 성능이 최대화되는 이상적인 요청 분배 상태로서, 버퍼 공유가 최대화되는 한편 공유 손실이 최소화되는 상태이다. 버퍼 공유는 어떤 요청을 처리하기 위해 버퍼에 저장한 데이터를 다른 요청의 처리에 재사용하는 기법으로서, 디스크 연산을 감소시켜 동시에 더 많은 요청을 처리할 수 있게 한다. 공유 손실은 클러스터 비디오 서버의 현재 서비스 상태가 TSS로 바뀌어가는 과정에서 일시적으로 버퍼 공유가 감소되는 현상을 말한다. 시뮬레이션에 기반한 다양한 실험을 통해 DynaBS가 다른 요청 분배 방법에 비해 우수한 성능을 가짐을 보인다.

Abstract

We consider a request distribution method that distributes incoming user requests among servers in a clustered video server, when each server in the cluster can store all video files in its local storage. The proposed method called DynaBS improves the performance of a clustered video server in environments where the popularity of videos dynamically changes. DynaBS monitors incoming user requests and periodically determines the Target Service State (TSS) based on them. The TSS

1) 교신저자

is an ideal service state of the clustered video server where its performance is maximized by maximizing buffer sharing while minimizing sharing loss. Buffer sharing shares previously buffered data for multiple requests, making it possible to support more number of concurrent requests by reducing costly disk operations. Sharing loss is a temporary degradation of buffer sharing incurred while the current service state is being changed to TSS. Through simulations driven by various dynamic workloads, we show that DynaBS outperforms the other methods.

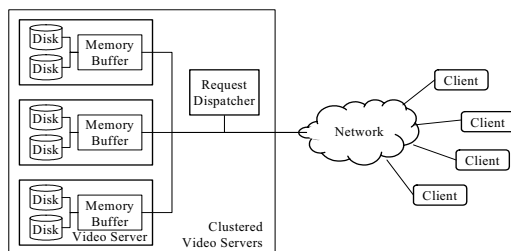
키워드: 요청 분배, 버퍼 공유, 공유 손실, 클러스터 비디오 서버, 동적 작업부하

Keyword: Request distribution, buffer sharing, sharing loss, clustered video server, dynamic workload

1. 서론

초창기의 비디오 서버는 고성능을 가진 단일 서버로 구성되었다. 하지만 사용자가 증가하면서 확장성을 갖추기 위해 여러 대의 서버로 구성된 클러스터 비디오 서버가 사용되고 있다 [1] [2]. 그림 1에서 볼 수 있듯이 클러스터 비디오 서버는 독립적인 저장 장치와 버퍼 메모리를 갖춘 다수의 서버 및 사용자의 요청을 받아 서버에게 분배하는 요청 분배기로 구성된다.

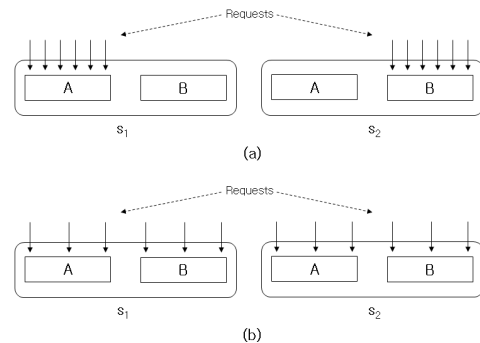
본 논문은 각 비디오의 인기도가 시간에 따라 동적으로 변하는 환경에서, 클러스터 비디오 서버의 성능이 최대화되도록 사용자의 요청을 서버들에 분배하는 방법을 논의한다. 예를 들어, 하루 사이에도 낮에 인기 있는 비디오(예: 만화나 어린이용 영화)는 밤에 인기도가 떨어지고, 밤에 인기 있는 비디오(예: 성인용 영화)는 낮에 인기도가 떨어진다.



(그림 1) 클러스터 비디오 서버 아키텍처

본 논문에서 제안하는 요청 분배 방법인 *DynaBS*

는 사용자의 요청을 분석하여 주기적으로 목표 서비스 상태(Target Service State, TSS)를 세운다. TSS란 어떤 비디오에 대한 요청을 어떤 서버로 보내야 클러스터 비디오 서버의 성능이 최대화될지를 나타내는 요청 분배 계획이다. 동적으로 변하는 비디오의 인기도를 반영하기 위해, DynaBS는 주기적으로 TSS를 갱신한다. TSS가 한 번 세워지면, 다음에 TSS가 갱신되기 전까지 사용자의 요청은 현재의 TSS에 따라 서버에게 분배된다.



(그림 2) 버퍼 공유의 예

DynaBS의 가장 큰 특징은 버퍼 공유(buffer sharing) [3]가 최대화되도록 요청을 분배한다는 것이다. 버퍼 공유는 버퍼에 올라간 비디오 데이터를 여러 요청의 처리에 재활용하는 기법

시맨틱 웹 데이터에서 키워드 검색을 위한 구조 질의 랭킹 기법

(A Method of Ranking Structured Queries for
Keyword Search on Semantic Web Data)

김 병 준 [†] 함 덕 민 ^{††} 송 인 철 ^{††} 이 기 용 ^{†††} 김 명 호 ^{††††}
(Byoung Jun Kim) (Deokmin Haam) (Inchul Song) (Ki Yong Lee) (Myoung Ho Kim)

요 약 시맨틱 웹에서 정보를 검색하기 위한 구조 질의(e.g., SPARQL)는 정보를 형식화 하여 다양한 의미의 질의를 정확하게 표현할 수 있지만, 질의 문법이 복잡하다. 그래서 구조 질의에 익숙치 않은 사용자들이 시맨틱 질의를 사용하는 것을 기대하기는 어렵다. 그러므로, 사용자가 손쉽게 시맨틱 질의 할 수 있는 환경을 제공하기 위해 사용자가 키워드 질의를 입력하면, 이를 시맨틱 질의로 변환해서 정보를 검색하는 연구가 활발히 진행 되고 있다. 키워드 질의는 사용하기 간편하지만, 하나의 키워드 질의가 다수의 구조 질의로 변환될 수 있다. 각 질의의 중요도 및 입력된 키워드 질의와의 연관성이 다르기 때문에 이를 측정하여 랭킹하는 방법이 필요하다. 본 연구에서는 시맨틱 웹 데이터에 존재하는 스키마 레벨에서의 트리플 간 연관성을 이용하여, 각 구조 질의를 랭킹하는 방법을 제안한다. 제안하는 방법은 구조 질의 로그에 기반하여 랭킹의 기준이 되는 연관성을 계산하고, 이를 적용하여 구조 질의 랭킹을 수행한다. 이 랭킹 방법은 사용자들의 선호를 반영하고 있기 때문에, 단순히 도메인 온톨로지의 정보만을 활용한 랭킹 방법보다 우수한 성능을 보인다.

키워드 : 시맨틱 웹, 온톨로지 구조 질의 랭킹, 온톨로지 트리플 패턴 연관도

Abstract To find information from Semantic Web, users need to use structured query languages like SPARQL. Using those languages, a user can express her information need exactly, but the usage is not familiar to ordinary users. To help the users find information easily, the methods of converting keyword queries issued by users into structured queries for Semantic Web are being researched. Keyword queries are easy to use, but it has ambiguity problem, in other words, a keyword query can be converted into a great number of structured queries. Each structured queries converted from a keyword query have different importance and relevance to the keyword query, thus, the ranking method for those structured queries is needed. In this paper, we propose a ranking method for structured queries converted from the same keyword query, which uses the affinities between schema-level triples in Semantic Web data. The proposed method extracts the affinities from the query log, and its ranking function for structured queries is based on the affinities. Because the proposed method reflects the preference of users, it can rank the structured queries more effectively than ranking methods using only the information of target domain ontology.

Key words : Semantic Web, Ontology Structured Query Ranking, Ontology Triple Pattern Affinity

· 이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2011-0020415)

[†] 정 회 원 : KAIST 전산학과
bjkim@dbserver.kaist.ac.kr

^{††} 학생회원 : KAIST 전산학과
dmhaam@dbserver.kaist.ac.kr
(Corresponding author임)
icsong@dbserver.kaist.ac.kr

^{†††} 정 회 원 : 숙명여자대학교 컴퓨터과학부 교수
kiyonglee@sookmyung.ac.kr

^{††††} 종신회원 : KAIST 전산학과 교수

mhkim@dbserver.kaist.ac.kr

논문접수 : 2011년 5월 12일

심사완료 : 2011년 11월 16일

Copyright©2012 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 데이터베이스 제39권 제2호(2012.4)

1. 서론

최근 시맨틱 웹(Semantic Web)이 주목 받고 있다. 시맨틱 웹이란 팀 버너스 리(Tim Burners-Lee)에 의해 제창된 지능형 웹으로, 웹 상에 존재하는 정보에 특정 코드를 추가하여 기계가 인간과 마찬가지로 정보를 인식하는 것을 목표로 한다. 시맨틱 웹에 존재하는 정보의 구조적/상호적 관계는 온톨로지(ontology)를 사용하여 정형화되어 기술되는데, 온톨로지에 존재하는 정보를 기술하기 위한 대표적인 언어로 RDF(Resource Description Framework)[1,2]가 있다. RDF에서는 각 자원을 구분하기 위해 URI(Uniform Resource Identifier)를 사용하고, 하나의 정보를 세 개의 자원(주어(subject), 술어(predicate), 목적어(object))으로 이루어진 트리플(triple)로 표현한다.

RDF로 표현된 온톨로지에서 원하는 정보를 검색하기 위해서는 형식화(formal)된 구조 질의 언어인 SPARQL[3], RDQL[4] 등을 사용해야 한다. 형식화된 구조 질의 언어를 사용하여 원하는 정보가 갖는 그래프 패턴을 서술하면, 질의처리 엔진이 그에 해당하는 서브그래프(sub-graph)를 전체 RDF 그래프(대상 온톨로지를 RDF로 표현한 그래프)에서 검색해서 반환한다. 이들을 사용하기 위해서는 그래프 패턴을 서술하기 위한 질의 언어 문법과 질의 대상인 도메인 온톨로지에 대한 지식이 필요하다. 하지만, 많은 사용자들은 구조 질의에 익숙치 않기 때문에, 이런 사전지식을 갖추어 질의하는 것을 기대하기

힘들다[2,3]. 이에 대한 해결책으로 사용자가 입력하는 키워드 질의(단순히 나열된 키워드들의 집합)를 형식화된 구조 질의로 변환하는 연구들이 진행되고 있다[5-10].

키워드 질의를 형식화된 구조 질의로 변환할 때 생기는 가장 큰 문제는 하나의 키워드 질의가 여러 개의 구조 질의로 변환될 수 있다는 것이다. 키워드 질의가 단순히 나열된 키워드 셋으로 구성되어 있는 것이 이런 모호성(ambiguity)을 유발한다. 예를 들어, 그림 1의 DBLP [http://www.informatik.uni-trier.de/~ley/db/] 온톨로지에 키워드 질의 $Q_1 = \langle \text{'semantic search'}, \text{'toronto'}, \text{'sigmod'} \rangle$ 를 수행하는 경우에 Q_1 은 다음과 같은 의미로 해석될 수 있다: (1) 'toronto'에서 개최된 'sigmod' 학회에서 발표된 'semantic search' 관련 논문, (2) 'sigmod' 학회에서 발표된 'semantic search' 관련 논문 중 저자 이름이 'toronto'인 논문, (3) 저자의 학교이름에 'toronto'가 포함되고 'sigmod' 학회에서 발표된, 'semantic search' 관련 논문. 그림 2는 Q_1 의 의미 (1)과 (2)에 해당하는 질의결과의 형태가 서로 다를 수 있음을 보여준다. 그림에서 색이 있는 타원은 클래스(class)에 해당하는 자원을 의미하고, 색이 없는 타원은 인스턴스(instance), 그리고 사각형은 텍스트 값을 의미한다. Q_1 은 위의 예시에서 언급된 세 가지 의미 이외에도 수많은 의미로 해석될 수 있다. 하나의 키워드 질의가 갖는 모든 의미를 고려하면 엄청난 수의 구조 질의(각 의미당 하나의 구조 질의)가 사용자에게 제시되는데, 이를 아무 기준 없이 사용자에게 반환할 경우 사용자에게 아무 도움이 되지 않는다. 그러므로 변환 결과로써 얻

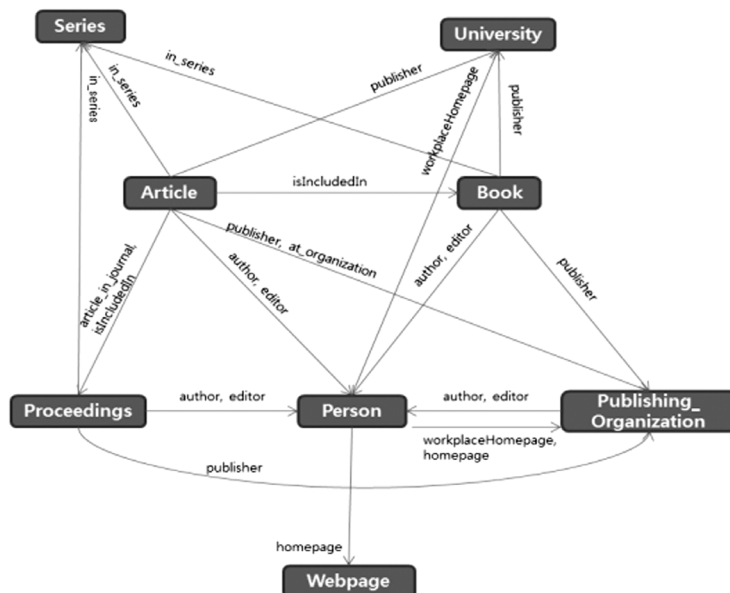


그림 1 DBLP 스키마

A Privacy Protection Method for Social Network Data against Content/Degree Attacks*

Min Kyoung SUNG[†], Ki Yong LEE^{††}, Jun-Bum SHIN^{†††}, *Nonmembers*, and Yon Dohn CHUNG^{†a)}, *Member*

SUMMARY Recently, social network services are rapidly growing and this trend is expected to continue in the future. Social network data can be published for various purposes such as statistical analysis and population studies. When social network data are published, however, the privacy of some people may be disclosed. The most straightforward manner to preserve privacy in social network data is to remove the identifiers of persons from the social network data. However, an adversary can infer the identity of a person in the social network by using his/her background knowledge, which consists of content information such as the age, sex, or address of the person and structural information such as the number of persons having a relationship with the person. In this paper, we propose a privacy protection method for social network data. The proposed method anonymizes social network data to prevent privacy attacks that use both content and structural information, while minimizing the information loss or distortion of the anonymized social network data. Through extensive experiments, we verify the effectiveness and applicability of the proposed method.

key words: privacy, social network, data publication, k -anonymity

1. Introduction

With the rapid increase of social network services, such as Facebook (facebook.com), Buzz (buzz.com), LinkedIn (linkedin.com), and Twitter (twitter.com), many researchers have paid much attention to social networks in recent years. These kinds of applications provide services based on the personal information of people and relationships among them. Researchers can analyze these social network data for the study of marketing, epidemiology, and so on. However, a social network may contain some sensitive information that discloses privacy of individuals [3], which causes social problems. For example, there are some personal information that is too sensitive to be disclosed, such as salary, disease, and credit rating. Moreover, a person can be the target of a crime that invades the person's privacy. A simple way to preserve the privacy of persons in social network data is to remove their identifiers (e.g., name or social security number (SSN)) from the social network data. However, even when the identifiers of persons are removed, there is still a risk of the privacy disclosure, as will be demonstrated in the following example.

Example 1. Let G be a social network (Fig. 1). Each node n_i ($i = 1, \dots, 6$) represents a person and each edge represents a relationship between persons. Each node contains content information, i.e., age, sex, and salary of a person. Note that the identifier of each person is not presented in G . Here, age and sex are non-sensitive information, whereas salary is sensitive one. Suppose an adversary knows that Tom is a 26 year old male and the number of persons having a relationship with him is 4. Then, the adversary infers that n_4 corresponds to Tom. Although the identifier of Tom is not presented in G , the adversary knows that the salary of Tom is \$ 40,000. $\square$

The above example shows that even when the identifiers of persons are removed from social network data, an adversary can disclose the privacy of a person by combining his/her background knowledge with the published social network data. Therefore, a data holder has to anonymize social network data before publishing them.

In general, a social network is represented as a graph, where each node represents a person and each edge represents a relationship between persons as described in the above example. The graph also contains content information, which corresponds to the values stored in each node. In order to achieve the anonymization of social network data, nodes and edges in the social network can be modified, added, deleted, or clustered. Several methods have been proposed to preserve the privacy of social networks in the past [6], [7], [9], [13], [15], [17], [18]. However, most of them considered only structural information [7], [9], [17], [18]. Some recent work [6], [13], [15] considered both structural information and content information, but have some limitations as described below.

Campan et al. [6] suggested a node clustering method that does not add or delete edges, but their method causes the loss of detailed structural information such as a relationship

Manuscript received March 15, 2011.

[†]The authors are with the Department of Computer Science and Engineering, College of Information and Communication, Korea University, Seoul, Korea.

^{††}The author is with the Department of Computer Science, Sookmyung Women's University, Seoul, Korea.

^{†††}The author is with DMC R&D, Samsung Electronics, Co., Ltd., Korea.

*This work was supported by Samsung Electronics.

a) E-mail: ydchung@korea.ac.kr (Corresponding author)

DOI: 10.1587/transinf.E95.D.152

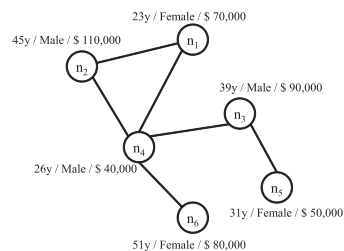


Fig. 1 An example of social network G .

between two specific persons. Zheleva et al. [15] proposed an edge clustering method which considers the content information attached to edges, but this method also causes the same problem as [5]. Wei et al. [13] suggested a method that produces unconnected social networks. However, unconnected social networks cause unnecessary confusion to a data receiver when he/she tries to reconstruct the original social network from the unconnected social networks. Furthermore, it considers only limited types of content information.

In this paper, we propose a novel privacy preservation method for social network data. For a given integer parameter k and social network G , the proposed method modifies G to an anonymized social network G_A by generalizing content information and adding extra edges to prevent privacy disclosure. As the structural information, we consider only the degree of each node (i.e., the number of edges connected to the node) and leave other types of structural information (e.g., sub-graphs) as future work because it complicates the problem more than necessary. In the anonymized social network G_A , every node has at least $(k - 1)$ indistinguishable nodes that have the same degree and content information. Thus, an adversary who knows the degree or content information of the target victim cannot correctly identify the target victim from G_A with more than $1/k$ probability. While preserving the privacy in the social network, our method minimizes the information loss caused by edge-addition and content-generalization as much as possible. The contributions of our paper are summarized as follows:

- We define information loss metrics that measure the loss of information incurred in social network anonymization.
- We propose a social network data anonymization method that can conserve the whole structural information as much as possible. That is, the proposed method can conserve more structural information than the previous methods [6], [13].
- Unlike the previous privacy preserving methods for the social network that consider only structural information [15], we consider both of the structural information and the content information of nodes in the social network, which contains a variety of person-specific information.
- Whereas the previous privacy preserving methods produce a number of unconnected social networks [13], our method produces a single connected social network.
- Through extensive experiments, we show that the proposed method performs better than the previous methods with respect to the level of privacy preservation and the loss of information.

The remainder of this paper is organized as follows. Section 2 introduces the previous privacy preserving techniques in relational databases and social networks. In Sect. 3, we formulate the data/attack models and information loss metrics we use in the paper. In Sect. 4, we describe our anonymization algorithm. We show our experimental results in Sect. 5, and conclude the paper in Sect. 6.

2. Related Work

When publishing data, a data holder such as hospitals has to anonymize the data to preserve the privacy of persons contained in the data. Generalization and suppression are popular techniques to anonymize the data. In those techniques, the original data are modified and/or changed to satisfy the required level of privacy protection. When anonymizing the data, we have to consider the following two factors: (1) the level of privacy preservation supported by the anonymization method and (2) the amount of information loss or distortion caused by the data anonymization.

Many researchers have worked for privacy preservation within the relational data [10], [11], [14], [16]. Sweeney et al. [11] pointed out that it is not sufficient for preserving the privacy in the published data to remove the identifiers of persons such as name or SSN, and proposed the notion of k -anonymity. The constraint of the k -anonymity indicates that every record in the published data must have at least $(k-1)$ other indistinguishable records. Machanavajjhala et al. [10] presented a model of ℓ -diversity which complements the shortcoming of k -anonymity, i.e., the lack of diversity in sensitive values. The ℓ -diversity requires that (1) the data satisfy the k -anonymity and also (2) the data have ℓ different sensitive values among k indistinguishable records. There were some considerations on dynamic data that is periodically updated [14], [16], where the data holder should ensure that the privacy is not disclosed by combining the previously released data and the current ones. In the data mining field, there was a study that attempts to maintain privacy while extracting useful information from databases [12].

However, although the above methods are suitable for relational data, they are not appropriate for social network data represented by graphs. For this reason, many researchers have studied methods for privacy preservation of the social network. The privacy leakage problems on social networks are classified as follows:

- **Sub-graph Attacks (Neighborhood Attacks):** Zou et al. [17] proposed an anonymization method to prevent structural attacks in the social network. They considered a sub-graph attack, called the d -neighborhood attack, where an adversary can identify a target victim in the social network by using his/her background knowledge, which is represented as a specific sub-graph. To preserve the privacy against such attacks, they modified the original social network so that the anonymized social network could contain at least k isomorphic sub-graphs.
- **Degree Attacks:** An adversary who knows the degree of a specific target victim in the social networks can acquire the sensitive data in the social network. The notion of k -degree anonymity [9] is proposed to prevent such attacks where every node in the social network should have at least $(k-1)$ indistinguishable other nodes by its degree. That is, an adversary who knows the degree of a target victim cannot iden-

멀티미디어과학과

Designing a Model for Context Awareness Based Reliable Auction-Recommendng System (CARARS) by Utilizing Advanced Information

Hwa-Jin Park and Sang-Beom Kim

Abstract This thesis aims to solve a problem that, even though the existing auction recommending system provides auction property based on the conditions and contexts users prefer, users cannot rely on the recommended property, but must analyze its investment value or request experts to analyze it. To solve this problem, advanced information in the real estate auction, an analysis of rights, an analysis of commercial quarters, and a development plan, are classified into 5 levels indicating investment value, which will be applied at the recommendation phase. This reliable recommending service is designed to be incorporated in the current context awareness system under a smart mobile environment. Therefore, we call it context awareness-based reliable auction recommending system (CARARS).

Keywords Context awareness system • Recommendation system • Real estate auction system

H.-J. Park

Department of Multimedia Science, Sookmyung.Women's University,
Hyochangwon-gil 52, Yongsan-gu, Seoul, South Korea
e-mail: phj2000@sm.ac.kr

S.-B. Kim (✉)

School of Real Estate Management, Sejong Cyber University, 111-1 Gunja-dong,
Gwangjin-gu, Seoul, South Korea
e-mail: dan@sjcu.ac.kr

K. J. Kim and K.-Y. Chung (eds.), *IT Convergence and Security 2012*,
Lecture Notes in Electrical Engineering 215, DOI: 10.1007/978-94-007-5860-5_14,
© Springer Science+Business Media Dordrecht 2013

111

1 Introduction

The auction information system has made a remarkable development as an essential element of a comprehensive information system for real estate, even since the early part of the 2000 s when private companies started to provide internet-based information up until the present, in which auction information in the Supreme Court is provided free of charge. In our current time, smart phones are diffused so prevalently that the information can be received at any time and in any place, and a context awareness system has been developed so thoroughly that real-time information that is constantly changing can be easily provided. Yet, most auction information systems for real properties offer personalized recommendation results on the basis of user profiles written at the time when customers joined, and most of the results are too multiple to be of any help. So, a user dissatisfied with the results has to select the region and kind of an auction property, every time he searches for the property. Although the results from content-based recommendations number fewer than 10, sometimes a user suffers a loss when he does not refer to advanced information comprising an analysis of rights and an analysis of commercial quarters. Subsequently, an ordinary person who is not an expert at auctioning has to refer to the advanced information of recommended properties in order to reduce such risk. Currently, however, advanced information is described in the form of a report and cannot be directly consulted.

Therefore, this research proposes turning the advanced information into indicators so that risks of investment may be reduced, that recommendations may be made according to the investment value of properties and that, at the same time, this system may be set up for recommending properties in compliance with current situation of a user. A recommendation system using the advanced information will be embedded in the context awareness auction information system suggested in [1] and will also be designed to acquire real-time contextual data such as a floating population when advanced information is analyzed.

2 Development of Advanced Information Analysis Indicators

2.1 Advanced Information Analysis

As for auction information, advanced information refers to well defined information regarding auction property derived from data which are collected and analyzed by an expert or a relating agency. The advanced information in real estate auction fields specifically means an analysis of rights to auction properties, an analysis of commercial quarters, and a city development plan. This advanced information is treated as a very valuable asset, provided by fee and mostly documented; so, it isn't reasonable to classify such an enormous amount of information into a few levels.

동적 경로 선정을 위한 효율적인 탐색 기법

최경미*, 박화진**, 박영호***

요 약

최근, ITS(Intelligent Transportation Systems)의 개발과 함께 차량용 내비게이션의 실시간 교통 정보를 이용하는 수요가 급증하면서, 경로탐색의 중요성이 더욱 가속화되고 있다. 그러나 기존의 경로탐색 알고리즘의 대부분은 최단경로 탐색을 위한 알고리즘으로, 정적인 거리 및 운행 시간정보를 사용하여 최적 경로를 계산하여 운전자에게 제공하기 때문에 교통량에 따라 동적으로 변하는 현 시점에서의 최적의 경로를 제공하지 못하는 문제가 있다. 따라서 본 논문에서는 이를 해결하기 위해 감속률과 거리에 기반한 동적 경로 선정을 위한 의미적 최단거리 알고리즘(Semantic Shortest Path algorithm with Reduction ratio & Distance, SSP_RD)과 감속률과 거리에 기반한 이동 경로 예측 모형화 및 동적 이동 경로 링크 맵을 제안한다.

An Efficient Search Mechanism for Dynamic Path Selection

Kyung-Mi Choi*, Hwajin Park**, Young-Ho Park***

Abstract

Recently, as the use of real time traffic information of a car navigation system increases rapidly with the development of Intelligent Transportation Systems (ITS), path search is getting more important. Previous algorithms, however, are mostly for the shortest distance searching and provide route information using static distance and time information. Thus they could not provide the most optimal route at the moment which changes dynamically according to traffic. Accordingly, in this study, Semantic Shortest Path algorithm with Reduction ratio & Distance(SSP_RD) is proposed to solve this problem. Additionally, a routing model based on velocity reduction ratio and distance and a dynamic route link map are proposed.

Keywords : Reduction Ratio, Distance, Dynamic Route, SSP_RD

1. 서론

최근 ITS(Intelligent Transportation Systems)의 개발과 함께 차량용 내비게이션의 사용이 급증하면서 경로탐색의 중요성이 더욱 가속화되고 있다. 경로탐색 알고리즘은 교통시스템, 통신 네트워크, 운송시스템, 이동 로봇의 경로 설정 등

다양한 분야에 사용되고 있다[1].

경로탐색과 관련된 기존의 경로탐색 알고리즘의 대부분은 최단거리 및 최소비용 탐색을 위한 알고리즘으로서, 경로 문제에 동적으로 변하는 상황정보를 주요하게 고려하지 않았다.

그러나 도로 교통량의 증가와 차량들의 이동이 증가하면서 교통상황이 이동 경로 비용에 미치는 효과가 매우 커지고 있으며, 따라서 교통상황을 고려한 이동 경로 비용을 최소화하는 요구가 대두되고 있다.

이러한 문제점을 해결하기 위해, 실제로 상황에 기반한 동적인 교통 정보의 제공은 운전자의 의사결정에 영향을 미친다는 연구결과가 있다.

미국 캘리포니아 주 샌프란시스코시에서 실시간 교통정보를 다양한 전달매체를 통해 시민들에게 전달했을 때 운전자들의 의사결정 형태를

※ 제일저자(First Author): 최경미

접수일:2012년 08월 29일, 수정일:2012년 09월 19일

완료일:2012년 09월 27일

* 숙명여자대학교 멀티미디어학과

kmchoi0990@sookmyung.ac.kr

** 숙명여자대학교 멀티미디어학과 교수

*** 숙명여자대학교 멀티미디어학과 교수(교신저자)

■ 본 연구는 숙명여자대학교의 2012학년도 교내연구비 지원에 의해 수행되었음

살펴보면, 교통정보를 받은 이용자 중 인터넷 사이트를 이용한 경우 81%, 전화를 이용한 경우 45%, 라디오나 TV를 이용하여 실시간 교통정보를 받은 이용자는 25%가 통행계획을 변경한 것으로 나타났다[2][3].

따라서 본 논문에서는 이러한 문제점을 해결하기 위해, 감속률과 거리에 기반한 동적 경로 선정을 위한 의미적 최단거리 알고리즘(Semantic Shortest Path algorithm with Reduction ratio & Distance, SSP_RD)을 제안한다.

동적 경로 이동 선정은 출발지 노드에서 목적지 노드까지의 도로혼잡 상황정보를 파악하고, 상황에 따른 이동시간을 추론하여 감속률을 구하고, 감속률과 거리에 기반해 경로를 선정한다. 이때, 감속률은 낮을수록 속도차감 정도가 줄어들어 좋다. 이러한 방법은 내비게이션 등의 경로 탐색 알고리즘에 적용할 수 있는 방법이다.

본 논문에서는 감속률과 거리에 기반한 이동 경로 예측 모형과 이동 경로 링크 맵을 제안한다. 본 논문의 공헌은 다음과 같다.

첫째, 감속률과 거리에 기반한 이동 경로 예측을 모형화하였다. 둘째, 감속률과 거리에 기반한 동적 이동 경로 링크 맵을 제안하였다. 셋째, 감속률과 거리에 기반한 동적 경로 선정을 위한 의미적 최단거리 알고리즘(Semantic Shortest Path algorithm with Reduction ratio & Distance, SSP_RD)을 제안하고, 구현하였다.

본 논문의 구성은 다음과 같다. 2장에서는 Dijkstra, DSPT, A*, ACO 알고리즘에 대해 설명한다. 3장 동적 경로 선정을 위한 탐색 기법에 대해 설명하고, 4장에서는 실험 결과에 대해서 설명한다. 마지막으로 5장에서는 결론 및 향후 연구로 논문을 마친다.

2. 관련연구

본 장에서는 경로 탐색 알고리즘인 Dijkstra, DSPT, A*, ACO 알고리즘에 대해 설명한다.

2.1 Dijkstra 알고리즘

Dijkstra 알고리즘[4]은 어떤 간선도 음수 값을 갖지 않는 방향 그래프에서 주어진 출발점과

목적점 사이의 최단 경로를 찾는 알고리즘이다. Dijkstra 알고리즘은 각 정점에 이르는 최단 경로를 출발점의 주변에서부터 하나씩 확장해가면서 서서히 범위를 넓힌 후 최종적으로 모든 정점에 이르는 최단 경로를 구하는 알고리즘이다.

하지만, Dijkstra 알고리즘은 Polynomial Time의 시간 복잡도를 가지는 최단거리 알고리즘으로 노드의 수가 증가함에 따라 경로 탐색 시간이 오래 걸린다는 단점이 있다.

Dijkstra 알고리즘의 수행과정을 살펴보면, 먼저 출발점에서 가장 인접한 정점을 찾는다. 그 정점까지의 거리가 최단거리가 되며, 최단거리의 정점들의 집합을 S라고 한다. 집합 S에 포함되지 않은 정점중에서 출발점으로부터 가장 가까운 정점을 찾고 새로운 정점은 집합 S에 이웃하는 정점들 중 하나가 된다. 그 정점까지 거리는 최단거리이며, 그 정점을 집합 S에 포함하게 된다. 모든 정점이 집합 S에 포함될 때까지 과정을 반복한다.

2.2 DSPT 알고리즘

최단 경로를 찾는 알고리즘으로 Dijkstra 알고리즘을 많이 사용한다. 하지만, Dijkstra 알고리즘은 일부 링크의 값이 수정될 때, 재사용하면 모든 노드에 대해서 반복 계산하여 계산량이 많아지고, 불필요한 수정이 발생하게 되는 문제가 있다[5].

이러한 Dijkstra 알고리즘의 단점을 보완하기 위한 동적 알고리즘인 DSPT(Dynamic Shortest Path Tree)가 있다. DSPT 알고리즘은 수정된 링크의 값에 영향을 주는 노드만을 계산하여 계산량을 줄이고, 불필요한 수정을 하지 않는다[6].

하지만, DSPT 알고리즘은 계산이 수행되는 자식노드들 사이에서의 증가 거리는 제외하고, 외부노드로부터 자식노드로 들어오는 증가거리만을 계산하기 때문에 일부 잘못된 계산을 하는 문제가 발생한다.

2.3 A* 알고리즘

A* 알고리즘은 출발지 노드에서부터 목적지 노드까지 가는 최단 경로를 찾아내는 그래프/트리 탐색 알고리즘 중 하나이다[4].

목표노드까지의 가장 좋은 경로를 추정하기 위해 각 노드에 랭킹을 부여하는 “Heuristic

장애우 및 고령자 측면에서의 모바일 SNS 앱접근성 평가

홍미희*, 박화진**

요 약

페이스북과 같은 모바일 SNS는 모든 사용자들이 시간과 장소에 구애받지 않고 타인과 협력하고 정보를 공유하기 위한 앱으로 유명하다. 그러나 일반인외에도 앱을 사용하는데 어려움을 가지고 있는 장애인과 고령자가 사용하기에 적절한 인터페이스를 제공하고 있는지에 대한 실태를 확인하기 위해 모바일 앱 접근성을 평가할 필요가 있다. 모바일 앱 접근성은 인지적, 신체적 특징에 상관없이 모든 사람에게 동등하게 모바일 애플리케이션에 대한 기회를 제공하고 있는가를 나타내는 지표이다. 이에 본 논문에서는 행정안전부에서 고시한 '모바일 애플리케이션 접근성 지침'과 한국정보화진흥원의 '모바일 애플리케이션(안드로이드) 접근성 점검 매뉴얼 (v1.0)'을 기반으로 페이스북과 카카오톡인 모바일 SNS의 모바일 앱 접근성을 평가하여, 모바일 앱 접근성 준수 현황을 파악하고, 개선방향을 제안하고자 한다.

키워드: 모바일 SNS, 모바일 앱 접근성

Mobile SNS Application Accessibility Evaluation for the Disabled and the Aged

Mee-Hee Hong*, Hwa-Jin Park**

Abstract

Mobile SNS like facebook is famous as an app which users cooperate others and share information without regard to time and place. But it is necessary to evaluate the mobile app accessibility in view of the disabled and the aged to check if they get appropriate interface service to use such a popular app. Mobile app accessibility is an indicator of equal opportunity to mobile app for all users regardless of cognitive and physical attribute. Therefore, this paper evaluates accessibility of facebook and kakaotalk mobile SNS, apprehend the status of conform to the manual, and suggest an improving way, based on 'mobile application accessibility guideline' announced by government and 'mobile application (Android) accessibility checking manual(v1.0)' by National Information society agency.

Keywords : mobile SNS, mobile app accessibility

1. 서론

휴대 단말 기술의 발전으로 인한 스마트폰들

※ 교신저자(Corresponding Author) : Hwa-Jin Park
접수일:2012년 11월 27일, 수정일:2012년 12월 14일
완료일:2012년 12월 24일

* 숙명여자대학교 멀티미디어학과
email: write-in@hanmail.net

** 숙명여자대학교 멀티미디어학과
Tel:+82-2-710-9204, Fax:+82-2-710-9704
email: phj2000@sm.ac.kr

■ 본 연구는 숙명여자대학교의 2011학년도 교내연구비 지원에 의해 수행되었음

이 출시되면서 모바일 환경의 콘텐츠들이 매우 빠른 속도로 보급되었다. 특히 모바일이 사람과 사람을 이어주는 교류의 공간으로 부각되면서, 관계형성을 도와주는 여러 서비스들과 애플리케이션들이 생겨났다. 이러한 서비스를 모바일 SNS(Social Network Service)라고 한다. SNS는 사회적 관계 개념을 인터넷 공간으로 가져온 것으로, 사용자 간의 관계 맺기를 통해 네트워크 형성을 지원하며 개인의 일상사나 관심사를 공유·소통시켜주는 서비스를 말한다[1]. SNS의 가장 큰 특징은 개인을 둘러싼 지인들이 어떤 제약도 없이 커뮤니케이션 할 수 있고, 그 지인들이 제공하는 관심사 및 정보를 함께 공유할

수 있다는 것이다. 이 모바일 SNS의 대표적인 국외 서비스로는 모바일 페이스북, 트위터가 있고, 국내 서비스로는 카카오톡, 라인, 미투데이가 있다.

모바일 SNS를 이용하는 이용자의 수도 지속적으로 늘어나고 있어, 모바일 SNS를 이용하는 사용자의 계층도 점점 더 다양화될 것으로 보인다. 그러나 장애인과 고령자는 신체적·인지적 특성의 문제점으로 인해 모바일을 사용하는 데 어려움을 겪고 있다. 이처럼 장애인과 고령자가 모바일 SNS를 이용하는 데에 어려움을 겪게 된다면, 온라인상에서 사람과 사람과의 관계를 맺고 유지할 수 있는 기회를 접하지 못하고, 여러 사람들과의 정보공유에서 제외되어, 정보 고립의 어려움도 겪으면서 자아감과 소속감을 상실할 것이다. 이 같은 어려움을 해결하기 위해 모든 사용자가 동등한 모바일 접근 기회를 제공받을 수 있도록 월드 와이드 웹 컨소시엄(W3C) 모바일 웹 이니셔티브(Mobile Web Initiative)에서 Mobile Web Best Practices 1.0을 2008년 7월 29일에 제정하였다. 이는 초창기 모바일 앱 접근성의 시발점이라 할 수 있다.

모바일 앱 접근성이란, 모바일 기기를 사용하여 모바일 애플리케이션을 이용하고자 하는 장애인, 고령자 등을 포함한 모든 사람들에게 이의 활용 가능성이 제공됨을 말한다[2].

모바일 SNS가 모바일 앱 접근성을 준수하여 제작되도록 유도하기 위해 먼저 모바일 SNS가 모바일 앱 접근성을 얼마나 준수하고 있는지 평가할 수 있어야 한다. 현재 국내에서 모바일 앱 접근성을 평가하는 기준으로는 '모바일 애플리케이션 접근성 지침'과 iOS, Android 각 운영체제별 '모바일 애플리케이션 접근성 점검 매뉴얼'이 있다. 현재, '모바일 애플리케이션 접근성 점검 매뉴얼'이 각 운영체제별로 1.0까지 발표되었으며, 이는 지체장애, 시각장애, 청력상실 등 보다 다양한 유형의 장애인과 고령자에 대한 모바일 앱 접근성 가이드라인을 제시하고 있다. 일반인보다 상실감과 고립감을 더 심하게 경험하는 장애인과 고령자 집단이 사람들과의 좀 더 활발한 정보공유를 하고 친밀한 관계를 유지하게 하기 위해서 모바일 SNS의 모바일 앱 접근성을 높여야 한다.

본 논문에서는 '모바일 애플리케이션 접근성

지침'에 따라 구글 안드로이드 프레임워크 기반의 '모바일 애플리케이션 접근성 점검 매뉴얼'을 기반으로 카카오톡과 모바일 페이스북의 모바일 앱 접근성을 평가하고, 부족한 점을 분석하여 개선안을 제시하고자 한다. 모바일 SNS의 모바일 앱 접근성 준수를 평가하기 위해 구글 안드로이드 프레임워크 기반의 '모바일 애플리케이션 (안드로이드) 접근성 점검 매뉴얼 (v1.0)' 기반으로 한 체크 목록을 이용해 안드로이드 운영체제에서 평가를 하였다. 이 평가를 통해 모바일 앱 접근성에 대해 논해봄으로써 모바일 SNS의 모바일 앱 접근성 문제점 및 개선점을 제안해 보고자 한다.

2. 관련연구

2.1. 안드로이드 프레임워크의 접근성 기능과 서비스

SNS(Social Network Service)란 사회 연결망으로, 가상공간인 인터넷에서 사람들의 활발한 참여를 기반으로 하는 유튜브, 페이스북, 카카오톡 등 각종 온라인-오프라인 매체 서비스를 일컫는다. 이 SNS는 온라인상에서 사용자와의 커뮤니케이션이 가능하도록 관계를 형성하는 동시에 사람과 사람들을 연결하여 그들 간의 관계 맺기를 지원하는 서비스이다[3].

안드로이드 1.6 버전(Donut)부터 텍스트를 음성으로 전환해주는 TTS(text to speech) 엔진이 탑재하여, 처음으로 접근성이 적용되었다. 이후 버전이 지속적으로 업데이트되어 TTS 엔진 위주로 접근성이 향상되었으며, 4.0 버전(Ice Cream Sandwich)부터 제품 최초 환경 설정시 접근성 기능 활성화 가능, 'Explore by touch' 기능 제공, 실시간으로 음성을 텍스트로 변환하는 기능 향상, 안면 인식 소프트웨어 탑재로 사용자의 안면 인식으로 잠금 해제 기능 등과 같이 다양한 접근성이 대폭적으로 확대되었다. 모바일 애플리케이션 접근성 점검은 안드로이드 2.3 (Gingerbread) 이후 버전에서 평가가 가능하다. 4.0 버전(Ice Cream Sandwich) 기준으로 안드로이드에서 제공하는 대표적인 접근성 기능에는 터치하여 탐색(Explore by touch), 스크롤, Directional Pad 탐색이 있다.

캐릭터스티커 기능을 활용한 창의력 증진용 모바일 e-동화 웹앱 구현

조희정*, 안다빈*, 김명원*, 박화진**

요 약

본 논문은 유아나 아동의 창의력을 키우기 위한 e-동화를 설계 구현하기 위한 연구로서 e-동화, 생각하는 질문, 캐릭터 그림판으로 구성되어있으며 특히 캐릭터 스티커를 사용하게 함으로써 그림을 쉽게 접할 수 있게 기능을 발전시킨 논문이다. 본 웹앱은 플래쉬로 제작된 동화를 보여주고 주어진 생각해보기 질문에 스티커를 변형시키면서 상상력있는 그림그리기로 답변하게 한다. 그림 그리기를 마치면 제목을 입력하고 저장할 수 있다. 질문을 통하여 한 번 더 생각하게 하고 생각한 내용을 표현할 수 있게 하는 교육용 콘텐츠로서 저장된 그림 목록을 통해 자녀의 생각이 얼마나 변화되고 다양해지는지 확인해 볼 수 있다.

키워드 : 창의성, e-동화, 캐릭터 스티커 그림판

An Implementaion of Mobile e-storytelling Application for Creativity Development using Character Sticker Function

Hee-Jung Cho*, Da-Bin Ahn*, Mung-won Kim*, Hwa-Jin Park**

Abstract

The purpose of this paper is to plan and implement e-storytelling for improving children's creativity, which consists of e-storytelling, thinking questions, and character sticker canvas. Especially It makes children draw paintings feel easier by character sticker. This webapp shows the flash animation e-storytelling and gives the question to answer by painting with a fertile imagination on the canvas as transforming the character sticker. A user can input the title of painting and store it when finish the drawing. This is an educational tool which makes educatees think more and express their ideas through questions. Therefore, we can check how much children's imagination is changed and diversified by the stored paintings.

Keywords : creativity, e-storytelling, character sticker canvas

1. 서론

요즘 시대에 가장 많이 요구되는 인간상은 아마도 “창의적 사고를 하는 인간”일 것이다. 21세

기는 정보화, 세계화 시대로 세계를 주도하는 경쟁력있는 인간을 효과적으로 양성해 내기 위하여 창의력 교육은 더욱 그 중요성이 커질 수 밖에 없다.

창의력이란 기존의 지식, 경험, 산물에 의거하여 놀랄 만한 새로운 발명이나 생산적, 독창적 사고 및 착상등의 고차원적인 사고 능력을 포함하여 일상생활에서 당면하는 여러 사태나 문제를 새롭게 참신한 방식으로 해결해 나가는 능력으로서 이런 능력은 대부분 확산적 사고에 속하며 이것은 수렴적 사고의 반대개념이다. 수렴적 사고는 사과의 결과가 하나의 정확한 해답이 되거나 또는 문제해결을 위한 가장 보편적인 방법이다. 이 두 가지 사고 방법은 모두 중요하지만

※ 교신저자(Corresponding Author): Hwa-Jin Park
접수일:2012년 12월 03일, 수정일:2012년 12월 14일
완료일:2012년 12월 31일

* 숙명여자대학교 멀티미디어학과

** 숙명여자대학교 멀티미디어학과

Tel: +82-2-710-9204, Fax: +82-2-710-9704

email: phj2000@sm.ac.kr

■ 본 연구는 숙명여자대학교의 2011학년도 교내연구비 지원에 의해 수행되었음

스토리텔링과 미술교육은 자라는 아동에게 있어 확산적 사고를 발달시킬 수 있는 매우 중요한 교육분야이므로 더욱 강조될 필요성이 있다.

이런 추세와 발맞추어 요즘 학부모들은 일찍부터 아이들의 창의력 향상을 위해 다양한 학습 도구와 서적을 이용하기를 원한다. 스마트 기기가 보편화 되어있는 이 시점에 창의력을 향상시켜주는 콘텐츠가 스마트 기기에서 사용된다면 학부모와 아동에게 도움을 줄 수 있을 것으로 보인다. 이러한 생각을 바탕으로 우리는 어린이가 동화를 읽고 동화의 내용을 토대로 주어진 질문에 생각하도록 유도하여 그 생각을 마침내 그림으로 표현하게 하는 콘텐츠를 구현하고자 한다. 또한 생각해 본 내용을 그림판에 그릴 때 빈 도화지에 시작하는 것보다 동화책에 등장하는 캐릭터를 스티커로 제공해 줌으로써 쉽게 표현하도록 하였다. 또한 스티커의 다양한 기능 즉 축소 확대 회전등을 추가하여 아이들이 자신의 생각을 쉽고 재미있게 표현할 수 있을 것이다. 따라서 생각해보는 질문이 있는 e-동화와 캐릭터 스티커 그림판을 통해 창의력을 향상시키는 것이 본 연구의 목표이다.

논문의 구성은 다음과 같다. 제 2장에서는 본 연구의 이론적 배경 및 창의력 향상을 위한 동화 및 그림그리기 앱에 대해 소개하고 분석한다. 제 3장에서는 본 논문에서 제안하는 캐릭터 스티커 e-동화 콘텐츠 서비스 및 기능에 대해 소개하고 제 4장에서는 본 콘텐츠 서비스의 구현 결과를 보인다. 마지막으로 제 5장에서는 향후 연구와 함께 결론을 맺는다.

2. 관련 연구

2.1 이론적 배경

2.1.1 창의력의 개념

창의력 개념에 대한 논의는 1952년 Ghiselin으로 거슬러 올라가 시작되어 다양한 관점에 따라 여러 개념이 제시되어 왔으나 인간의 능력이나 인성까지 종합한 고등정신으로써 명확하게 정의하기에는 다소 곤란한 개념이다. 단지 과거의 창의성에 대한 관점은 선천적이고 측정이 곤란하며 교육이 거의 영향을 주지 못하다는 것이었으나 현대에 접어들수록 모든 사람이 지닌 보

편적인 잠재능력으로 측정가능하고 교육에 의해 증진될 수 있는 것이라는 관점으로 변화되었다[2]고 말한다. 또한 창의적 행동은 개인의 잠재 능력과 개인의 동기 또는 욕구가 직접적으로는 가정, 학교와 같은 일차적 환경과 간접적으로는 사회, 문화와 같은 이차적 환경의 영향을 받아 나타나게 된다[3]. 즉, 여러 학자들의 의견을 종합하면 인간의 보편적 능력으로서의 창의력은 개인의 독특성에서 비롯되는 내부의 힘으로서 가정과 학교 및 사회, 문화를 포함한 외적 환경에 의해 영향을 받아 당면하는 과제를 해결하거나 자신에게 가치있는 새로운 생각을 산출하기 위하여 기존의 정보를 유용하게 이용하거나 새로운 아이디어를 만들어내는 능력으로 정의할 수 있다.

2.1.2 동화와 미술교육이 창의력 증진에 미치는 영향

1950년대 이후 인간의 창의성에 대한 개념이 논의되면서 최근에는 창의성의 개발가능성에 뜨거운 관심을 보이며 활발한 연구가 이루어지고 있다.

Guilford는 창의력을 구성하는 하위요인으로서 ①문제에 대한 감수성 ②어휘, 관념, 연상, 표현을 포함한 사고의 유창성 ③ 자발성, 적용성과 관련된 사고의 융통성 ④비범성, 원격연합, 기교성을 포함하는 사고의 독창성 ⑤ 면밀성 ⑥제구성력 ⑦집요성을 제시하였으며 정의적 영역으로는 호기심 위험감수, 복잡성 및 상상력을 창의성의 하위영역으로 제시하였다[4]. 즉, 창의성을 증진시킨다는 의미는 다양한 하위 요인 중 일부 요인에서 의미 있는 효과가 있다는 것을 의미한다. 창의성 개발 가능성에 대해 관심이 고조되면서 교육효과에 미치는 요인들, 즉 교육의 형태 및 시기, 교육환경등과의 연관성을 조사하는 연구도 활발해지고 있다. 이 중 동화들려주기 교육은 어휘, 관념, 영상과 같은 유창성에 효과가 있었음을 증명했고, 이 외에 정의적 영역에서의 호기심이나 상상력등의 지적인 욕구를 만족시킬 수 있었다. 특히 유아기에 부모나 교사가 들려주는 경우, 동화이야기에서 오는 효과 외에 정서적 안정과 발달측면에 큰 기여를 하게 된다[5].

미술교육의 중요성도 매우 크다고 할 수 있다. 어린이 미술교육은 아동의 지적, 정서적, 신체적,

디지털 말하기 책을 위한 음성 주석달기 시스템

(A Voice-Annotation System for Digital Talking Book)

이 경 희 ^{*} 이 종 우 ^{**} 임 순 범 ^{**}
(KyungHee Lee) (Jongwoo Lee) (Soon-Bum Lim)

요 약 독서장애인의 손쉬운 독서활동을 위해 개발된 디지털음성도서는 네비게이션 기능을 제공하여 시력 없이도 장, 절 등의 자유로운 이동을 가능하게 하였다. 그러나 북마크, 노트 등 사용자가 임의로 입력하는 어노테이션은 모두 화면을 중심으로 제공되고 있어 청력만을 사용하는 시각장애인의 이용에 한계가 있다. 이에 본 논문에서는 음성도서를 이용하는 사용자가 청각과 음성만으로도 하이라이트 할 수 있는 음성 어노테이션 시스템을 제안한다. 제안된 어노테이션 시스템은 입력, 분석, 렌더링, 저장, 출력 단계로 구성된다. 안드로이드 환경에서 구현된 시스템은 눈가리개를 착용한 실험자들을 통해 청력과 음성만으로도 효과적으로 어노테이션 할 수 있음을 확인하였다.

키워드 : 디지털음성도서, 음성어노테이션, 독서장애인

Abstract Digital talking books have been developed to enlarge the reading experiences for reading-disabled people. They also provide convenience through a chapter-to-chapter or section-to-section navigation functionality. The existing digital talking books support, however, bookmarks or notes only through a screen interfaces. This screen annotation interface is useless for reading-disabled people. In this paper, we suggest a voice annotation system that can insert highlights or notes during playing a talking book by using only hearing sense and voice. Our annotation system implemented on the android platform consists of 5 steps of input, analysis, rendering, storing, and output. We found by performing blindfolded users group's usability tests that our voice annotation system can effectively create annotations only through hearing sense and voice.

Key words : Digital Talking Book, Voice Annotation, Reading-disabled People

1. 서 론

디지털 문서의 어노테이션은 본문 내용 자체의 정보를 습득하는 데 그치지 않고 사용자들 간의 추가적인

정보를 공유 및 교환할 수 있게 해주므로 일회성 정보가 아닌 재사용될 수 있는 중요한 정보라 할 수 있다 [1]. 그러나 기존의 어노테이션 서비스들은 단지 일반 전자문서에서 화면을 중심으로 제공되고 있어 시각장애가 있는 사용자들에겐 무용지물인 기능이다. 이들을 위한 디지털음성도서에서조차 어노테이션은 화면을 통해서만 입력할 수 있다. 이에 본 논문에서는 디지털음성도서의 독서환경을 효율적으로 개선하기 위한 음성 어노테이션 시스템을 제시한다. 시각보다 청각 의존도가 높은 독서장애인에게 음성 어노테이션 입력을 지원함으로써 전자책-사용자 간 상호작용과 사용자들 간의 정보공유를 돕고, 효율적인 어노테이션 콘텐츠의 재사용을 위한 기술을 구현한다.

2. 관련연구

디지털음성도서(Digital Talking Book)는 아날로그 방식의 음성도서를 디지털 파일로 변환·발전시킨 형태

· 본 연구는 숙명여자대학교의 2012학년도 교내연구비 지원에 의해 수행되었음

^{*} 학생회원 : 숙명여자대학교 멀티미디어학과
khleewow@gmail.com

^{**} 종신회원 : 숙명여자대학교 멀티미디어학과 교수
bigrain@sm.ac.kr
(Corresponding author임)
sblim@sm.ac.kr

논문접수 : 2011년 9월 28일

심사완료 : 2012년 1월 26일

Copyright©2012 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지 : 컴퓨팅의 실제 및 레터 제18권 제4호(2012.4)

로, 대표적 전자책 표준 형식 DAISY(Digital Accessible Information System)[2]가 있다. DAISY는 디지털음성도서 재생 시스템의 어노테이션 구현에 필요한 북마크, 하이라이트, 노트에 대해 정의하고 있으나 구체적인 어노테이션의 입력 및 출력에 대한 내용은 언급하지 않고 있다. 표준에서는 어노테이션 파일의 저장 및 조작 등에 대해 제한을 두지 않으나 어노테이션을 지원하는 플레이어는 반드시 북마크 DTD[3]를 따라야 한다. 북마크 DTD에서는 어노테이션 유형으로 북마크 <bookmark>, 노트 <note>, 하이라이트 <hilit>를 정의하고 있다.

일반 전자문서에서 음성을 통해 어노테이션 하는 방법은 저장 방식에 따라 크게 두 가지로 구분할 수 있다. 음성 녹음 후 미디어 파일을 그대로 저장하는 방식과 음성인식을 기반으로 텍스트 형태로 변환하여 저장하는 방식이다. 녹음된 음성 미디어 파일을 그대로 저장하고 사용하는 방식은 시스템 구현은 비교적 간단하나, 녹음 당시 주변 소음으로 인한 음질 훼손의 변수와 추후 어노테이션 콘텐츠 정보를 재사용 하고자 할 때 비효율적이다. 이 방식은 단순히 오디오 노트의 기능만 제공하기 때문에 하이라이트 등 어노테이션의 다양성은 기대할 수 없다. 또한 녹음 파일이 많아질수록 텍스트 파일 대비 저장 공간이 낭비되기 때문에 전체적으로 시스템의 성능이 저하될 수 있다. 전자책 리더에서 음성 어노테이션을 지원하는 미국의 특허 E-BOOK READER WITH VOICE ANNOTATION/US2010-0324709[4]는 사용자가 특정 부분을 선택하여 하이라이트하고, 해당 하이라이트에 음성 노트를 입력하여 어노테이션 한다. 즉 하이라이트된 텍스트에 음성 노트 파일이 링크된다.

두 번째로 음성인식을 기반으로 음성노트 입력 후 변환된 텍스트 파일을 저장하는 방식이 있는데, 기존의 음성인식 기반 연구들은 주로 노트, 하이라이트 등과 같은 전자책 환경에서의 어노테이션이 아닌, 음성 검색이나 태깅을 위한 어노테이션으로 사용된다. 또한 이 방식의 연구 및 서비스들 역시 사운드 환경을 고려하지 않은 일반 전자문서를 기준으로 설계되었기 때문에 음성도서에 그대로 사용하기에 적합하지 않다. 시맨틱(의미) 기반의 이미지에 대한 음성 어노테이션 시스템[5]은 음성 기술이 사용되지 않았던 기존의 이미지 태깅 기술에 음성인식 기술을 적용한 것이다. 이 시스템은 자동 음성인식을 지원하고 음성인식의 정확도를 높였지만, 단지 효율적인 이미지 검색만을 위한 음성 노트를 다루고 있어 전자책이나 웹 환경에서의 어노테이션과는 그 의미와 목적이 다르다고 할 수 있다.

일반 전자문서와 음성도서에서의 어노테이션 입력에 있어 가장 큰 차이점은 도서 콘텐츠 열람의 접근 방식이다. 일반 문서는 시력을 사용하기 때문에 한 번에 여러

가지의 정보를 받아들이지만, 청각만을 사용하는 음성도서는 한 번에 한 개의 정보만을 입력하는 선형적 재생 환경을 가지고 있다. 이와 같은 차이로 일반 전자문서에서 지원되고 있는 다양한 어노테이션의 입력 방식은 음성도서 환경에서 동일하게 적용하는 데 한계를 가진다.

본 논문에서 제안하는 시스템은 재생 중인 음성도서에서 음성으로 어노테이션을 입력하고, 입력된 어노테이션의 위치를 정확히 분석하여 출력하는 데 중점을 두고 있다. 어노테이션의 대표적인 속성인 노트와 하이라이트를 주로 다루고 있으며 어노테이션 입력, 저장, 재생 모듈을 설계 및 구현한다. 구현된 전자책 재생기는 스마트폰 플랫폼인 안드로이드 2.3을 기반으로 XML(Extensible Markup Language) 형식의 전자도서 파일과 이것을 읽어주는 음성파일로 구성되는 음성도서를 지원한다.

3. 디지털음성도서 음성 어노테이션 시스템

본 논문에서는 전자책 환경에서 전맹인을 비롯한 독서장애인의 노트 및 하이라이트를 가능하게 하는 음성 어노테이션 시스템을 구축하였다. 제안된 시스템은 음성 인식 및 합성을 통하여 어노테이션 콘텐츠를 효율적으로 이용할 수 있도록 설계하였다. 본 논문에서 제안하는 기법은 오디오 재생이 지원되는 독서장애인을 위한 디지털음성도서 환경에서 음성 어노테이션을 가능케 하고, 음성 인식 및 변환 과정을 통해 저장되므로 어노테이션 콘텐츠의 재사용 및 저장 공간에 있어 효율성을 높였다.

또한 기존의 어노테이션들은 모두 비주얼 중심의 화면 포지셔닝(display positioning) 기법을 사용하였으나, 본 논문에서 제안하는 음성 어노테이션은 100% 청각만으로 어노테이션 할 수 있도록 사운드 환경에 입각한 시간 포지셔닝(time positioning) 기법을 적용하였다. 이것은 음성도서 미디어 파일의 각 문장에 대한 시간 정보를 이용한 것이다.

어노테이션 시스템의 모델은 그림 1과 같이 기능에 따라 크게 5개의 모듈로 구성되는데, 순차적으로 입력(INPUT), 분석(ANALYSIS), 렌더링(RENDERING), 저장(STORE), 출력(OUTPUT)이 그것이다. 입력 받은 음성은 텍스트 형식으로 변환되어 분석모듈로 전달된다. 분석모듈에서는 입력받은 어노테이션의 형식과 위치를 분석한다. 분석된 결과를 기반으로 렌더링 모듈에서 각 어노테이션 형식에 따라 사운드 및 디스플레이 효과를 적용한다. 어노테이션의 모든 정보를 저장하고 데이터베이스를 관리하는 저장단계를 거쳐 출력 모듈은 사용자에게 어노테이션의 결과물을 출력한다.

그림 2는 본 논문에서 제안된 음성 어노테이션 시스템의 전체 흐름도이다. 모듈에 따라 크게 다섯 부분으로 나눌 수 있고, 각 단계 별 생성하는 파일과 처리 방식에

내용 MathML에서의 수식 독음 규칙 및 변환*

서승희⁰ 박수현 임순범

숙명여자대학교 멀티미디어학과

sluna1215@gmail.com, katepark.818@gmail.com, sblim@sookmyung.ac.kr

The Reading Rule and Transformation of Mathematical Expression in the Content MathML

SeungHee Seo⁰ Soohyun Park SoonBum Lim

Multimedia Science, Sookmyung Women's University

요 약

최근 독서 장애인을 위한 국제 디지털 음성 도서 표준인 DAISY를 통해 장애인의 고등 교육을 위한 연구가 활성화되고 있으나 그림이나 수식과 같은 특수요소는 음성도서로 만들어 활용하기 어렵다. 이를 위해, 본 연구에서는 중학교 교과과정에 포함된 내용 MathML 엘리먼트에 대한 한글 독음 규칙을 정의하였다. 또한, 정의한 한글 독음 규칙에 따라 내용 MathML로 표현된 수식을 독음 텍스트로 변환하는 프로그램을 XSLT로 구현하였다.

1. 서 론

독서 장애인의 독서 환경 개선을 위해 다양한 디지털 음성 도서 기술이 개발되고 있고, 이 중 DAISY가 국제 디지털 음성도서의 표준으로 널리 사용되고 있다. 또한 DAISY를 활용한 장애인의 고등 교육을 위한 전자 교과서 연구가 활성화되고 있다. 그러나 DAISY에서 그림이나 수식 등의 특수 콘텐츠의 독음에 대한 연구가 제대로 이루어지지 않고 있다. 특히 DAISY 내의 수식 콘텐츠의 독음에 대한 국내외 연구가 매우 미흡한 상태이다.

DAISY 내의 수식 콘텐츠는 웹에서도 널리 사용되는 MathML으로 표현된다. 본 연구에서는 중학교 교과 과정에 포함된 수식을 대상으로 MathML로 표현된 수식 콘텐츠의 독음을 위해 수식 독음 규칙을 정리하고, 이에 따른 독음 변환 기법을 연구하였다.

2. 관련연구

2.1 MathML

MathML은 XML의 응용 중 하나로 수학 수식을 표현하기 위한 마크업 언어이다. MathML은 단순한 수식의 표기뿐만 아니라 수식의 의미도 표현할 수 있다.

MathML은 두가지 마크업 언어로 구성되는데, 표현(Presentation) MathML과 내용(Contents) MathML이 있다. 표현 MathML은 수식의 표기에 중점을 둔 마크업이다. 내용 MathML은 수식의 의미를 표현하는 데에 중점을 둔 마크업 언어로 다른 응용 프로그램에 의미를 전달 가능하다. 내용 MathML은 대부분 전위표기법(prefix notation)으로 되어있다.

2.2 수식의 독음 서비스

해외의 수식 독음 서비스로는 DesignScience사의 MathPlayer에 적용된 MathSpeak의 문법 규칙이 있다. MathSpeak는 2004년에 시작된 gh사의 MathSpeak 프로젝트를 통해 고안된 독음스타일이다. MathSpeak의 문법규칙은 충분한 독음(verbose), 중간 단계(brief), 간략 단계(super-brief)의 세 가지 단계의 독음 스타일을 제공하여 청자의 상황이나 요구조건에 따라 선택하여 수식을 들을 수 있도록 하였다[1].

2.3 MathExpression Reader

국내에서는 일반 문서 편집기 내의 수식 편집기에서 사용되는 수식 표현법을 독음텍스트로 추출하는 연구를 본 연구진에서 진행하였다. 이 연구는 중학교 교과과정에 포함되는 수식을 대상으로 하였다. 이 연구에서는 국내의 환경에 맞게 수식 독음 규칙을 정리하고, 정리된 독음 규칙에 따라 수식 편집기의 식에서 독음 텍스트를 추출하는 프로그램을 구현하였다[2].

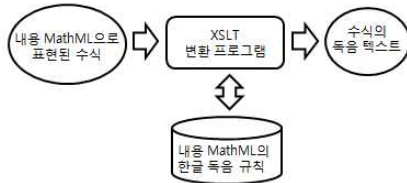
3. 연구 목표 및 범위

본 연구에서는 내용 MathML의 한글 독음 규칙을 정의하고, 내용 MathML으로 표현된 수식에서 독음 텍스트를 추출하는 것을 목표로 하였다.

기존 연구 중 MathSpeak의 문법 규칙을 분석한 결과, 절대값이나 역함수 등의 수식은 보이는 순서대로 독음하도록 되어 있어 의미에 따른 수식 독음이 미흡한 것으로 나타났다. 한편 본 연구진에서 선행한 MathExpression Reader 연구의 경우, 현재 DAISY나 웹에서 활용되고 있는 MathML에서의 별도의 작업이 필요하다. 따라서, 본 연구에서는 150

* 이 논문은 2011년도 정부(교육과학기술부)의 재원으로
한국연구재단의 지원을 받아 수행된 연구임(No.20110026951)

여 개의 내용 MathML 엘리먼트 중 중학교 교과과정에 포함된 엘리먼트에 대한 한글 독음 규칙을 정의하였으며 그룹핑 독음 등 특수한 경우는 별도로 처리하였다. 또한, 내용 MathML으로 표현된 수식을 독음 텍스트로 변환하는 프로그램을 XSLT로 구현하였다.



[그림 1] 내용 MathML 수식의 독음 과정

4. 내용 MathML 독음 규칙

중학교 교과 과정에 포함된 엘리먼트를 분석하여 메타 엘리먼트, <apply> 엘리먼트, 피연산자, 연산자 총 4가지로 분류하였다. 이 각각에 대하여 표현 MathML 독음 규칙을 정의하였다.

4.1 메타 엘리먼트 (Meta Element)

루트 엘리먼트인 <math>와 수식의 부가정보를 표현하는 <annotation>, <semantics>, <annotation-xml>은 한글독음으로 변환하지 않는다.

4.2 <apply> 엘리먼트

<apply>엘리먼트의 시작 태그와 끝 태그는 연산자의 범위를 결정한다. 전위표기법을 사용하기 때문에 첫 번째 자식 엘리먼트(child element)에는 연산자가 나머지 자식 엘리먼트에는 그 연산에 대한 피 연산자가 온다. <apply>엘리먼트는 한글독음으로 변환하지 않는다.

4.3 피연산자 (Operand)

<apply>의 나머지 자식 엘리먼트에는 첫 번째 자식 엘리먼트인 연산자에 대한 피 연산자가 온다. 피 연산자에는 토른 엘리먼트가 단말 노드로 오거나 <apply>엘리먼트가 형성한 식 트리(expression tree)가 중첩되어 올 수 있다.

토른 엘리먼트에는 수를 나타내는 <cn> 과 변수를 나타내는 <ci>가 있으며 토른 엘리먼트는 그 텍스트 값을 그대로 출력한다. 또한 중첩된 <apply>에는 그룹핑 독음인 ‘괄호 열고’와 ‘괄호 닫고’를 매핑하여 연산자의 범위를 명확하게 한다. <표 1>은 <apply>의 독음규칙이다.

<표 1> apply 엘리먼트의 독음 규칙

Contents Markup	한글 독음
<apply> operator argument(s) </apply>	괄호 열고 operator argument(s) 괄호 닫고

4.4 연산자 (Operator)

<apply>의 첫 번째 자식 요소에는 연산자가 온다. 연산자마다 피 연산자의 수가 다르며 한글 독음은 <표 2>와 같다.

<표 2> 내용 MathML 독음 규칙

Contents Markup	수학 기호	한글 독음	비고
Basic Element 기본 요소			
<fn>[f] </fn>[x]	f(x)	함수 [f] [x]	
...(생략)			
Arithmetic, Algebra and Logic 산술 논리 대수			
<minus>[a]	-a	마이너스 [a]	마이
<minus>[a][b]	a-b	[a]마이너스 [b]	너스
<plus>[a][b][n]*	a+b(+n)*	[a] 플러스 [b] (플러스 [n])*	
<times>[a][b][n]*	ab(n)*	[a]곱하기 [b] (곱하기 [n])*	
<abs>[a]	a	절대값 [a]	
...(생략)			
Relations 관계(기호논리학)			
<eq>[a][b][n]*	a=b(=n)*	[a]은/는 [b](이 고 [n])*이다	조사
<neq>[a][b]	a! =b	[a]은/는 [b]가 아니다	조사
...(생략)			
Theory of Sets 집합 이론			
<in>[a][A]	a ∈ A	[a]은/는 [A]의 원소이다	조사
<union>[A][B][n]*	A ∪ B(Un)*	[A] 합집합 [B] (합집합 [n])*	
<gt>[a][b][n]*	a > b(>n)*	[a]은/는 [b]보다 (크고 [n]보다)* 크다.	조사
<lt>[a][b][n]*	a < b(<n)*	[a]은/는 [b]보다 (작고 [n]보다)* 작다.	조사
...(생략)			
Elementary Classical Functions 기본 고전 함수			
<sin>[x]	sin	사인 [x]	
<cos>[x]	cos x	코사인 [x]	
...(생략)			
Statics 통계			
<sdev>[x][y]*	V(X, Y, Z)	[x] [과/와 y]*의 표준편차	조사
...(생략)			

*단말노드: <cn>이나 <ci>노드/ 비단말 노드: <apply>나 <reln>

[a], [A]: 비단말노드 a, A: 단말노드 A : 집합 a : 원소

4.5 특이 사항

앞서 정리한 독음 규칙 중 조사, <minus> 엘리먼트 등은 추가적인 처리가 필요하다.

첫째, 한글에서 많은 조사가 은/는, 이/가, 과/와, 을/를 등과 같이 앞 음절의 종성에 따라 달라지는 변이형을 가진다. 따라서 앞 음절의 종성의 유무에 따라 다르게 독음해준다. 둘째, 내용 MathML에서 <minus> 엘리먼트는 피연산자 수

스마트폰 음악 플레이어에서의 주변환경을 고려한 멀티모달 자동제어기법

곽주은[○] 박찬영 송병인 유정현 이주은 임순범

숙명여자대학교, 세종대학교, 숭실대학교, 숭실대학교, 서울대학교, 숙명여자대학교
kwakje88@gmail.com, univstar@naver.com, byunginhb@gmail.com, desireso@gmail.com,
ljefamily@gmail.com, sblim@sm.ac.kr

Multi-modal automatic control of the smartphone music player considering the noisy environment

Kwak JooEun[○], Park chanyoungk, Song Byungin, Yu Jeongheon, Lee Jueun, Lim Soonbum
Sookmyung Univ., Sejong Univ., Soongsil Univ., Soongsil Univ., Seoul Univ., Sookmyung Univ.

요 약

기존의 스마트폰 음악 플레이어는 이어폰을 끼고 장소를 이동할 때 주변의 경고소리를 잘 듣지 못하는 위험함이 있고, 플레이어 컨트롤을 위해서는 화면을 보아야만 하는 불편함이 많았기 때문에 이를 해결하고자 멀티모달 자동제어 기법을 사용한 새로운 플레이어를 제안하였다. 주변의 dB(데시벨)을 입력받아 자동으로 볼륨을 제어해주며, 음성 명령과 터치 제스처 모드를 통해 화면을 보지 않고도 플레이어의 일부 기능을 제어할 수 있도록 하였다.

1. 서 론

최근 다양하게 출시되는 음악 플레이어 어플리케이션으로 인해 스마트폰으로 음악을 듣는 사용자가 증가하면서 이어폰 혹은 헤드폰 사용자도 증가하였다[1]. 그러나 이어폰을 통해 음악을 크게 듣는 경우, 주변의 청각적 환경을 인지하지 못하여 위험상황에 노출되는 일들이 빈번하게 일어나고 있다. 또한, 추운 겨울날 등 손으로 화면을 터치하기에는 불편한 상황들이 자주 일어난다.

본 논문에서는 이러한 기존 음악 플레이어의 문제점들을 해결하기 위해 음성과 터치 제스처의 멀티모달 인터페이스를 제안하고자 한다. 이를 통해 dB측정으로 위험상황 판단 시 볼륨을 줄여줄 수 있고, 이어마이크를 통한 음성인식과 터치 제스처 인식으로 플레이어의 일부 기능을 제어할 수 있게 한다.

2. 관련 연구

2.1 음성인식 기술 동향

멀티모달 인터페이스의 대표적인 음성인식 기술은 애플사의 아이폰 Siri를 예로 들 수 있다[2]. 사용자가 말하는 내용에 따라 해당 기능이 제공되는 Siri는 단순히 사용자의 음성을 인식하는 것뿐 아니라 상황맥락과 사용시간, 이용사전호도를 파악하여 답변을 제시하는 기능도 갖추고 있다[3].

Moonblink는 안드로이드 기반 스마트폰으로 음성의 크기

를 측정할 수 있는 라이브러리이다[4]. 스마트폰의 마이크를 통해 실시간 dB단위의 수치들을 측정한다. 최근 이 라이브러리를 통해 Sound Meter, Audalyzer 등의 음성관련 어플리케이션이 지속적으로 출시되고 있다.

LPC(Linear Predicative Coding)는 선형예측 코딩 기법으로써 과거의 일정개수의 샘플들을 이용해 현재의 값을 예측하기 위한 코딩 기법이다[5]. 과거의 여러 개의 샘플들에 적절한 계수들을 곱하면 현재의 값에 유사한 값을 유추할 수 있는데, 음성을 전송할 때 전송할 데이터의 양을 감축시키기 위해서 개발된 코딩 기법이다. 음성처리 분야에서는 이 LPC계수의 특징성을 이용해 두 개 음성의 유사성을 판단할 때 사용하기도 한다.

2.2 제스처 인터페이스 동향

멀티모달 인터페이스로 많이 쓰이는 또 다른 기술은 제스처 인터페이스이다. 기존의 터치 인터페이스 기술은 스크린을 직접 터치하여 기기와 인터랙션을 주고받는다. 점에서 사용자에게 직관적인 사용 경험을 제공하였다. 그러나 스크린을 보지 못하는 상황 혹은, 손을 쓸 수 없는 상황에서 기기를 컨트롤하기에는 불편한 문제점이 있다. 이로 인해 제스처를 활용한 인터페이스 활용사례가 늘고 있다.

마이크로소프트에서 출시한 키넥트는 사용자의 제스처를 기반으로 NUI(Natural User Interface) 라이브러리를 따로 제공하기도 한다.

3. 연구 목표

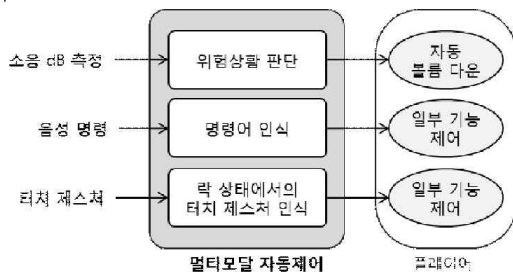
멀티모달 자동제어기법으로 소음 dB측정과 음성인식, 터치 제스처 인식을 통해 일부 기능을 자동으로 제어 가능한 플레이어를 만든다.

주변 소음 dB 측정을 통한 플레이어 볼륨 제어는 이어마이크를 통해 주변 소음을 실시간으로 수집하여 이루어진다. dB의 값이 일정 수치보다 높아질 경우, 플레이어는 해당 소음을 경고성 알림으로 판단하여 위험상황으로 인지한다. 위험상황으로 인지된 플레이어는 볼륨을 자동으로 줄여 사용자가 주변의 경고성 알림을 들을 수 있도록 한다.

음성인식을 통한 플레이어 제어는 이어마이크를 통해 실시간으로 입력받는 소리 중 사용자의 명령어를 구분하여 이루어진다. 일반 소음과 달리 명령어로 판단된 음성은 사용자가 미리 녹음해놓은 단어들과 비교되어 해당 기능을 제어 할 수 있다.

플레이어의 일부 기능을 제어하기 위한 또 다른 방법으로 터치 제스처모드를 제공한다. 터치 제스처는 스마트폰의 락 상태에서 첫 터치 위치와는 상관없이 인식되어 화면을 보기 힘든 상황에서도 플레이어를 쉽게 제어 한다.

[그림 1]은 소음 dB 측정과 음성명령, 터치 제스처의 멀티모달로 플레이어의 일부 기능이 자동 제어가 되는 모습이다.



[그림 1] 멀티모달 자동제어를 통한 플레이어 제어

4. 설계 및 구현

멀티모달 자동제어기법을 사용한 플레이어는 Android 2.2 플랫폼에서 Java와 C, C++를 사용하여 개발하였다.

4.1 주변 상황에 따른 볼륨 자동 컨트롤

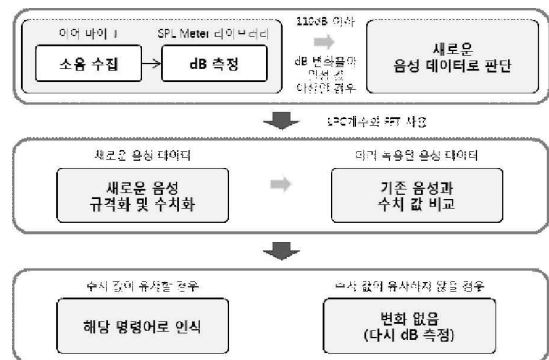
주변의 소음 dB는 SPL Meter 라이브러리를 사용하여 측정한다[6]. 실시간으로 측정된 dB의 일정 값(110dB으로 설정)을 넘어갈 경우 위험상황으로 인식하여 2초간 볼륨을 자동으로 줄여준다. [그림 2]는 위험상황으로 판단하는 과정이다.



[그림 2] 위험상황으로 판단하는 과정

4.2 사용자의 음성을 통한 플레이어 컨트롤

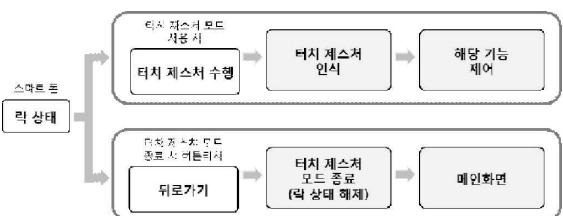
SPL Meter 라이브러리를 사용하여 측정한 dB의 값이 110dB 이하이며 변화율이 일정 값 이상일 경우, 플레이어는 해당 소음을 새로운 음성 데이터로 판단한다. 새로운 음성 데이터는 수치화 되어 기존에 녹음되었던 음성데이터 수치와 비교된다. 이 때 LPC계수와 FFT(Fast Fourier Transform)를 통해 두 개의 음성데이터 값을 계산한다[7]. 새로운 음성 데이터가 녹음된 음성 데이터의 수치 값과 유사할 경우, 해당 명령어로 인식되어 플레이어 제어가 가능하다. [그림 3]는 이러한 과정을 나타낸다.



[그림 3] 음성데이터 수치화를 통한 명령어인식 과정

4.3 터치 제스처 모드에서의 플레이어 컨트롤

터치 제스처 모드는 스마트폰의 락 상태에서 사용한다. Flicking과 Double tap, Long press의 사용으로 미리 지정된 플레이어의 일부 기능을 제어한다. 터치 제스처 모드를 사용하지 않을 경우는 스마트폰 기기의 '뒤로가기' 하드 버튼을 터치하여 일반 메인 화면으로 넘어간다. [그림 4]는 이러한 과정을 나타낸다.



[그림 4] 터치 제스처 모드 사용 과정

5. 실험 결과 및 분석

멀티모달 자동제어기법이 적용된 음악 플레이어에서는 위험상황으로 인지할 dB 값이 고정되어 있어야하며, 음성 명령과 터치제스처의 인식률이 높아야한다. 5.1과 5.2, 5.3은 이를 실험한 결과이다.

독서장애인 전자책을 위한 음성인식을 이용한

어노테이션 브라우징 기법⁺

박주현[○], 이종우, 임순범

숙명여자대학교 멀티미디어학과

e-mail : park.joohyun5@gmail.com, bigrain@sm.ac.kr, sblim@sm.ac.kr

An Annotation Browsing Technique in e-book for Reading-disabled People Using Voice Recognition

Joo Hyun Park[○], JongWoo Lee, Soon-Bum Lim

Dept. of Multimedia Science, Sookmyung Women's University

요 약

본 연구에서는 독서장애인을 위한 전자책용 어노테이션의 탐색 및 재생 기법을 제안하고 이를 음성 어노테이션 브라우징 시스템이라 칭하였다. 제안된 음성어노테이션 브라우징 시스템은 명령 입력, 중요도 분석 및 추천, 검색, 출력 단계로 구성된다. 특히 본 연구에서는 대상 사용자가 청각 의존도가 높은 독서장애인들이기 때문에 완전히 청각에 의존해서 사용할 수 있도록 모든 단계에서 음성인식 기능을 제공한다. 제안된 음성 어노테이션 브라우징 시스템의 효율성을 검증하기 위해 안드로이드 환경에서 실행되는 전자책 소프트웨어와 음성 어노테이션 브라우징 시스템을 설계하고 구현하였다.

1. 서 론

디지털 토크북의 개발은 청각 의존도가 높은 독서장애인들에게 환영 받고 있다. “책에 사용자가 노트나 주석 등을 기록하는 것”을 의미하는 어노테이션(annotation) 기능 역시 독서장애인이 사용하려면 음성 어노테이션이 필수적이다. 또한 기록된 어노테이션을 독서장애인이 찾아볼 수 있는 기법도 제공 되어야만 온전한 독서장애이용 전자책이라 할 수 있다. 음성 어노테이션의 기록 방법에 대한 연구는 현재 진행되고 있지만, 기록된 음성 어노테이션을 독서장애인이 탐색하는 방법에 대한 연구는 거의 없는 형편이다. 이에 본 연구에서는 독서장애인의 효율적인 독서환경 개선을 위한 음성인식을 이용한 어노테이션 브라우징 시스템을 설계하고 구현한다.

제안하는 시스템은 음성지령과 인식을 통해 사용자가 원하는 어노테이션을 단어 검색 등 다양한 방법을 통해 빠르고 정확하게 찾아가는데 중점을 두고 있다. 기록된 어노테이션을 찾아가는 방법으로 키워드 검색 방법과 추천어 검색 방법을 지원하며, 이를 달성하기 위한 어노테이션 키워드 입력, 검색, 출력, 중요도 분석 및 추천 등의 모듈 등을 설계하고 구현한다.

2. 관련연구

기존의 어노테이션에 관한 연구들은 어노테이션 표현 방식에 따라 텍스트, 스타일, 펜, 컨텍스트 순서로 발전 [1]되어 왔으나 이 모든 발전은 사운드 환경을 고려하지 않는 비장애이용 전자책 기반 어노테이션 기법이었다. 이를 개선하기 위하여 음성 기반 어노테이션 입력도 지원하는 연구가 진행되었는데 [2]가 대표적이다. 이 연구에서는 음성 인식 기술을 이용하여 사용자가 음성 어노테이션을 입력하고 이를 텍스트로 변환하여 저장 한다. 사용자는 본문의 내용을 오디오로 듣는 도중 ‘노트’라고 음성으로 지령 하면 음성인식기가 실행되면서 사용자가 말한 내용을 텍스트로 변환 하는 방식으로 동작한다. 본 연구에서는 연구 [2]의 텍스트로 변환 된 어노테이션 정보를 음성 인식 기술을 이용하여 독서장애인이 브라우징 할 수 있도록 하였다.

3. 어노테이션 브라우징 시스템 설계

3.1 사용자 요구 사항

다음 3가지 표를 통해 표1, 2, 3은 어노테이션 브라우

⁺ 이 논문은 2011년도 정부(교육과학기술부)의 재원으로
한국연구재단의 지원을 받아 수행된 연구임(No.20110026951)

정 시스템에 대한 사용자의 요구 사항을 정리한 것이다. 사용자 분류의 기준은 잔존시력의 유무와 콘텐츠에 대한 이해도이다.

<표 1>은 전자책 콘텐츠를 열람하는 방식에 따른 독서장애인의 주요 요구사항[2]이며, <표 2>는 사용자별 어노테이션 브라우징 방식을 나타내고 있으며, 이것을 기준으로 시스템을 설계 하였다. 전맹인은 음성으로만 어노테이션을 입출력 하지만, 저시력인 및 학습장애인은 음성과 동시에 화면을 이용할 수 있다[2]. 따라서 모든 독서장애인군에서 음성 인터페이스가 사용 가능하다는 결론을 도출 할 수 있다.

<표 1> 독서장애인 사용자별 주요 요구사항

사용자분류	주요 요구사항	콘텐츠 열람방식
전맹인	화면 필요 없음	오디오
저시력인	고대비 화면	화면, 오디오
학습장애인	단어-오디오 간 동기화	

<표 2> 사용자별 어노테이션 탐색/재생 방식

사용자 분류	탐색인터페이스	재생인터페이스
전맹인	음성	음성
저시력인/노인	음성/텍스트	음성/텍스트
난독증/학습장애인	음성/텍스트	음성/텍스트

또한, 독서장애인이 직접 모바일 전자책 어노테이션 서비스를 사용할 경우 실질적으로 필요한 요구 사항을 분석한 결과 <표 3>과 같은 요구 사항이 만족되어야 할 것으로 판단된다.

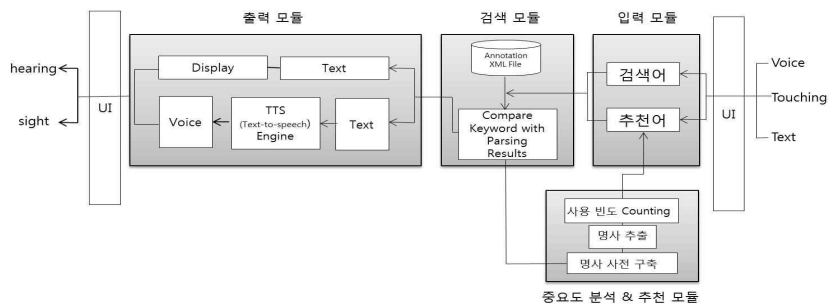
<표 3> 독서장애인용 음성 어노테이션 사용 시 요구 사항

실행 단계	요구 사항
모바일 전자책 사용시 요구 사항	멀티미디어 형식의 콘텐츠 내용을 텍스트로 제공 색상을 사용한 의미의 전달이 흑백 화면에서도 동등하게 제공 폰트의 크기 조절, 확대 기능 제공
모바일 전자책 어노테이션 브라우징 시 요구 사항	기록한 노트의 유무 기록한 노트의 개수 정확한 화면 터치에 제약 음성만으로도 브라우징 가능해야 함 어노테이션내 키워드 검색 기능 페이지 전환, 음성 인식기 시작 등 화면 전환시 알림 제공

3.2 시스템 구성

어노테이션 브라우징 시스템의 설계는 각각의 기능에 따라 4개의 모듈로 구성된다. 입력(Input), 중요도 분석 & 추천(Ranking & Recommendation), 검색(Search), 출력(Output)이 그것이다. 시스템에서 사용하는 모든 입력 행위는 음성지령을 통해 이루어지도록 설계 하였다. [그림 1]은 음성 어노테이션 브라우징 시스템의 전체 구성 모듈을 보이고 있다.

입력 모듈은 사용자로부터 음성 및 터치 지령을 받는다. 사용자가 찾고자 하는 어노테이션의 키워드를 음성으로 입력하면 이를 텍스트로 변환하여 검색 모듈로 전달한다. 검색 모듈에서는 기존에 저장된 어노테이션 XML파일을 파싱한 후, 파싱된 내용과 입력 받은 검색어의 매칭 작업을 실행한다. 매칭 된 결과를 바탕으로 그 결과를 출력 모듈로 전달하여 TTS 엔진을 거쳐 음성으로 출력하게 한다. 중요도 분석 및 추천 모듈은 입력 모듈과 검색 모듈 사이에 존재하는데, 저장되어 있는 어노테이션 파일의 내용을 분석하여 사용자가 원하는 특별한



[그림 1] 음성 어노테이션 브라우징 시스템의 전체 구성 모듈

국내 독서장애인을 위한 Math Expression Reader의 구현 및 사용성 평가

이재화[†], 이종우^{††}, 임순범^{†††}

요 약

국내에서 제작되는 전자도서들은 현재 문서 내에 작성된 수식 및 수학 기호들을 음성으로 변환하지 못하여 독서 장애인들에게 제약적인 음성서비스를 제공하고 있다. 본 논문에서는 국내 독서 장애인들을 위해 일반 문서에 삽입되어 있는 수식표현을 한글로 읽어줄 수 있는 'Math Expression Reader'를 구현하였다. 그리고 'Math Expression Reader'를 통해 생성된 한글 수식 독음을 일반인그룹과 시각장애인들에게 각각 들려줌으로 생성된 한글수식독음을 얼마나 정확히 이해하고 알아들을 수 있는지 평가하고 그 결과를 비교 분석하였다.

Implementation & Usability Evaluation of Math Expression Reader for Domestic Reading Disables

Jae-hwa Lee[†], Jong-Woo Lee^{††}, Soon-Bum Lim^{†††}

ABSTRACT

E-books produced in the country provide limited audio service for reading disables. The reason is that those books cannot translate the mathematical expressions and symbols in the context. In this paper, the 'Math Expression Reader' was implemented that can translate the expressions and symbols in the document into Korean speech for those who have reading disabilities. The math to speech generated by this program has been tested to both the public and reading disables and the results of this test has been compared whether they can exactly understand the speech and evaluated the reading rules.

Key words: Reading disables(독서 장애인), Math to speech(수식의 음성 서비스)

1. 서 론

현재 독서 장애인들의 정보 접근성을 높이기 위해 대부분의 독서 장애인 지원 기관들 및 도서관들은 음성 도서 서비스를 제공 받을 수 있는 디지털 파일 형태의 전자도서를 제공하고 있다. 그러나 대부분 국내에서 제작되는 전자 도서는 본문 텍스트만 음성

으로 변환하며, 전자 책 내에 있는 수식 및 수학 기호들은 음성으로 변환하지 못하고 아예 생략 해버리거나, 그림이나 이미지와 같이 간단한 텍스트로 제공하고 있다[1-4]. 이미 외국에서는 수식을 음성으로 변환하여 서비스하는 다양한 시스템들이 개발이 되어 있지만, 이러한 외국 시스템을 그대로 국내에 적용하게 될 경우, 언어 구조의 차이로 인해 영어로 표현된

※ 교신저자(Corresponding Author): 임순범, 주소: 서울시 용산구 청파동 2가 53-12 숙명여자대학교, 전화: 02)710-9424, FAX: 02)710-9704, E-mail: sblim@sookmyung.ac.kr

접수일: 2011년 7월 18일, 수정일: 2011년 9월 28일
완료일: 2012년 5월 29일

[†] 준회원, 숙명여자대학교 멀티미디어학과
(E-mail: envywiki@gmail.com)

^{††} 종신회원, 숙명여자대학교 멀티미디어학과 교수
(E-mail: bigrain@sookmyung.ac.kr)

^{†††} 종신회원, 숙명여자대학교 멀티미디어학과 교수
※ 본 연구는 숙명여자대학교 2011학년도 교내연구비 지원에 의해 수행되었음.

수식을 다시 한국어 표현에 맞게 재가공해야 하는 문제가 발생한다.

이에 본 논문에서는 국내 독서 장애인들을 위해 일반 문서에 삽입되어 있는 수식표현을 한글로 읽어 주는 수식의 음성 변환 기법을 제안한다. 이를 위해 현재 중학교에서 사용되고 있는 20개의 수학교과서를 분석 및 분류하여, 각 영역에서 쓰이는 수학 기호를 정리하여 한글 독음 규칙을 정의하였으며 이러한 한글 독음 규칙을 바탕으로 일반 문서에 입력된 수식을 한글 표현으로 읽을 수 있는 ‘Math Expression Reader’를 구현하였다.

그리고 ‘Math Expression Reader’를 통해 생성된 한글 수식 독음을 일반인과 시각장애인들에게 각각 들려줌으로 생성된 한글수식독음을 얼마나 정확히 이해하고 알아들을 수 있는지 평가하고 그 결과를 비교 분석하였다.

본 논문의 구성은 다음과 같다. 2장에서는 현재 외국에서 사용되고 있는 수식의 음성 서비스에 대해 알아보고 3장에서는 한글 독음 규칙을 이용한 Math Expression Reader 구현에 대해 설명한다. 4장에서는 이를 바탕으로 일반인그룹과 시각장애인들을 대상으로 실시한 평가 및 그 결과를 기술하며 5장에서 결론과 향후연구에 대한 정리로 본 논문을 끝맺도록 한다.

2. 관련연구

2.1 Design Science사의 MathDaisy & Math-Player

MathPlayer[5]는 웹브라우저에서 MathML[6]을

포함하는 웹 페이지를 브라우징 할 때, 해당 페이지의 수식을 디스플레이 해주며, 수식 읽어주기(speak expression) 기능을 통해 MathML로 표현된 수식을 읽어준다.

MathDaisy[7]는 MS Word에서 작성한 수식을 DAISY[8] 포맷으로 변환하여, 음성 서비스가 가능하도록 한다. MathDaisy는 수식 입력기인 MathType을 MS Word에 플러그인하여 수식을 표기하며, 작성된 수식은 DAISY 포맷으로 저장되어 DAISY 플레이어 혹은 DAISY 포맷을 지원하는 플레이어로 음성 서비스를 가능하게 한다.

2.2 Design Science 사의 Speech Style

현재 Design Science사의 MathPlayer의 최신 버전인 MathPlayer 3.0에서는 MathSpeak[9]와 SimpleSpeech의 두 가지 독음스타일을 제공하고 있다. MathSpeak는 수식 독음에 대한 표준의 부재로 많은 독서 장애인들이 수학 및 과학 공부에 어려움을 겪는다는 사실에 착안하여 gh사가 MathSpeak™프로젝트를 통해 고안한 독음스타일로 Braille Nemeth Code[10](수식을 점자로 나타내는 표준코드)를 바탕으로 수식의 독음을 정의하고 있다. MathSpeak는 수식을 구어로 표현 할 때 생기는 모호성을 최소화 하고 수식의 명확성을 제공하기 위해 명확성을 표현하는 구문들을 사용하며 미국에서 사용되는 대부분의 수식 음성 프로그램에서 채택하고 있는 독음 스타일로 현재 DAISY/NISO 표준에도 적용되고 있다. SimpleSpeech는 현재 MathPlayer 3.0의 기본 설정 독음 스타일로 MathSpeak의 룰을 대부분 적용하고 있지만 수식을 표현하는 부분에 있어 약간 다르다.

표 1. 수식의 일반구어체와 MathSpeak Rule을 적용시킨 예

	수 식 독 음	독음에 따른 수식
일반적 구어체	x equals a over b plus 1	$x = \frac{a}{b+1}$ 또는 $x = \frac{a}{b} + 1$ (모호성 발생)
Math Speak	x equals BEGIN FRACTION a OVER b END FRACTION plus1.	$x = \frac{a}{b} + 1$

표 2. 수식 $2^{x^n+1} - 1$ 을 MathSpeak Rule과 SimpleSpeech Rule을 적용시킨 예

수 식	MathSpeak Rule	SimpleSpeech Rule
$2^{x^n+1} - 1$	2 superscript x super superscript n superscript plus 1 baseline minus 1	2 super x to the n plus 1 end super minus 1

대학 교과서 콘텐츠를 위한 사용성 중심의 UI 설계 방안

손원성*, 임순범**, 우성호***, 김재경****

*경인교육대학교 컴퓨터교육과

**숙명여자대학교 멀티미디어학과

***숙명여자대학교 환경디자인과

****경인교육대학교 스마트교육 연구소

e-mail : sohnws@ginue.ac.kr

Usability-Centered UI Design Plan For University Textbook Contents

Won-Sung Sohn*, Soon-Bum Lim**, Sung-Ho Woo***, Jae-Kyung Kim****

*Dept of Computer Education, Gyeongin National University of Education

**Dept of Multimedia Science, Sookmyung Women's University

***Dept of Environmental Design, Sookmyung Women's University

****Smart Education Research Center, Gyeongin National University of Education

요 약

과거에는 기능적 중심의 어플리케이션 개발이 주로 이루어져 왔으나 최근에는 사용성을 중요시하는 사용자 경험 중심의 설계 기법으로 전환하고 있다. 특히 스마트 기기의 활성화와 함께 보다 편리한 사용자 인터페이스를 추구하는 사용자들의 성향에 따라 이러한 사용성의 가치는 더욱 중요해지고 있다. 그러나 대학 교육 환경에서는 이러한 기술적 및 디자인적 요소를 체계적으로 적용하지 못하고 있으며, 각 전공 분야에 따라 개별적으로 사용성에 대한 방법론을 연구 및 교육하고 있는 실정이다. 어플리케이션의 사용성은 공학, 심리학, 미술학등 다양한 분야의 요소가 융합되어 완성되는 것으로 어느 한 전공 분야에 국한되어 설계될 수 있는 것이 아니다. 이에 본 연구에서는 융합적 관점에서의 사용자 경험을 기반으로 하는 소프트웨어 및 콘텐츠 사용자 인터페이스를 설계하기 위한 과목을 개발하였다. 본 과목을 통하여 IT 관련학과에서는 산업적 요구를 충분히 반영할 수 있도록 하였으며, 교육대학 환경에서는 기획 및 문제해결 능력 배양을 도모하였다.

1. 서론

최근 IT 산업의 중심이 하드웨어에서 소프트웨어·콘텐츠로 변화함에 따라 UI(User Interface, 사용자 인터페이스) 설계의 중요성이 대두되고 있다 [1,4,5]. 이는 제품의 성능보다는 사용성이 제품의 가치를 결정하는 기준이 되기 때문이다

특히, 스마트 모바일 환경의 발전에 따라 UI 설계 [2]의 중요성이 더욱 부각되고 있으며, 애플의 아이폰, 아이패드 등과 같은 모바일UI[3]에 대한 관심이 증대되고 있으나 전문 인력 및 노하우가 선진국에 비하여 크게 부족한 실정이다[6].

따라서 본 연구에서는 아래와 같은 UX 기반 UI 설계를 위한 교육 과정을 개발하고자 한다. 먼저 본 연구에서는 산업계에서 중요시하는 UX 기반 UI 설계 기법을 반영한 교과목 개발하고자 하며, 특히 문제 해결 능력, 제품 기획력, 설계 능력을 갖춘 인력 양성을 위한 교과목 개발을 중점적으로 고려한다.

또한 기존 유사 과목과는 차별적인 전략을 토대로 한다. 현재 산업 공학 및 일부 IT 학과에서 교육하는 HCI 과목은 인간과 컴퓨터 간의 상호작용 기법을 중심으로 하며, 입출력 도구, 인지 과학 및 가상현실 기술까지 포함 (IT 관련 학과에서는 기술 중심으로, 인문학과에서는 인지적 이론 중심)하고 있다.

디자인 관련 학과의 인터페이스 설계 과목에서는

본 연구는 한국연구재단의 2010년도 '대학교육과정개발연구지원사업'의 지원으로 수행되었습니다.

표 1. 강의의 구성

주	이론 강의(1시간)	프로젝트 실습(2시간)	과제
1	강좌소개	<UX 사례 소개>	
2	UX의 기본 개념	<UI 사례 조사 및 분석>	UI 사례 조사 및 분석하기
3	UX 디자인과 사용성	<UI 사례 발표 및 팀구성>	
4	사용자중심 디자인	<UX 기반 UI 요구분석>	프로젝트 제안서 작성
5	참여 디자인	<제안서 발표>	
6	시나리오기반 디자인	<작업분석 기법 실습>	작업분석 및 시나리오 설계
7	디자인 스케치 기법	<작업분석 발표>	
8	HCI 역사	<비디오 프로토타입 작성방법>	비디오 프로토타입 작성
9	인간의 인지과정	<비디오 프로토타입 발표>	
10	상호작용	<Low-fi 프로토타입 작성기법>	Low-fi 프로토타입 작성
11	비주얼 디자인 요소	<Low-fi 프로토타입 발표>	
12	휴리스틱 평가 기법	<Interactive 프로토타입 작성>	Interactive 프로토타입 작성
13	사용성 평가 기법	<Interactive 프로토타입 발표>	
14	데이터 수집 기법	<사용성 평가 기법 실습>	사용성 평가 및 최종발표 준비
15	데이터 분석 기법	<최종발표>	

심미적 요소를 중시한 레이아웃 디자인을 중심으로 교육하고 있으며, 제품의 기능적 설계보다는 그래픽 요소의 모양 및 배치, 색상의 선택 등 외형적 디자인에 집중하고 있는 상황이다.

따라서 본 교과목에서는 HCI 분야의 UI 설계에 대한 기본 이론을 바탕으로 하고, 그래픽 디자인과 관련한 심미적 요소를 반영하여, 산업계의 요구사항인 UX를 향상시킬 수 있는 UI 설계 과정을 제안한다.

2. 사용성 중심의 교과목 UI 설계 방안

본 연구에서는 UX 기반 사용자 인터페이스 설계 교과목을 개발하고 실제 현장에 적용하도록 한다.

특히 급변하는 인터넷, 모바일 환경에 대응할 수 있는 IT(소프트웨어, 콘텐츠, 멀티미디어 등), 디자인, 산업 공학 등이 융합된 신규 교과목 개발하는 것이 주요 목표이다. 세부적으로는 문제 해결 능력, 제품 기획력, 설계 능력을 갖춘 인력 양성을 위한 교과목을 개발하고자 한다.



그림 1. UX 기반 UX 설계 교과목의 주요 방법론

본 교육과정 개발을 위하여 기반이 되는 핵심 이론

및 방법론을 선별하여 적용하였으며 이는 다음 그림 1과 같이 전체 5가지의 항목으로 구분된다.

본 과목은 이론과 프로젝트로 구성된 프로젝트 학습 모형과 유사한 혼합 모형을 적용하였다(표 1). 이는 이론 수업을 즉시 프로젝트 기반의 실습과 연계하여 학습의 이해도를 증진하기 위함이다.

3. 적용 결과

본 과목은 숙명여자대학교 멀티미디어학과 및 경인교육대학교 컴퓨터교육과에 공통적으로 적용하였다. 숙명여자대학의 경우 2011년 1학기, 학부 2학년 대상으로 40명, 2강좌에 적용하였으며, 경인교육대학교는 2011년 2학기, 학부 3학년 25명, 2강좌에 시범 적용하였다.

3.1 설문을 통한 수업 분석

2011년 2학기 현재 경인교육대학교 2강좌에서 도출된 수업에 대한 정성적 의견은 표 2와 같다. 설문을 통한 수업 분석을 수행한 결과 대체적으로 수업 설계의 목표를 달성한 것으로 판단되었다. 다만 이론 강의의 내용이 다양한 학문적 배경을 포함하기 때문에 상세하고도 간결한 강의 진행이 요구된다.

또한 수업 프로젝트 설계시 범위를 축소화하여 학생들에게 보다 구체적인 결과물을 도출할 수 있도록

키네틱 타이포그래피의 활용을 위한 메신저 용 모션 라이브러리 설계

곽주은[○] 임순범

숙명여자대학교

kwakje88@gmail.com, sblim@sm.ac.kr

Design of Motion Library for using the Kinetic Typography in Messengers

Joo-Eun Kwak[○], Soon-Bum Lim

Sookmyung Women's University

요 약

기존의 메신저는 정적인 텍스트로 인해 표현의 한계를 지니고 있는 반면, 본 논문에서는 키네틱 타이포그래피를 활용하여 의사전달의 표현력을 보완하고자 하였다. 메신저에서 키네틱 타이포그래피를 효율적으로 사용하기 위하여 모션 라이브러리를 체계적으로 설계하였다. 키네틱 타이포그래피 모션의 구성요소를 정의하여 이에 따라 설계된 모션 템플릿을 선호도 조사를 통하여 선정된 모션으로 라이브러리를 구성하였다. 모션 라이브러리에서 모션 선택 후 해당 모션의 속성을 설정할 수 있도록 HTML5로 인터페이스를 구현하였다.

1. 서 론

최근 다양하게 출시되는 메신저들은 정적인 텍스트와 함께 다양한 이모티콘을 제공하고 있다. 이는 커뮤니케이션 시 즐거움과 재미요소를 함께 제공하기 위한 것이다 [1]. 그러나 이모티콘 역시 의사전달과 감성전달에 한계를 갖기 때문에, 본 논문에서는 텍스트에 움직임 변화를 주는 키네틱 타이포그래피를 활용하여 메시지의 내용 표현을 보다 효과적으로 전달하고자 하였다. 본 논문에서의 키네틱 타이포그래피는 기존의 정적인 텍스트에 움직임을 부여함으로써 일정한 모션을 갖는 텍스트를 뜻한다.

키네틱 타이포그래피의 수많은 모션들 중에서도 메신저 환경을 고려하여 대표적인 모션들을 라이브러리에 선택하였다. 모션 라이브러리에서 제공하는 기본 모션들은 그대로 사용 가능하나, 파라미터 설정을 통해 빠른 시간 내에 누구나 새로운 키네틱 타이포그래피를 만들 수도 있도록 하였다.

2. 관련 연구

2.1 키네틱 타이포그래피 동향

2.1.1 애니메이션 제작 도구

현재까지 키네틱 타이포그래피를 제작하기 위해서는 Adobe사의 Flash와 After Effect, Apple사의 Motion 등 전문적인 소프트웨어 도구를 사용하고 있다. 그러나 이러한 도구들의 사용으로 제작되는 키네틱 타이포그래피는

활용 목적이 제한적이고, 일반 사용자들에게는 사용 방법이 다소 어렵고 전문적이라는 단점을 지닌다.

2.1.2 기술 동향

키네틱 타이포그래피의 기술연구는 MIT와 CMU 등 국외에서 활발하게 진행되고 있다. 이들의 내용은 키네틱 타이포그래피의 제작을 위한 엔진 구현 연구[2]와 시간을 기반으로 한 타이포그래피의 제작 기술 연구[3] 등이 있다. 이러한 연구는 키네틱 타이포그래피가 다양한 응용 서비스 시스템에서 사용될 수 있고, 이는 인간의 표현력이 보다 풍부해질 수 있음을 보여주고 있다.

3. 연구 목표

본 연구에서는 메신저에서 효과적으로 사용할 수 있는 모션 라이브러리 설계를 목표로 하였다. 사용자가 메신저 채팅 시 대화 내용을 입력하고, 모션 라이브러리에서 모션 선택과 해당 모션의 파라미터 설정을 통해 키네틱 타이포그래피를 적용할 수 있도록 하였다.

4. 설계 및 구현

4.1 전체 시스템 구성도

[그림 1]은 키네틱 타이포그래피에를 적용한 메신저의 전체 시스템 구성도이다. HTML5로 구현된 메신저 인터페이스에서 입력된 텍스트에 모션과 파라미터를 설정하면 그 결과가 SVG로 변환되어 화면에 디스플레이 되고 상대방에 SVG 파일로 전송이 된다[4].

* 본 연구는 지식경제부 산업원천기술개발사업의 과제번호 10041794, 소셜미디어 환경에서 감성전달을 위한 키네틱 타이포그래피 기술 개발 과제의 연구결과로 수행되었음.



[그림 1] 시스템 구성도

4.2 모션 라이브러리 설계

메신저에서의 키네틱 타이포그래피 제작을 위한 모션 라이브러리 설계는 크게 두 단계로 진행하였다. 첫째, 라이브러리 내 모션과 파라미터 속성들을 정의하는 키네틱 타이포그래피 모션 개념 설계 단계와 둘째, 이를 토대로 한 모션 라이브러리 설계이다.

4.2.1 키네틱 타이포그래피의 모션 개념 설계

키네틱 타이포그래피의 구성 요소는 모션과 모션에 따른 속성으로 나뉜다. <표 1>은 이를 설명한다.

모션은 키네틱 타이포그래피의 이동 방법과 방향에 따라 기본 모션과 세부 모션으로 나뉜다. 기본 모션은 제자리, 직선, 곡선의 움직임으로 구분되고, 세부 모션은 기본 모션에서의 방향에 따라 다시 여러 가지로 구분된다.

모션에 따른 속성은 키네틱 타이포그래피의 움직임에 영향을 주는 속성들로 이루어진다. 본 논문에서는 텍스트 움직임의 속도, 위치, 크기, 투명도, 글자분리, 타이포그래피의 기본 요소로 구분 하였다.

<표 1> 키네틱 타이포그래피의 구성 요소

기본 Motion	세부 Motion	Motion에 따른 속성
이동 방법에 따라	이동 방향에 따라	이동 동작(움직임)에 영향을 주는 속성
제자리 움직임	제자리	속도 / 위치 / 크기 / 투명도 / 글자분리 / 타이포그래피 기본 요소
직선 움직임	상/하/좌/우/앞/뒤	
곡선 움직임	포물선/물결/마운드/동심원	

4.2.2 모션 라이브러리 설계

모션 라이브러리 설계 단계에서는 본 논문에서 정의한 키네틱 타이포그래피 모션 개념 설계를 바탕으로 메신저에서 사용될 모션 30가지를 우선 제작하였다. 이들 모션 중에서 실제 타겟 사용자들을 대상으로 한 메신저에서 사용될 가능성과 선호도 평가에 따라 빈도가 높은 모션들을 중심으로 그룹핑하여 8가지로 나누었다. [그림 2]는 그룹핑 한 대표 모션 8가지의 모습과 설명이다.

커졌다 작아지는 반복 모션	나타났다 사라지는 반복 모션	글자 날개가 순서대로 커졌다 작아지는 모션	점점 커진 후 원래 크기로 다시 나타남
모션 라이브러리	모션 라이브러리	모션 라이브러리	모션 라이브러리
좌우로 빠르게 떨림	아래로 내려오며 서서히 사라지는 모션	상하로 빠르게 떨리는 모션	포물선을 그리며 글자 이동하는 모션
모션 라이브러리	모션 라이브러리	모션 라이브러리	모션 라이브러리

[그림 2] 사용 가능성과 선호도에 따른 8가지 대표 모션

4.2.3 키네틱 타이포그래피의 모션과 속성의 설정 방법

본 논문에서의 메신저 용 키네틱 타이포그래피 제작은 <표 1>에서처럼 모션 선택과 속성 설정으로 이루어진다. [그림 3]는 대표 모션 중 1번 모션(크기반복)을 대상으로 세부 모션과 속성의 설정 과정을 설명하고 있다. 직선의 기본 모션에서 앞과 뒤로 움직이는 세부 모션의 선택과 속도, 크기, 투명도 등 속성을 설정하여 모션을 완성되었다.



[그림 3] 메신저 용 키네틱 타이포그래피의 모션 제작과정

5. 인터페이스 구현 결과

메신저에서의 키네틱 타이포그래피 제작은 대화 내용을 입력하며 모션을 확인하고 전송이 가능하다. [그림 4]는 본 논문에서 제시하는 메신저 용 모션 라이브러리 인터페이스 구현 결과이다. 글자 입력 후 모션을 선택하고, 속성을 설정함에 따라 키네틱 타이포그래피가 완성된다.



[그림 4] 모션 라이브러리 인터페이스 제공 모습

6. 결론

본 논문에서는 메신저에서 키네틱 타이포그래피로 활용될 모션 개념을 정의하고, 이를 바탕으로 사용자 평가를 통해 모션 라이브러리를 설계하였고, 인터페이스를 구현하였다. 이를 통해 메신저에서 효과적으로 사용될 키네틱 타이포그래피 모션을 선정하였다. 이는 모션 라이브러리를 활용함으로써 메신저 환경에서 키네틱 타이포그래피 제작을 더욱 쉽고 간편하게 할 수 있을 것이라 기대한다.

7. 참고 자료

- [1] 연합뉴스, “카카오톡, ‘말하는 이모티콘’ 출시”, <http://news.naver.com>, 2012
- [2] Zhizuan Yeo, and Scott. E. Hudson, “KTE2: An Engine for Kinetic Typography”, ACM CHI 09, 2009
- [3] S.Ford, J.Forlizzi, S.Ishizaki, “Kinetic Typogeaphy: Issues in time-based presentation of text”, ACM CHI 97, 1997
- [4] W3C, HTML5, <http://www.w3.org/TR/html5/>

UX 기반 소프트웨어 UI 설계를 위한 교과목 개발

손원성¹⁾, 임순범^{2)O}, 우성호²⁾

경인교육대학교¹⁾, 숙명여자대학교²⁾

sohnws@ginue.ac.kr, sblim@sookmyung.ac.kr, shwoo@sookmyung.ac.kr

Curriculum development of UX based UI design for software

Won-Sung Sohn, Soon-Bum Lim, Sungho Woo

Dept. of Computer Education, Gyeongin National University of Education

Dept. of Multimedia Science, Sookmyung Woman's University

Dept. of Architecture Design, Sookmyung Woman's University

요 약

최근 대다수의 소프트웨어 및 콘텐츠 업계에서는 제품의 성능보다는 사용성을 중요시하여 사용자 경험 중심의 설계 기법으로 전환하고 있다. 한편 대학에서는 이러한 변화를 적극적으로 반영하지 못하고 있으며, 각 학문분야에 국한된 인터페이스 설계, 시각 디자인, 산업공학 방법론을 개별적으로 교육하고 있는 실정이다. 이에 본 연구에서는 융합적 관점에서의 사용자 경험을 기반으로 하는 소프트웨어 및 콘텐츠 사용자 인터페이스를 설계하기 위한 과목을 개발하였다. 본 과목을 통하여 IT 관련학과에서는 산업적 요구를 충분히 반영할 수 있도록 하였으며, 교육대학 환경에서는 기획 및 문제해결 능력 배양을 도모하였다.

1. 서 론

최근 IT 산업의 중심이 하드웨어에서 소프트웨어, 그리고 콘텐츠로 옮겨감에 따라 UI(User Interface) 설계의 중요성이 대두되고 있다[1,2]. 이제는 제품의 외형적 성능보다는 사용성이 제품의 가치를 결정하는 기준이 되기 때문이다. 이러한 사용자 중심의 체계적인 사용자 인터페이스 설계는 개발 기간은 물론이고 유지보수 비용을 단축시켜 궁극적으로 제품 및 서비스의 경쟁력을 확보할 수 있다는 인식이 보편화되고 있다.

그러나 대학의 현행 교육 과정은 위와 같이 급변하는 시장 트렌드에 부응하지 못하는 실정이다. 아이폰, 안드로이드 등 스마트폰 시장이 활성화되고 모바일 단말기 기반 어플리케이션 및 콘텐츠의 UI/UX(User eXperience) [3,4] 인력에 대한 수요가 급증하고 있으나 대학 정규 교육 과정에 이것을 적절히 반영하지 못하고 있다.

특히 소프트웨어 및 콘텐츠에 대한 UI 설계 교과목이 절대 부족한 실정이다. 현재 미술/디자인 전공 분야에서는 제품 디자인, 웹 기획, 인터페이스 디자인 등의 유사 과목이 교육되고 있으나, 컴퓨터-정보공학으로 대표되는 IT·SW 분야와의 융합이 제대로 이루어지지 않고 있다. 오히려 IT 전공 분야에서는 개발자 중심의 UI 설계 방법론, 사용자 인터페이스 개론 수준의 과목이 교육되고 있다. 따라서 본 연구에서는 아래와 같은 UX 기반 UI 설계를 위한 교육 과정을 개발하고자 하였다.

2. UX 기반 UI 설계 교과목 개발 방안

본 연구는 한국연구재단의 2010년도 대학교육과정개발 연구지원사업의 지원을 받아 수행되었음

본 연구에서는 UX 기반 사용자 인터페이스 설계 교과목을 개발하고 실제 현장에 적용하도록 한다. 특히 급변하는 인터넷, 모바일 환경에 대응할 수 있는 IT(소프트웨어, 콘텐츠, 멀티미디어 등), 디자인, 산업 공학 등이 융합된 신규 교과목 개발하는 것이 주요 목표이다. 세부적으로는 문제 해결 능력, 제품 기획력, 설계 능력을 갖춘 인력 양성을 위한 교과목을 개발하고자 한다. 본 연구에서 개발한 'UX 기반 소프트웨어 및 콘텐츠 UI 설계' 과목에 대한 개요는 다음 <표 1>과 같다.

<표 1> 강좌의 개요

구분	교과 내용
과목명	UX 기반 소프트웨어 및 콘텐츠 UI 설계
학점/시간	3학점/45시간(주당 3시간, 15주)
교과목개요	- 최근 중요시되고 있는 UX(사용자 경험)를 기반으로 하는 사용자 인터페이스 설계 방법을 배운다. - 사용자 인터페이스 설계는 소프트웨어 및 멀티미디어 콘텐츠 개발하는 첫 단계로서 특히 웹 환경이나 모바일 환경에서는 더욱 중요시 되고 있다. - 사용자 특성에 대한 이해에서 시작하여 사용자 인터페이스의 설계의 순환 단계를 배우고 실습을 통해 모바일 단말기 및 서비스 UI를 설계하고 개발할 수 있는 실천 능력을 습득한다.
교육목표	- 사용자 인터페이스의 분석, 설계, 프로토타입 개발, 평가 기법까지의 일련의 순환 단계를 배우고, 이러한 기법을 팀 프로젝트를 통해 UI설계 능력을 배양한다. - o UX에 기반한 사용자의 요구 사항 및 작업 분석 - o UI 프로토타입 설계 및 개발 기법 - o UI 설계에 대한 사용성 평가 기법 - o UI 설계에 영향을 주는 UX 및 인지적 요소

3. 적용 결과

본 과목은 숙명여자대학교 멀티미디어학과 및 경인교육대학교 컴퓨터교육과에 공통적으로 적용하였다. 숙명여자대학교의 경우 2011년 1학기, 학부 2학년 대상으로 40명, 2강좌에 적용하였으며, 경인교육대학교는 2011년 2학기, 학부 3학년 25명, 2강좌에 시범 적용하였다.

3.1 설문을 통한 수업 분석

2011년 2학기 현재 경인교육대학교 2강좌에서 도출된 수업에 대한 정성적 의견은 <표 2>와 같다. 설문을 통한 수업 분석을 수행한 결과 대체적으로 수업 설계의 목표를 달성한 것으로 판단되었다. 다만 이론 강의의 내용이 매우 생소한 점과 대부분의 용어가 다양한 학문적 배경을 포함하기 때문에 상세하고도 간결한 강의 진행이 요구된다 또한 수업 프로젝트 설계시 범위를 축소화하여 학생들에게 보다 구체적인 결과물을 도출할 수 있도록 유도할 필요가 있다. 특히 최근 추세인 모바일 앱과 관련된 내용 스마트 교과서 설계 및 분석 등에 대하여 관심을 갖는 것으로 나타났다.

<표 2> 수업 분석 결과

의견	세부 내용
기획능력 배양에 대한 긍정적 의견	- 콘텐츠 기획에 대한 새로운 방법론 습득 - S/W 기획 및 설계에 대한 자신감 - 교육현장에서 적용가능한 방법론
문제 분석 및 해결에 대한 의견	- 문제 분석에 대한 다양한 방법론 습득 - 문제 해결을 위한 구체적 공동작업 방법론 학습 - 디자인 기법을 통한 문제해결 방법 능력 배양
프로토타입에 대한 의견	- 프로토타입을 통한 구체적 설계 가능 - 프로그래밍 보다 유연한 방법론 - 코딩 능력이 부족해도 구체적인 결과물 작성이 가능 - 컴퓨터교육에 대한 자신감 고취
User Research 방법론에 대한 의견	- Insight 분석기법을 통한 사용자 내면의 사고방식 도출 가능 - 수업설계 및 분석에 적용가능 - 정보설계, 콘텐츠 설계와 관련된 핵심 방법론
수업 방식에 대한 긍정적 의견	- 수업 초반에 어려웠던 내용들을 실습을 통하여 매우 구체적으로 습득한 점이 매우 인상적임 - 이론 학습과 프로젝트 학습을 병행하여 학습효과를 유지할 수 있음 - 이론 강의의 양이 적절하였음 - 가정학습주년을 적절히 활용한 수업 방식이 효과적이었음
수업 방식에 대한 개선 의견	- 프로젝트의 범위를 구체적으로 세분화하여 제시하여야 함 - 과제의 양이 부담스러움 - 최종 프로젝트의 수준을 정해주셨으면 좋겠음 - 모바일, 앱, 디지털교과서, 교과수업등에 적용 가능한 프로젝트 주제가 필요함

3.2 수업 프로젝트 소개

2011년 2학기 본 과목에서 도출된 프로젝트 중 요약된 내용은 다음과 같다.

<표 3> 프로젝트 리스트

프로젝트 내용
- 농협 ATM 개선 방안/Beyond 위키피디아
- 교보문고 e-Book 인터페이스 개선
- 포토샵 CS2 UI 개선방안
- OK Cashbag 개선방안/아프리카 TV 버전 업 프로젝트
- 네이버 웹툰 개선 프로젝트
- 우리의 지누앱: 경인교대 앱 개선 전략
- Daum 검색엔진 개선 방안
- 스크래치 변화의 날개를 달다: 스크래치 UI 개선 방안
- 경인교대 수강신청 프로그램 개선방안
- 뉴런뉴런 새로운 해답을 찾다
- 안드로이드 사용자를 위한 페이스북 UI 개선



<그림 2> 페이스북 UI 개선 프로젝트 화면

4. 결론

최근 산업계 및 IT 분야에 불고 있는 UX 열풍은 순간의 유행이 아닌, 새로운 설계 패러다임에 가깝다. 본 연구에서 제안한 ‘UX 기반의 UI 설계’ 과목을 통하여 학생들은 다양한 사용자 경험을 도출하기 위한 사회학적 분석 기법을 익히고 이것을 설계 및 개발 과정에 적용할 수 있는 능력을 습득할 수 있었다. 향후 본 연구팀이 수행한 경험을 바탕으로 보다 개선된 교과목을 개발할 필요가 있으며, 이는 일반 IT 계열 학과뿐만 아니라, 교육대학교 현장에서 매우 효과적으로 활용될 수 있음을 확인하였다.

5. 참고문헌

- [1] Helen Sharp, Yvonne Rogers and Jenny Preece, Interaction Design: beyond human-computer interaction, 2nd edition, John Wiley & Sons, 2007
- [2] Ben Shneiderman and Catherine Plaisant, Designing The User Interface: Strategies for Effective Human-Computer Interaction, 5th edition, Addison-Wesley, 2009
- [3] Russ Unger and Carolyn Chandler, A Project Guide to UX Design: For User Experience Designers in the Field or in the Making, New Riders, 2009
- [4] Hokyung Ryu, Mobile User Interface Analysis and Design: A Practitioner's Guide to Designing User Interfaces for Mobile Devices, NOVA, 2009

CSS3를 활용한 웹 콘텐츠의 3D 표현 기법 연구

이희진[○] 심영빈 임순범 임현정 정원식숙명여자대학교, 숙명여자대학교, 숙명여자대학교, 한국전자통신연구원, 한국전자통신연구원
wmfvv@hanmail.net, rindasim@naver.com, sblim@sookmyung.ac.kr, hjyim@etri.re.kr, wscheong@etri.re.kr

3D Representation Techniques of Web Content by Applying CSS3

Hee-Jin Lee[○] Yeong-Bin Shim Soon-Bum Lim Hyun-Jeong Yim Won-Sik Cheong

SookMyung Univ., SookMyung Univ., SookMyung Univ., ETRI, ETRI

요 약

최근 웹 콘텐츠 관련 3D 표현기술에 대한 연구는 다양한 분야에서 요구된다. 본 연구는 HTML5로 구성된 웹페이지를 CSS 스타일시트를 통해 3D 공간상에 재구성하여 표현하는 것을 목표로 한다. 차세대 웹 언어인 HTML5로 구성된 웹페이지를 요소별, 의미별로 분리하고 CSS 스타일시트를 적용하여 3차원 공간상에 배치할 수 있는 효과적인 3D 콘텐츠 표현 기법을 연구하였다. 제안된 스타일시트의 적용방법은 스포츠 중계, 일기예보, 다큐멘터리, 독립형 데이터 서비스 등의 다양한 분야에 3D 콘텐츠를 구성하는데 활용이 가능하다.

1. 서 론

최근 웹 3D 관련 연구는 활발히 진행되고 있으나 2차원의 웹 콘텐츠를 3D로 구성하는 연구는 다양한 분야에서 필요로 하고 있으나 활발히 진행되지 않고 있다. 이에 따라 본 연구는 웹 환경에서 차세대 마크업 언어인 HTML5(HyperText Markup Language)로 표현되어 있는 웹페이지에 CSS(Cascading Style Sheet) 스타일시트를 적용하여 웹 콘텐츠를 3D 공간에 재구성한다.

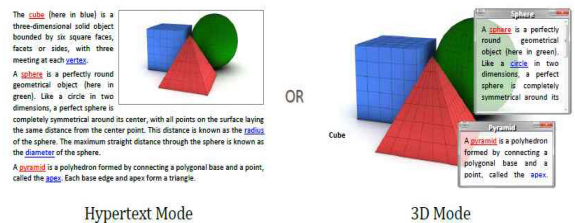
2. 3D 콘텐츠 서비스를 위한 관련 기술

2.1. HTML5를 위한 스타일시트

차세대 웹 표준인 HTML5[1]는 웹페이지의 구조를 정의하며 Javascript와 같은 스크립트, XML, DOM과 같은 데이터 모델, 레이아웃을 구성하는 스타일시트 CSS[2]로 이루어져 있다. 기존의 HTML은 웹 문서의 다양성 제공 및 변경에 제약이 존재하였으며, 이를 보완하기 위해 만들어진 것이 레이아웃과 스타일을 자유롭게 설정할 수 있는 스타일시트이며 CSS3 스타일시트로 보강되었다.

2.2. 관련 연구

웹 브라우저에서 3D 콘텐츠를 표현하는 연구로는 3D 콘텐츠로의 접근을 위한 'Dual-Mode 사용자 인터페이스[3]'와 '3D 위키 엔진[4]'이 있다. [그림 1]은 3D 콘텐츠로의 접근을 위한 'Dual-Mode 사용자 인터페이스' 연구에 소개된 하이퍼텍스트와 3D 데이터 정보를 통합하여 제공하는 하이퍼텍스트 모드와 3D 모드이다. 하이퍼텍스트와 3D 데이터 정보로 구성된 2차원 화면을 3차원 공간에 3D 모드로 재구성하여 보다 효과적인 정보전달을 가능하게 한다.



[그림 1] 하이퍼텍스트와 3D 데이터 정보의 통합

'3D 위키'는 [그림 2]와 같이 웹 페이지에서의 그림을 3차원 배경으로 구성하고 그 앞에 텍스트로 내용을 보여준다. 사용자가 관찰 시점을 자유롭게 설정할 수 있기 때문에 해당 3D 콘텐츠를 다양한 각도에서 탐색할 수 있다.



[그림 2] 3D 위키 엔진

3. 연구 목표

본 연구는 HTML5로 구성된 웹페이지를 CSS 스타일시트를 적용하여 3D 공간상에 재구성하여 표현하는 것을 목표로 한다. 단계별로 2D로 작성되어 있는 웹 콘텐츠를 3D로 표현할 수 있도록 관련 CSS3 모듈의 기능과 속성들을 적절히 활용하여 스타일시트를 정의한다. 그리고 제안된 스타일시트

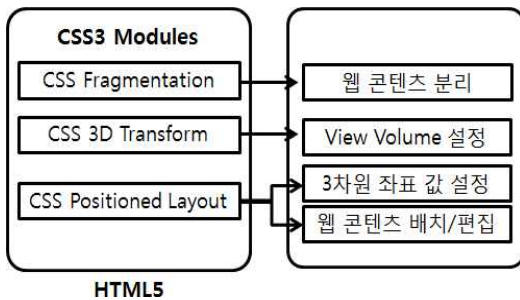
※ 본 연구는 방송통신위원회의 지상파 양안식 3DTV 방송시스템 기술 개발 및 표준화 사업의 연구결과로 수행되었음(KCA-2012-10912-02001).

를 확인하기 위해 요소별, 의미별로 웹페이지를 분리하고 3차원 공간에 배치하는 샘플 콘텐츠를 구현해 본다.

4. 웹 콘텐츠의 3D 표현 과정

4.1. 3D 표현을 위한 CSS3 관련 모듈

웹페이지를 요소별, 의미별로 분리하고 3차원 공간을 설정하고 배치하는 데 필요한 CSS3의 모듈을 탐색하고 이를 적절히 활용하여 스타일시트를 정의한다. CSS Fragmentation 모듈을 통해 웹페이지를 요소별, 의미별로 분리하고 CSS Positioned Layout 모듈로 3차원 좌표 값을 설정하여 3차원 공간상에 웹 콘텐츠를 배치하며 CSS 3D Transform 모듈을 사용하여 뷰 볼륨을 설정한다.

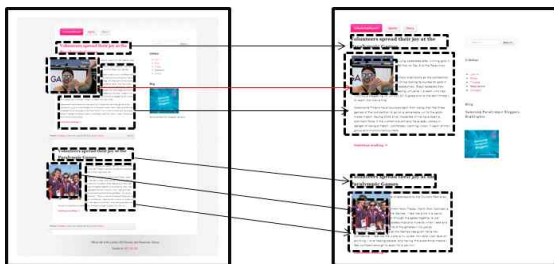


HTML5

[그림 3] 3D 표현을 위한 모듈 정의

4.2. 웹 페이지의 콘텐츠 분리

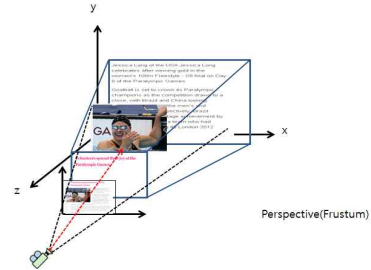
텍스트, 이미지, 동영상 등의 다양한 콘텐츠로 구성된 웹 페이지의 효과적인 재구성을 위해 CSS3 Fragmentation 모듈을 사용하여 요소별, 의미별로 분리한다. HTML5로 구성된 웹페이지의 구역을 나누는 'break-before', 'break-after', 'break-inside' 속성을 사용하여 [그림 4]와 같이 제목, 이미지, 본문 내용 등을 원하는 사이즈로 분리한다.



[그림 4] 웹페이지의 요소별, 의미별 분리

4.3. 3D 뷰 볼륨 설정

CSS3 3D transform 모듈의 perspective 속성을 통해 뷰 볼륨을 설정한다. [그림 5]와 같이 공간 내에 분리된 웹 콘텐츠를 배치하고 2차원 화면에 투영한다. 또한 3D transform의 perspective-origin 속성을 통해 관찰자의 시선 방향과 동일하게 투영 방향을 설정할 수 있다.



[그림 5] 3D 뷰 볼륨 설정[5]

4.4. 3D 공간에 콘텐츠 배치

Positioned Layout 모듈을 사용하여 요소별, 의미별로 분리한 웹 콘텐츠들의 3차원 좌표 값을 설정하고 3차원 공간상에 배치한다. [그림 6]과 같이 3D 공간에서 배치된 웹 콘텐츠를 왼쪽과 오른쪽으로 회전하여 결과를 확인하였다.



[그림 6] 3D 공간에 배치한 결과의 확인

5. 결론

본 연구는 요소별, 의미별로 웹 페이지를 분리하여 공간에 배치할 수 있는 효과적인 3D 콘텐츠 표현 기법을 제공하였다. 본 논문에서 제안한 방법으로 정의한 스타일시트를 기반으로 생성된 3D 콘텐츠는 3차원 공간상에 배치할 수 있다. 제안된 스타일시트의 적용방법은 추후 텍스트, 비디오, 애니메이션 등과 같은 다양한 요소로 구성된 웹 콘텐츠를 3D 공간에 표현 하는데 적용할 수 있으며 스포츠 중계, 일기예보, 다큐멘터리, 독립형 데이터 서비스 등의 다양한 분야에 3D 콘텐츠를 구성하는데 활용 가능하다.

참고문헌

- [1] W3C, "HTML5", <http://ko.wikipedia.org/wiki/HTML5>, 2012.
- [2] W3C, "Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification", <http://dev.w3.org/html5/spec/single-page.html>, 2011.
- [3] Jacek Jankowski and Stefan Decker, "A Dual-Mode User Interface for Accessing 3D Content on the World Wide Web", WWW 2012-Session, User Interfaces and Human Factor, 2012.
- [4] Jacek Jankowski, Marek Jozwicz, Yolanda Cobos, Bill McDaniel, Stefan Decker, "3DWiki: The 3D Wiki Engine", ACM, 2009.
- [5] 최윤철, 임순범, "컴퓨터그래픽스 배움터", pp.220-221, 생능출판사, 2006.

인스턴트 메신저에서 키네틱 타이포그래피의 모션 라이브러리 및 인터페이스 설계

곽주은*, 임순범*
*숙명여자대학교 멀티미디어학과
e-mail : kwakje88@gmail.com

Development of Motion Library and Interface for using the Kinetic Typography in Instant Messengers

Joo-Eun Kwak*, Soon-Bum Lim*
*Dept of Multimedia Science, Sookmyung Women's University

1. 연구 필요성 및 문제점

기존에 출시된 메신저들은 텍스트가 정적으로써, 의사전달과 감성전달에 한계점을 지닌다. 이에, 본 논문에서는 글자가 동적으로 움직이는 키네틱 타이포그래피를 활용하여 이러한 문제점을 해결하고자 한다. 그러나 키네틱 타이포그래피의 모션을 어플리케이션에서 표현하기 위해선 전문적인 저작도구를 필요로 하기 때문에[1], 인스턴트 메신저 사용자가 키네틱 타이포그래피를 쉽게 표현할 수 있는 인터페이스 개발을 목표로 하였다. 이를 위해 키네틱 타이포그래피를 구성하는 요소들을 체계적으로 분류하고, 메신저에서 필요한 모션들을 그룹지어 템플릿 형태의 모션 라이브러리로 제공하였다.

2. 연구내용과 방법

키네틱 타이포그래피의 체계적인 모션 개념은 {기본모션, 세부모션, 모션의 속성}의 요소 집합으로 정리하였다. 이후, 모바일 환경에서 적용될 예상 모션들 30개를 바탕으로 사용자 선호도 조사를 통해 기본모션에 따른 모션들을 <표1>과 같이 그룹핑하였다. 예를 들어, 제자리 그룹 내 모션 속성들 중 크기변화에 따라 커졌다 작아지는 모션 혹은 텍스트가 점점 작아져 없어지는 모션을 제작할 수 있다.

<표1> 라이브러리 구성 모션들의 그룹핑

기본모션	세부모션	모션을 이루는 속성	
제자리	제자리	공동속성	시간, 글자분리
		세부속성	크기, 위치, 회전, 투명도
직선	상, 하, 좌, 우, 앞, 뒤	공동속성	시간, 글자분리, 크기, 투명도
		세부속성	이동(이동방향에 따라)
곡선	물결, 포물선, 바운스, 동심원	공동속성	시간, 글자분리, 크기, 투명도
		세부속성	패스(곡선 모양에 따라)

* 본 연구는 지식경제부 산업원천기술개발사업의 과제번호 10041794, 소셜미디어 환경에서 감성전달을 위한 키네틱 타이포그래피 기술 개발 과제의 연구결과로 수행되었음.

모션 라이브러리의 인터페이스는 사용자의 능숙도에 따라 세 단계로 구분한다. 초급자는 모션 라이브러리의 모션을 선택하여 바로 적용시키며, 중급자는 모션을 선택한 후 해당 파라미터를 변경함으로써 모션을 재편집 및 저장한다. 고급자는 모션의 선택 없이 자신만의 새로운 모션을 제작할 수 있다.

채팅창 구현은 phpwebsocket을 사용하여 HTML5 환경에서 제작하였고, 키네틱 타이포그래피의 모션은 Canvas를 사용하였다. [그림1]은 모션 라이브러리에서 모션을 선택한 후, 해당 파라미터를 변경하며 키네틱 타이포그래피를 제작하는 모습을 나타낸다.



[그림1] 모션 라이브러리와 속성값 변경 모습과 결과

3. 결론 및 향후 연구

인스턴트 메신저에서의 키네틱 타이포그래피 제작을 위한 모션 라이브러리 제공은 메신저 환경에서 누구에게나 쉽고 재미있게 키네틱 타이포그래피를 제작할 수 있도록 효과적인 도움을 줄 것이다.

참고문헌

[1] Zhiuan Yeo, Scott. E. Hudson, "KTE2: An Engine for Kinetic Typography", ACM CHI 09, 2009

N-스크린 환경에서의 웹 콘텐츠 재구성을 위한 콘텐츠 분리 방법

이희진*, 임순범*, 임현정**, 정원식**

*숙명여자대학교 멀티미디어학과

**한국전자통신연구원

e-mail : wmfvv@hanmail.net

The Method of Contents Separation for Web Contents Reconstruction in N-Screen Environment

Hee-Jin Lee*, Soon-Bum Lim*, Hyun-Jeong Yim**, Won-Sik Cheong**

*Dept of Multimedia Science, SookMyung University

**Dept of Electronics and Telecommunications Research Institute

1. 연구 필요성 및 문제점

스마트 기기의 보급 이후 N-스크린 환경에 들어서면서 서로 다른 디바이스에서도 동일한 서비스를 제공할 수 있게 되었다. 이에 따라 현재의 웹 콘텐츠들은 N-스크린 환경에서 대부분 웹 표준 기술인 HTML5로 구성되어 다양한 플랫폼에 적용할 수 있도록 서비스 된다. 따라서 본 연구에서는 웹 페이지를 서로 다른 디바이스에서 재구성할 수 있도록 보다 효율적으로 웹 콘텐츠를 분리하고 배치하는 방법을 제공하고자 한다.

2. 연구내용과 방법

본 연구는 CSS 스타일시트를 활용하여 기존의 다양한 2D 기반 웹 페이지를 구성하고 있는 콘텐츠를 재구성하기 위해 웹 콘텐츠를 분리하고 배치하는 방법을 설계한다. 기존의 웹 페이지들은 각각 다른 태그들로 구성되어 있기 때문에 사용자가 일일이 태그 별로 스타일을 지정해야 하므로 비효율적이다. 보다 효율적인 작업을 위해 CSS Fragmentation Level 3 모듈[1]에서 콘텐츠를 분리할 위치를 지정하는 break-before, break-after, break-inside 속성을 사용하여 콘텐츠를 분리하는 두 가지 방법을 제안한다.

첫째, 웹 페이지를 구성하는 다양한 멀티미디어 데이터를 정의하는 태그들을 원하는 형태로 Grouping하여 분리한다. 사용자는 멀티미디어 데이터 종류 별로 콘텐츠를 분리할 수 있으며, 원하는 레이아웃을 쉽게 구성할 수 있다. <tag name>{<break-before>|<break-after>|<break-inside>}로 분리할 위치를 지정하고 auto, always, avoid 등의 속성을 함께 설정한다. 둘째, 하나의 태그로 정의된 콘텐츠를 내부적으로 분리한다. 이 방법은 웹 페이지에서 많은 양으로 구성된 본문 내용과 같은 콘텐츠를 문단별로 쉽게 분리할 수 있다.

※ 본 연구는 방송통신위원회의 지상파 양안식 3DTV 방송시스템 기술 개발 및 표준화 사업의 연구결과로 수행되었음(KCA-2012-10912-02001).



(그림 1) 웹 페이지를 구성하는 콘텐츠 분리

다음으로 분리된 콘텐츠가 위치될 영역을 지정해 주는 'Fragmentainer'를 사용하여 Page, Column, Region 단위로 배치할 수 있다. 일반적으로 웹 페이지는 구조적인 레이아웃으로 구성되어 있기 때문에 가장 관련이 깊은 CSS Regions Level 3 모듈[2]을 사용하여 배치한다. #<id value>{<flow-into>|<flow-from>:<ident>}을 사용하여 스타일을 적용할 범위를 지정하고, #region{}으로 영역의 크기와 위치를 설정한다. 추가로 <region-overflow:<속성>;>으로 overflow된 콘텐츠를 다룬다.



(그림 2) 분리한 콘텐츠를 지정한 영역에 배치

3. 결론 및 향후 연구

본 연구는 N-스크린 환경 기반의 웹 페이지를 구성하는 콘텐츠를 원하는 형태로 분리하고 재배치하는 방법을 제안하였다. 나아가 사용자는 웹 페이지를 구성하는 콘텐츠들의 스타일을 정의하는 태그에 의미를 부여하여 사용자 설정에 따라 레이아웃을 자유롭게 구성할 수 있는 방안을 연구할 예정이다.

참고문헌

- [1] W3C, "CSS Fragmentation Module Level 3",
"http://dev.w3.org/csswg/css3-break/", 2012.
- [2] W3C, "CSS Regions Module Level 3",
"http://dev.w3.org/csswg/css3-regions/", 2012.

스마트 교육환경을 위한 e-교과서 포맷 표준화 방안 연구

임순범*, 손원성**, 김재경***

숙명여자대학교 멀티미디어학과* · 경인교육대학교 컴퓨터교육과**

경인교육대학교 스마트교육연구센터***

요 약

스마트 기기의 보급과 Apple사의 iBooks 2 보급과 더불어 교육 환경에서의 디지털 교과서의 중요성이 높아지고 있다. 국내에서도 정부는 스마트 교육 추진 계획을 최근 발표하였으며, 디지털 교과서 적용 및 클라우드 교육 시스템 구성 등의 내용을 포함하고 있다. 미래의 교육환경은 이러닝을 넘어 다양한 디바이스에서도 정확한 학습 및 소통이 가능한 스마트 교육 환경을 필요로 하고 있다. 현재 정부의 디지털교과서 사업은 중장기 목표로 추진되고 있으며, 단기적인 목표로 전자책과 유사한 형태의 모델로서 배포가 용이한 PDF 기반 e-교과서 사업을 추진한 바 있다. 그러나 차후 스마트 교육을 위해서는 다양한 종류의 단말기에서도 변형이 용이하고 원본 콘텐츠에 대한 중립성이 강화된 e-교과서 전략이 요구된다. 따라서 본 논문에서는 전자책 기반의 e-교과서 포맷에 대한 표준화 방안을 제안한다. 이를 위하여 먼저 전자책 요소 기술의 동향에 대하여 조사 분석을 하였으며, 기존 교과서의 분석을 바탕으로 서책형 교과서의 디지털화를 위한 전자책 포맷 표준화 방안을 연구하였다. 또한 제안된 표준 방안에 근거하여, 실제 e-교과서 시범 콘텐츠를 제작하였다. 그 결과 제안 기법은 향후 스마트 교육 환경에 적용 가능하며 디지털교과서 중장기 사업 목표에 효과적으로 응용 가능할 것으로 예상된다.

주제어 : 스마트 교육, e-교과서, 디지털교과서, 표준문서형식, ePub

A Study of e-Textbook Format Standardization Scheme for Smart Education Circumstance

Soon-Bum Lim*, Won-Sung Sohn** and Jae-Kyung Kim***

Sookmyung Women's University, Dept. of Multimedia Science*

Gyeong-In Educational University, Dept. of Computer Education**

Gyeong-In Educational University, Smart Education Research Center***

Abstract

The Korea government has recently announced 「A Master Plan for Smart Education」, including application of digital textbooks and composition of education system using cloud computing. Our education system in future circumstance, over the conventional e-learning methods, needs the smart education solutions which enable students to study and communicate on various types of devices. The ongoing government project related with the digital textbook has been performed as mid- and long-term goals, whereas PDF-based e-textbook project, similar to e-book model and, has been already completed for the short-term goal. For the purpose of improved future smart education circumstance, however, a specific strategy is required in the following areas: flexibility of format conversion and independency of original text sources among the multiple device platforms. Therefore, in this paper, we propose a standardization scheme for e-textbook format based on e-book structure. To do this, we survey trends in e-book technologies, and research on standardization of e-book format for digitalization of textbooks, based on the analysis of existing textbooks. Moreover, we produce an example e-book content using our proposed standard method. As a result, our approach can be applied to the future smart education circumstance, and we may say that it will be efficiently applicable to the long-term digital textbook project.

Key Words : Smart-Learning, e-textbook, digital-textbook, standard document format, ePub

논문투고: 2011-09-30

논문심사: 2011-10-05

심사완료: 2012-09-21

1. 서론

최근 교육기술과학부가 발표한 「스마트 교육 추진전략」[4]에서는 디지털교과서의 확대 적용 및 클라우드 교육 시스템 구축 등의 내용을 담은 미래 교육 전략을 포함하고 있다. 특히 시범사업으로 진행되어 왔던 디지털교과서 사업을 초·중·고 분야에 까지 확대하는 등의 미래교육과 관련된 주요한 내용을 제안하고 있다.

국내 디지털교과서 사업은 ‘플랫폼 구축 및 시범학교 적용’ 단계를 지나 다양한 형태의 사례 연구를 진행한 바 있다. 또한 2010년 발표한 ‘교과서 선진화 방안’[1]을 통하여 전자책과 유사한 형태의 모델인 e-교과서를 활용함으로써, 젊은 학습세대에게 보다 친밀한 형태인 태블릿 형태의 단말기 또는 개인용 모바일 형태로 교과서를 제공하기 위한 중기적 전략을 추진한 바 있다. 현재는 일선현장에 PDF 형식의 CD롬 기반 교과서를 배포한 상태이며, 차후 보다 가볍고 중립적 포맷형식을 지원하는 발전된 e-교과서 개발이 시급히 요구되고 있다[3][5]. 특히 스마트 교육 및 클라우드 서비스 제공을 위해서는 과거에 개발된 디지털교과서 포맷을 다양한 형태의 단말기기에 정확하게 배포할 수 있어야 하므로 표준 포맷에 기반한 e-교과서 사업 전략은 스마트 교육의 성공을 위한 주요한 기술적 이슈로 제기되고 있다.

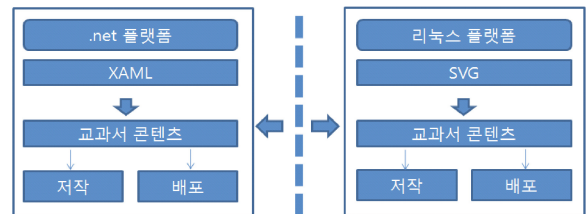
본 논문에서는 현재 전자책 시장에서 통용되는 표준 포맷 개발 사례를 살펴보고, 또한 e-교과서에서 필요한 학습자 상호작용을 분석하기 위하여 초·중·고 국·영·수 교과서의 구조 및 기능 요소 그리고 학습 상호작용을 도출하였다. 또한 분석된 상호작용 요소를 표준 포맷에 어떠한 형태로 적용 및 확장되어야 하는지를 분석 하였으며, 분석 결과를 토대로 서책형 교과서의 디지털화를 위한 전자책 포맷 표준화 방안을 제안하였다. 마지막으로 제안된 포맷에 기반하여, 실제 e-교과서 시범 콘텐츠를 제작 및 적용을 한 후 이에 대한 장단점 파악을 통하여 표준화 가능성을 제시하였다.

2. 디지털 교과서 및 전자책 동향

2.1 디지털 교과서 플랫폼 현황

디지털교과서 플랫폼의 가장 중요한 이슈는 콘텐츠와 관련된 문서 포맷 및 출력방안이다. (그림 1)과 같이 기존의 제작된 디지털교과서 콘텐츠는 각각의 플랫폼에 따라 서로 상이한 형태로 제작되었다. 즉 윈도우의 경우 MS에서 정의

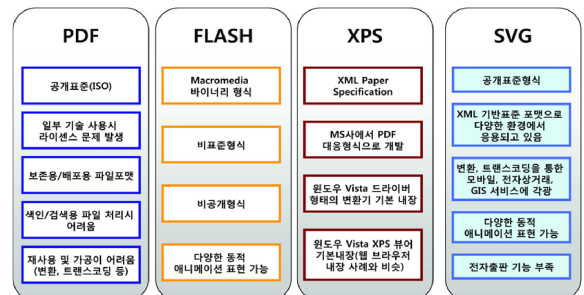
한 XAML 기반의 MS Silverlight 출력방식을 사용하며, 리눅스 기반 플랫폼의 경우 W3C 웹 표준인 SVG를 이용하여 콘텐츠가 제작되었다. 따라서 각각의 플랫폼에서 제작된 교과서 콘텐츠는 서로 호환이 되지 않기 때문에 학습자들은 해당 운영체제의 특정 플랫폼을 통해서만 교과서를 지원받게 된다. 또한 교과서를 제작하는 업체의 경우 XAML 혹은 SVG 형태로 콘텐츠를 직접 코딩하여 제작하고 있다.



(그림 1) 디지털교과서 플랫폼에서의 저작 및 배포 현황

2.2 전자출판 출력 기술 동향

현재 출판물 표기를 위한 다양한 출력 기술로는 다음 (그림 2)와 같은 기술들을 살펴볼 수 있다.



(그림 2) 출판물 표기를 위한 다양한 기술

특히 PDF나 Flash 등은 상업적으로도 크게 사용되고 있는 포맷이며 다양한 장점들을 포함한다. 그러나 이러한 포맷들은 특정 업체의 기술로서 내부 포맷에 대한 정보는 알 수 없을 뿐만 아니라 이를 출력하기 위해서는 특정업체의 엔진을 사용해야 한다. 따라서 대한민국 디지털교과서를 출력하기 위해서는 특정 업체의 출력기술에 의존하여 플랫폼을 구성해야 한다.

특히 현재 XAML은 높은 출력물을 보장하며 관련된 다양한 API를 Microsoft사에서 제공하는 등의 장점을 포함함에도 불구하고 표준기반의 디지털교과서 출력방안은 반드시 지속적으로 연구되어야 한다.

H5Station: HTML5-기반 효과적인 멀티미디어 파일 관리 시스템

정다운*, 원지혜*, 김수정*, 이종우**

요 약

최근 다양한 스마트폰들이 급속도로 퍼지면서 이를 이용하여 멀티미디어 콘텐츠를 감상하는 사용자들이 늘고 있다. 하지만 점점 콘텐츠의 수가 증가하면서 사용자들은 자신 소유의 멀티미디어 파일을 더 빠르고 쉽게 검색/관리하여 감상하고 싶다는 필요성을 느끼게 되었다. 따라서 본 연구에서는 자신이 소유한 사진, 음악, 동영상 같은 멀티미디어 콘텐츠를 통합 관리/감상할 수 있는 시스템을 제안한다. 나만의 사진, 음악, 동영상 통합 관리 시스템인 H5Station은 웹 표준 기술인 HTML5을 기반으로 개발하였기 때문에 클라이언트 브라우저 종류에 관계없이 동작하고, 아울러 클라이언트가 컴퓨터 웹 브라우저인지 아니면 모바일 웹 브라우저인지를 구분하여 차별화된 사용자 인터페이스를 지원한다. 주된 기능에는 회원 정보 변경 및 파일 분포 현황, 파일 업로드, 조건 검색 및 관리, 파일 감상, 사용자 맞춤 기능 등이 있다. H5Station의 사용자는 자신 PC에 접속하여 원하는 파일을 검색할 수 있으며, 특히 모바일 웹 애플리케이션으로 언제 어디서나 감상할 수 있다.

H5Station: An Effective HTML5-based Multimedia File Management System

Jeong daeun*, Won jihye*, Kim sujung*, Lee jongwoo**

Abstract

As various smartphones users are increasing rapidly, the more contents they keep in their devices, however, the larger necessity for the effective and easy management of their files they want. To satisfy this need, we propose an integrated multimedia files management system, H5Station. By using H5Station client, users can connect to the H5Station server running on their own PC, and manage their multimedia files effectively, and enjoy every file they have. The H5Station client is developed by using HTML5 standard technology. Therefore H5Station client can run on any HTML5-supported browser. Main functions of the H5Station are as follows: configuring user settings, charting the distribution of user files, upload/download/deletion of files, synchronizing files' metadata between client and server, playing user-selected files, searching files users want, and finally creating user-customized buttons. Through these functions, users can easily find, play, and manage their multimedia contents at any place and anytime.

Keywords : HTML5, Client-Server, Multimedia File Management, Mobile Browser

1. 서론

※ 제일저자(First Author) : 정다운

접수일:2011년 12월 04일, 수정일:2012년 01월 25일

완료일:2012년 03월 06일

* 숙명여자대학교 멀티미디어학과

{jdaeun0426, wonone43, diem880521}@gmail.com

** 숙명여자대학교 멀티미디어학과 부교수(교신저자) bigrain@sm.ac.kr

■ 본 연구는 숙명여자대학교 2012학년도 교비연구비 지원에 의해 수행되었음

스마트 시대를 맞이하여 스마트폰, 스마트 태블릿, 카메라 등 다양한 기기들의 보급이 본격화되면서 우리는 주변에서 스마트 기기를 쉽게 찾을 수 있다. 사람들은 스마트 기기를 사용하여 수많은 콘텐츠를 접할 수 있게 되었고, 가족과 친구들의 사진, 녹음한 음성 파일, 좋아하는 가

수의 음악 파일, 특별한 날의 동영상 파일 등을 컴퓨터에 저장한다. 하지만 시간이 흐른 뒤 많은 양의 파일 중에서 원하는 파일을 찾을 때 여러 가지 문제점들이 발생한다.

첫째, 사람은 많은 양의 파일 이름을 기억할 수 없으므로 직접 파일을 검색하고 찾는 데에 많은 시간이 소요될 수 있다. 다양한 기기에 혼재되어 있는 파일들을 일일이 분류하여 컴퓨터에 저장하기란 쉽지 않은 일임은 분명하다.

둘째, 스마트 기기 내에 저장된 파일들을 컴퓨터 또는 다른 기기에서 불러오기가 불편하다. 왜냐하면 한 사람이 여러 개의 기기를 소유하는 경우, 기기 사이에 파일 이동은 동기화 과정이 필수적이기 때문이다. 뿐만 아니라 여러 개의 기기 안에 동일한 파일이 용량을 차지하므로 불필요한 저장 공간이 낭비된다.

셋째 모바일 기기는 한정된 저장 용량을 가지고 있기 때문에 수많은 콘텐츠를 모두 보관하기에는 제한이 있다.

위와 같은 문제점들을 해결하고자 본 연구에서는 사용자들이 언제 어디서나 감상이 가능한 나만의 사진, 음악, 동영상 통합 관리 시스템을 개발하였다. 이 시스템의 명칭은 H5Station이고, 호스(hose)처럼 멀티미디어가 분출하는 통로란 의미와 HTML5로 구현했다는 의미를 담고 있다.

H5Station(이하 H5S)의 주된 기능으로는 업로드, 검색, 파일 감상 및 관리 그리고 사용자 맞춤형 기능 등을 들 수 있다. 사용자가 서버에 업로드한, 또는 이미 서버에 있는 파일들에 대해 웹 클라이언트와 모바일 클라이언트를 통해 H5S에 접속하여 언제 어디서나 감상할 수 있다. 또한 회원 소유의 사진, 음악, 동영상 파일들을 다양한 검색 조건을 통해 효율적으로 찾을 수 있게 해준다. 검색 결과로 출력되는 파일들은 바로 감상 및 삭제가 가능하며 회원이 직접 파일에 대한 정보를 수정할 수도 있다. 그리고 회원이 가장 많이 감상한 파일들, 가장 최근에 본 파일들 그리고 중요한 파일들을 맞춤형 기능으로 편리하게 제공한다. 마지막으로 H5S가 웹 표준 기술로 대두되고 있는 HTML5를 기반으로 개발되었다는 점도 본 연구가 기여한 바라 할 것이다.

2. 관련 기술 및 관련 연구

2.1 HTML5 웹 표준

2.1.1 HTML5 개요

HTML5는 W3C(World Wide Web Consortium)를 중심으로 개발되고 있는 HTML 4.01, XHTML 1.0, DOM Level 2 HTML에 대한 차세대 웹 표준이며, 플러그인 기반이 아닌 웹 애플리케이션 개발을 위한 표준 개발을 목표로 하고 있다[1].

HTML5 표준 중 마크업 부분은 기존 HTML4보다 훨씬 더 명확하게 의미 표현이 가능하도록 28개의 새로운 마크업이 추가되었다. 이를 통해서 검색 엔진 등 웹 콘텐츠를 기반으로 하는 다양한 서비스나 응용은 개선된 기능을 제공할 수 있을 것으로 기대된다. 현재 주요 브라우저들이 지원하는 HTML5의 주요 기능으로는 웹 폼, 캔버스, SVG(Scalable Vector Graphic), 멀티미디어 요소(Video/Audio), Geolocation, 오프라인 웹 애플리케이션, 로컬저장소, 웹 소켓, 웹 워커, 웹 SQL 데이터베이스가 있다.

2.1.2 HTML5 표준화 동향

2010년부터 세계적인 주요 IT 기업들은 HTML 기술 개발, 표준화 활동 및 서비스 적용을 시작하고 있다. 먼저, 구글은 안드로이드 웹 플랫폼에서 HTML5의 많은 기능을 제공하고 있다. 이와 더불어 애플도 아이폰과 아이패드에서 어도비 플래시를 배제하고 HTML5를 적용하고 있으며, 이미 아이튠즈 등 주요 소프트웨어에 웹 표준 기술을 적극적으로 도입하고 있다. 마이크로소프트의 경우도 실버라이트와 IE9과의 역할을 분명히 정리하면서 IE9에 HTML5 기능을 빠르게 추가하고 있다.

HTML5 지원 정도를 수치(점수)로 알려주는 사이트[2]에 따르면 표 1과 같이 웹 브라우저마다 HTML5 지원의 차이가 있어 아직 브라우저 간 호환성 문제가 남아있는 실정이다.

2.2 아이튠즈(iTunes)

아이튠즈는 애플에서 만든 멀티미디어 플레이어 및 동기화 프로그램이다. 주요 기능은 ID 태그로 컴퓨터 내의 동영상 및 음악을 감상, 검색,

독서장애인용 모바일 전자책뷰어 인터페이스 설계

이 경 희, 김 태 은, 이 종 우, 임 순 범

숙명여자대학교 멀티미디어학과

A Design of Mobile e-Book Viewer interface for the Reading Disabled People

KyunHee Lee, TaeEun Kim, Jongwoo Lee, Soon-Bum Lim

Division of Multimedia Science, Sookmyung Women's Univ.

- 논문지 구분 : 한국멀티미디어학회(분야: 멀티미디어 콘텐츠 및 응용)
- 한글 키워드 : 독서장애인, 전자책, DAISY, 어노테이션
- Keyword : Reading-disables, eBook, DAISY, Annotation
- 교신 저자 : 이종우
주소 : 서울시 용산구 청파동 2가 숙명여자대학교 멀티미디어학과
우편번호 : 140-742
- TEL : 02-710-9952
- FAX : 02-710-9704
- E-mail : bigrain@sm.ac.kr
- 본 연구는 숙명여자대학교 2011학년도 교내연구비 지원에 의해 수행되었음
- 본 연구는 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2010002415)

독서장애인용 모바일 전자책뷰어 인터페이스 설계

A Design of Mobile e-Book Viewer interface for the Reading Disabled People

요 약

최근 전자책 시장이 활성화됨에 따라 전자책 단말기에서부터 스마트 기기의 소프트웨어 리더까지 각종 전자책뷰어가 등장하고 있다. 하지만 시각장애인, 난독증, 학습장애인과 같은 독서장애인을 위한 모바일 전자책 인터페이스에 대한 개발과 연구는 부족한 실정이다. 비장애인을 대상으로 만들어진 전자책뷰어는 독서장애인에게 그대로 적용할 수 없기 때문에 독서장애 사용자의 특성에 따라 차별화된 인터페이스가 요구된다. 이에 본 논문에서는 독서장애인용 전자책 표준 형식을 지원하는 모바일 전자책 뷰어 인터페이스 모델을 제안한다. 제시 모델은 전맹인, 저시력인, 학습장애인 등 사용자의 특성 및 상황(context)에 따라 차별화된 인터페이스를 제공한다. 아울러 독서장애인용 어노테이션 시스템을 지원함으로써 기존의 독서장애인용 오디오북과는 다른 사용자-전자책 간의 상호작용을 지원한다. 또한 본 모델을 이용하여 스마트폰 플랫폼인 안드로이드 환경에서의 독서장애인용 전자책뷰어 프로토타입을 구현하고 그 활용 가능성을 제시한다. 본 연구의 결과는 국내 인구 10%에 해당하는 독서장애인의 효율적인 독서활동을 지원할 수 있다.

Abstract

As the eBook market grows fast recently, various eBook viewer solutions such as hardware viewers and software readers came out to the market. We can, however, hardly find mobile eBook interfaces for the reading disabled people who have difficulties in reading for their visual impairment or learning disabilities, or dyslexia. An eBook viewer interfaces for the reading disabled people should be carefully and distinctively designed because the reading disabled people cannot use normal versions of eBook viewer. In this paper, we suggest a eBook viewer interface model to make the reading disabled people read eBooks easily. Depending on the type of the reading disabled people: the full blind, the almost blind, the just learning disabled, our model provides an adaptive interface to make them read eBooks effectively. In addition, unlike the existing simple audio books, we also support annotation systems to make the reading disabled people interact with eBook viewer. To show the effectiveness of our model, we implemented an eBook viewer prototype on an android-based mobile device. We are sure that our model and implementation can make the reading disabled people, who is 10% of all the domestic people, read eBooks effectively.

Keyword : eBook, reading disabled people, android, interface

사진 콘텐츠의 분류를 위한 의미적 유사도 기반 구조적 태그 클러스터링 기법

원지현*[○] 박희민** 이종우*

*숙명여자대학교 멀티미디어학과 **연세대학교 컴퓨터과학과
blessing_of@naver.com, parkheemin@gmail.com, bigrain@sookmyung.ac.kr

A Structured Tag Clustering Method using Semantic Similarities for Photo Categorization

Jihyeon Won*[○] Heemin Park** Jongwoo Lee*

*Dept of Multimedia Science, Sookmyung Women's University

**Dept of Computer Science, Yonsei University

요 약

개인이 사용할 수 있는 스마트 기기가 다양해지면서 여러 기기로 생산된 사진 콘텐츠가 어떤 기준이나 규칙 없이 분산되어 있어 콘텐츠를 관리하고 원하는 콘텐츠를 검색하는 것이 어려워졌다. 따라서 본 논문에서는 개인 사진 콘텐츠를 효과적으로 분류하기 위하여 의미적 유사도를 기반으로 한 태그 클러스터링 기법을 제안한다. 태그들 사이의 유사도를 계산하여 서로 관련이 있다고 판단되는 태그들을 클러스터링 하는데, 태그가 같은 클러스터에 포함되어 있으면 그 태그를 가진 사진들도 유사성을 가진다고 볼 수 있으므로 개인 사진들을 의미에 따라 분류하는데 이용할 수 있다.

1. 서 론

스마트 기기의 대중화로 방대한 양의 사진 콘텐츠가 개인에 의해 생산되고 소비되고 있다. 개인용 컴퓨터, 스마트 폰, 태블릿 등을 통해 생성된 멀티미디어 콘텐츠의 양이 늘어갈수록 (개인 사진을 관리하고 효율적으로 관리, 검색할 수 있는 방안이 필요하게 되었다.) 개인 사진이 다수의 기기와 저장장치에 분산되어 있으므로 사진 콘텐츠 모음의 파편화, 중복성이 발생하여 효율적인 검색, 열람이 어려워지게 되었다. 이러한 문제를 해결하고 효율적으로 사진을 관리하기 위해서는 사용자가 일일이 수작업으로 분류하고 저장하는 방법보다는 자동으로 사진을 분류하고 관리하는 방법을 필요로 한다. 이 논문에서는 계층적 및 구조적인 태그 분류를 이용하여 개인 사진을 자동으로 분류할 수 있는 태그의 클러스터링 기법을 제안한다. 웹에 있는 개인 컬렉션의 사진이 태그를 가지고 있을 때, 그 사진에 있는 태그를 이용하여 의미에 따라 관련 사진끼리 클러스터링하여 사진을 분류하게 된다.

2. 관련 연구

사진 분류와 태그에 관한 기존 연구는 연관된 태그를 추천해주어 사진 분류에 도움을 주는 연구와 태그 자체의 유사도 분석에 관한 연구가 있다. 먼저 사진 분류에 관한 기존 연구는 지도를 타일로 나누어 위치

태그(Geo-tag)에 따라 비슷한 위치에서 찍힌 사진을 분석하고 시각적 유사도(visual similarity)에 따라 같은 물체나 사물이 있는 사진을 모아주는 방법[1]과 사용자의 웹 페이지 브라우징 행동을 통해 키워드를 뽑아내 각 키워드간의 연관성에 따라 태그를 추천하고 분류하여 사진에 적용하는 연구가 있다.[2] 이런 연구 중에 태그를 이용한 사진 분류에 초점을 맞추어 살펴보면 태그 클러스터링 관련 연구들을 볼 수 있다. 먼저 SheepDog[3]에서는 미리 정해놓은 컨셉 중에 하나를 키워드로 검색하여 사진과 그룹을 클러스터링하여 학습시키고 학습된 데이터를 통해 새로운 사진의 특징(feature)을 추출(extraction)하여 그룹과 태그를 추천하는 방법을 제시한다. 또한 의미적 태그 클러스터링 방법을 제안하고 태그를 클러스터링하여 의미에 따라 계층적으로 트리를 만든 연구도 있다.[4] 이 연구에서는 주어진 조건에 맞춰서 모든 태그에 대해 각 클러스터에 포함될 확률을 계산하고 실제적 클러스터(effective cluster)가 될 때까지 클러스터를 분할한다. 실제적 클러스터가 되면 그 클러스터를 대표하는 태그를 리딩 태그(Leading tag)로 지정하고 의미적 클러스터링을 하게 된다. 또한 태그들의 동시 발생 빈도수(co-occurrence)에 따라 유사성을 측정하는 연구도 있는데[5] 어떤 한 페이지에서 태그들 사이의 동시 발생 빈도수를 세어서 높게 나타나는 태그들을 강한 연관성이 있는 태그(strongly related tag)라고 정하여 태그의 의미 사이의 연관성을 알아보는 연구이다. 이를 통해

¹이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (2011-0014991)

태그들간의 유사성이나 관련성을 알아내어 의미 있는 계층적 태그 구조를 만드는데 활용할 수 있다. 본 논문에서는 태그를 구조적으로 클러스터링 하기 위하여 랜덤으로 태그 사이의 의미적 유사성을 측정하고 그것을 이용하여 전체 클러스터를 2^n (n 는 분할 횟수)개로 분할해 나가는 방법을 제시한다.

3. 태그 클러스터링 기법

태그는 사진의 내용을 대표하는 단어이기 때문에 사진의 의미를 알려주는 중요한 데이터이다. 따라서 이 데이터를 이용하면 사진의 의미적 분류도 가능해진다. 또한 태그를 구조적으로 클러스터링하게 되면 1차원적인 나열이 아닌 계층적인 관계가 만들어지기 때문에 의미적인 포함관계를 알 수 있게 되어 좀 더 명확한 의미적 클러스터링이 가능하다. 태그의 의미적 클러스터링에는 상위 태그의 개념이 생기기 때문에 한 클러스터를 대표하는 하나 혹은 몇 개의 태그가 생기고 그에 따라 사진을 의미적으로 분류하는 것이 용이해진다.

태그를 의미에 따라 클러스터링 하기 위해서 다음과 같은 기법을 사용한다. N 개의 태그를 가진 클러스터가 있을 때 먼저 두 개의 클러스터로 나누기 위해 랜덤으로 태그를 한 개(a) 뽑는다. 그 후 나머지 태그에서 $N/10$ 개 만큼 랜덤으로 태그를 뽑아서 a 와의 유사도를 계산한다. 계산한 결과 중 가장 거리가 먼 태그(b)를 선택하여 나누어질 2개 클러스터의 기준 태그로 둔다. 그 후 기존 클러스터 안의 전체 태그에 대해 a 와 b 와의 유사성을 계산하여 유사도가 더 높은 태그가 있는 클러스터에 포함시킨다. 나누어진 클러스터의 원소의 개수가 $N/10$ 보다 적으면 더 이상 나누지 않고 그렇지 않으면 같은 과정을 반복하여 클러스터링한다. 유사도를 계산하는 방법은 [5]에서 제안된 방법을 이용한다. 한 페이지 안에서 동시에 나타나는 횟수를 측정하여 높게 나타나는 것을 연관성이 강하다고 보고 점수를 주는 방법이다. 이를 이용하여 주어진 사진의 모든 태그들에 대해 동시 발생 빈도수를 계산하여 동시 발생 빈도수가 높을수록 유사도가 높아지도록 점수를 부여한다. 이 방법을 사용하면 실제 사용자가 많이 사용하는 태그들의 유사성을 좀 더 정확히 측정할 수 있기 때문에 사용자 별 맞춤형(customized) 분류를 하는데 도움을 준다.

그림 1의 의사코드(pseudo-code)를 설명하면 먼저 N 개의 태그를 가진 하나의 클러스터 C 가 있을 때 처음 클러스터 C 를 queue Q 에 넣는다. 그 후 Q 에서 C 를 pop하여 for문을 시작한다. 먼저 종료조건에 만족하는지 검사하고 더 이상 나눌 필요가 없으면 최종 클러스터 집합인 queue F 에 넣는다. 그렇지 않은 경우 앞에 언급한대로 태그 a 와 태그 b 의 후보 $N/10$ 개 태그를 비교하여 가장 거리가 먼 두 개의 태그를

선택하여 클러스터 A 와 B 의 기준 태그(a, b)로 정하고 나머지 모든 태그에 대해 기준 태그 a, b 와의 거리(유사도)를 계산하여 더 가까운 기준 태그의 클러스터에 포함시킨다. 분할이 완료된 클러스터들은 다시 Q 에 넣는다.

```

Initialize Q.
Q는 분류하기 전 cluster들의 queue.
Initialize F.
F는 클러스터링이 완료된 final cluster들의 queue.
N개의 태그를 가진 클러스터 C = {t1, t2, ..., tn}
While Q is not empty do
  Pop C from Q
  for each cluster C do
    Initialize cluster A, B = {}
    if (C의 원소의 개수 <= n/10)
    {
      Push C into F
      break;
    }
  else
  {
    random pick tag A from C
    random pick tag B[0~([N/10]-1)]
    for i=0 에서 > [N/10]-1 까지
    { d = d(B[i], A)
      if (max<d)
        max = B[i]
    }
    a = A; // cluster A의 기준태그 a
    b = max; // cluster B의 기준태그 b
    for j=1 부터 j>N 까지
    {
      d(tj,a) compare d(tj,b)
      {d(tj,a)가 더 작으면 cluster A
       d(tj,b)가 더 작으면 cluster B
      }
    }
    Push A, B into Q
  }
}

```

그림 1 태그 클러스터링 알고리즘

4. 예상 결과

본 논문에서 제안한 클러스터링 기법의 구현이 완료되지 않아 본 절에서는 제안한 방법의 유효성을 보이기 위해 수작업에 의한 클러스터링 결과를 보인다.

안드로이드 기반 환경오염 신고 앱 설계 및 구현

이종우 한승희[○]

숙명여자대학교

bigrain@sm.ac.kr, noxhsh@naver.com

An Environmental Pollution Report application on Android Platform

Jongwoo Lee Seunghee Han[○]

Sookmyung Women's University

요 약

최근 스마트폰의 사용이 보편화 되고 다양한 어플리케이션이 등장하고 있다. 본 논문에서는 간편하게 환경오염 신고 기능을 제공하는 스마트폰 앱을 설계하고 구현한다. 구현된 앱에서는 기존의 번거로운 환경오염 신고방법을 개선하여 사진촬영과 위치 정보 등의 기능을 통해 간편하고 정확하게 환경오염을 신고할 수 있도록 하였다. 아울러 환경신고자에게 주어지는 각종 혜택 등에 대한 정보도 제공함과 동시에 환경오염 신고를 할 때마다 자라는 환경나무 애니메이션 기능도 제공한다. 본 서비스를 통하여 정확하고 간편한 환경오염 신고가 가능해진다면 나아가 사용자의 환경보호 의식도 높일 수 있을 것으로 기대된다.

1. 서 론

최근 끊임없는 산업발전으로 생활이 더욱 편리해 지고 있다. 하지만 이러한 발전에도 불구하고 항상 대두되는 문제가 바로 환경문제이다. 하지만 현재 환경오염을 신고하기 위해서는 개인이 직접 지역별 환경부홈페이지[1]에 접속하거나 환경부에 전화를 걸어야 하는 번거로움이 있다. 이러한 번거로움과 신고하고자 하는 시민들의 의식부족으로 제대로 된 환경오염신고가 이루어지지 못하고 있다. 이에 스마트폰을 이용하여 홈페이지에 접속해야 하는 번거로움 없이 환경오염 현장에서 빠르고 간편하게 신고할 수 있는 앱을 기획하게 되었다. 이 앱에서는 사용자가 환경오염 현장에서 바로 위치정보를 받아와 사진촬영과 함께 신고할 수 있다. 또한 환경오염신고에 대해 부족한 시민의식을 높이고자 신고자에게 제공되는 포상금제도에 대한 설명을 상세하게 제공한다. 이를 통한 기대효과는 다음과 같다.

- 간편한 어플리케이션으로 정확하고 간편한 환경오염 신고 제공
- 시민들의 환경문제 의식 고취

본 논문의 구성은 다음과 같다. 제 2장에서는 기존 환경오염 신고 시스템을 설명한다. 제 3장에서는 본 시스템의 설계 내용, 제 4장에서는 개발환경과 구현결과를 보인다. 마지막 제 5장에서는 본 논문의 결론을 맺는다.

2. 관련 연구

본 장에서는 기존 사이버 환경오염 신고 방법들에 대하여 설명한다.

그림 1은 부산광역시 홈페이지의 환경신고를 위한 페이지이다. 기본 구성은 제목, 작성자, 주관부서, 처리현황, 공개여부, 작성일로 되어있다.

번호	제목	작성자	주관부서	처리상황	공개	작성일	조회
617	매연차량	김○○	환경보전과	답변완료	X	2011-09-14	0
616	매연차량신고	김○○	환경보전과	답변완료	X	2011-09-14	4
615	불안토도장	김○○	환경보전과	불매(답장자차량)	X	2011-09-07	5
614	대기오염	홍○○	환경보전과	타기관이송	X	2011-08-28	0
613	매연행는 그레이스	김○○	환경보전과	답변완료	X	2011-08-25	0
612	삼한 석회 공장	윤○○	환경보전과	답변완료	○	2011-08-25	90
쓰레기거름 국책기해 계속 세우는대 녹가우 및 안지, 약 이 때문에 못 심었습니다!							
610	영도구 동화단 무단투기가 의심됩니다.	박○○	자정순화과	타기관이송	X	2011-08-17	0
609	영도구 동화단 무단투기가 의심됩니다.	박○○	자정순화과	타기관이송	X	2011-08-17	0
608	KTL의 대곡로 불태우차량 전술을 적당히주세요	정○○	건축주책담당관	타기관이송	○	2011-08-16	0

[그림 1] 부산광역시 홈페이지

신고 내역 작성 시에는 민원내용을 작성할 수 있고 첨부파일을 추가할 수 있다. 하지만 이러한 사이버환경신고 페이지는 신고를 위해 사용자가 홈페이지에 접속하여야 하기 때문에 신고지역에서 바로 신고할 수 없다는 단점을 가지고 있다. 때문에 이러한 기본 구성을 바탕으로 하되 언제 어디서나 바로 신고할 수 있는 본 모바일 앱을 기획하였다.

3. 시스템 설계

3.1 주요기능

본 시스템은 환경오염 현장을 즉시 신고할 수 있는 것이 핵심기능이다. 또한 현재까지 신고 된 내역을 볼 수 있는 신고내역확인, 포상금제도 설명, 자신의 E지수 보기의 기능이 있다.

3.1.1 환경신고

환경오염을 신고할 수 있는 네 가지의 기능을 제공한다. 먼저 환경오염 현장에서 바로 사진 촬영 후 신고할 수 있도록 한다. 두 번째로는 자신의 갤러리에 있는 사진을 불러와 신고할 수 있도록 한다. 세 번째로 사진이

없는 경우 내역 작성만으로 신고할 수 있도록 한다. 마지막으로 작성 없이 전화로 바로 신고하고자 할 때 환경 신고 대표전화인 128로 연결되도록 기능을 제공한다.

3.1.2 신고내역 조회

현재까지 신고한 내역을 확인할 수 있다. 처음에는 신고유형과 신고제목 그리고 처리현황만으로 한 눈에 신고내역을 리스트형태로 확인할 수 있다. 상세 신고내용은 해당하는 행을 선택하여 확인할 수 있다.

3.1.3 포상금제도

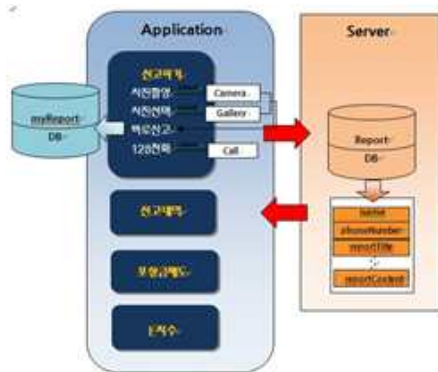
환경부의 '환경오염행위신고 포상금 운영지침' 보고서에 따라 신고 시 신고자에게 제공되는 포상금제도에 대한 설명을 지급기준, 지급시기, 지급기관 등 자세하게 제공한다.

3.1.4. E지수

자신의 신고횟수에 따라 지구나무를 키운다는 콘셉트로 신고횟수가 많을수록 지구나무가 점점 자라게 된다. 또한 카운트하여 자신이 신고한 수를 볼 수 있다. 사용자에게 흥미요소를 제공함으로 지루한 환경신고가 아닌 즐거운 신고의 이미지를 주고자 한다.

3.2 시스템 구성도

그림 2은 앱의 전체 구조도이다. 클라이언트, 서버, 그리고 DB의 각 기능과 데이터 흐름을 보여주고 있다.



[그림 2] 전체 구조도

클라이언트, 즉 EcoReporter에서는 신고하기, 신고내역, 포상금제도, E지수의 기능을 가지고 있다.

서버에서는 클라이언트의 각기 다른 요청에 따라 다양한 기능을 수행한다. 먼저 클라이언트에서 신고된 내역은 서버로 전송되고 서버에서는 MySQL을 통해 저장한다. DB테이블은 신고제목, 신고내용, 오염위치, 신고유형, 사진데이터, 신고자, 신고자연락처로 구성된다. 클라이언트에서 신고내역보기를 선택하면 서버에서 DB에 있는 데이터를 받아와 보여준다. 신고내역을 저장할 때에는 insert SQL명령문을 사용하여 DB에 저장하고, 신고내역을 불러올 때에는 select 문을 사용하여 DB에서 특

정 스키마 값을 받아와 보여준다[2].

또한 클라이언트 안에서는 따로 SQLite를 이용하여 내부로 자신의 신고정보를 저장한다. 또한 저장한 개수를 카운트하여 전체 E지수에서 활용한다.

4. 개발 환경 및 구현

4.1 개발 환경

EcoReporter의 개발환경은 다음과 같다. 서버는 Tomcat을 이용하여 구축하였고, Database는 MySQL을 이용하였다. 클라이언트는 자바 통합 툴인 Eclipse Indigo에서 안드로이드 개발자 도구 SDK를 설치하여 개발하였다. Android 2.2. 프로요를 사용하고 에뮬레이터와 갤럭시S2를 이용해 실행하였다.

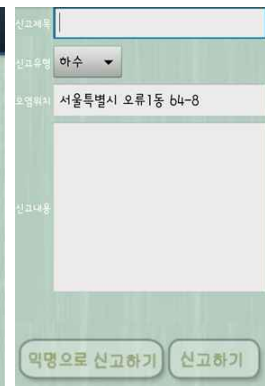
4.2 구현

4.2.1 메인화면

앱의 메인화면에는 신고하기 페이지와 함께 상단에 4개의 탭 메뉴가 있다(그림 3). 첫 번째 탭인 신고하기 탭에서는 환경오염을 신고하기위한 네 가지의 버튼을 제공한다. 첫 번째는 사진촬영 버튼으로 선택하면 바로 카메라 화면이 나타나 사진을 촬영할 수 있다. 두 번째는 사진선택 버튼으로 선택하면 자신의 갤러리가 나타난다. 여기서 신고자는 신고하고자 하는 사진을 선택하여 신고할 수 있다. 세 번째 바로신고 버튼은 사진 없이 바로 신고페이지가 나타나 신고할 수 있다. 마지막 128전화는 환경신고 대표전화인 128로 바로 전화가 연결되게 해준다.



[그림 3] 메인 화면



[그림 4] 신고내역 작성

4.2.2 신고하기

첫 번째 탭인 신고하기를 선택하면 신고할 수 있는 화면이 나타난다. 여기에서 사진촬영이나 사진선택, 바로 신고를 선택하였을 때 최종적으로 나오게 되는 화면은 그림 4와 같다.

신고내역 작성 화면에서는 신고제목, 신고유형, 오염위치, 신고내용을 작성할 수 있다. 오염위치는 GPS 위치정보를 얻어와 Geocoder를 활용해 주소를 자동으로 변환

Efficient techniques on retrieving bio-information for active U-healthcare

Young-Ho Park

Received: 24 August 2011 / Accepted: 30 December 2011
© Springer-Verlag London Limited 2012

Abstract Recently, active prevention healthcares are needed for potential patients to be suffered in the future as the forecasted diseases inherited from ancestors. We call active U-healthcare, for providing active, periodic, and continuous medical treatments depending on inherited heterogeneous states in DNAs of patients, such as diabetes, heart diseases, and female diseases. However, the bottleneck of the aggressive active U-healthcare is memory overhead in DNA sequence analysis of each patient since the sequences of DNAs have massive volume. Thus, the efficient retrieve of the many disease patterns in originally recorded on DNAs of potential patients is a major problem. This paper focuses on a novel method for efficient retrieving of disease patterns using a suffix tree in memory. The suffix tree is widely used in the similarity search for sequences consisting of limited characters. It is efficient when the occurrence frequency of a common prefix is high. Since in-memory suffix tree construction algorithms do not scale up, a large-scale disk-based suffix tree construction algorithm, TRELLIS, has been proposed recently. However, the algorithm requires a large amount of memory, disk space, and disk I/Os in order to merge sub-trees having a common prefix. In this paper, we propose a new non-merging method, called NST. The experimental results show that NST constructs an index using less memory than TRELLIS.

Keywords Active U-healthcare · Suffix tree · Variable-length prefix

1 Introduction

U-health refers to health management and medical services that utilize wired/wireless networking such as remote monitoring and are capable of use at “any given” time and place [1]. The purpose of the research is developing a U-healthcare system that helps health management users to manage their health for themselves as precautionary measures. Generally, the system consists of three modules such as sensing, management, and analysis. The goal of the process includes proper treatments according to the patients. However, nowadays, a problem of U-healthcare is passive actions depending on the status of sensed data. Recently, in contrast to the concept, active prevention healthcares are needed for potential patients to be suffered in the future as the forecasted diseases inherited from ancestors. For this, this paper proposes a term, Aggressive U-healthcare, for providing active, periodic, and continuous medical treatments depending on inherited heterogeneous states in DNAs of patients, such as diabetes, heart diseases, and female diseases. However, the bottleneck of the aggressive U-healthcare is overhead in DNA sequence analysis of each patient since the sequences of DNAs have massive volume [2]. Thus, the efficient retrieve of the many disease patterns in originally recorded on DNAs of potential patients is a major problem. This paper focuses on a novel method for efficient retrieving of disease patterns using an index in memory. Through the efficient indexing, we can actively notify individual active U-healthcare actions such as food on demand, kinds of exercises, and mental treatments depending on each potential patient. The efficient pattern retrieval acts as a DNA-aware disease prevention system.

The similarity search for large sequences consisting of limited and repeated characters is important in information

Y.-H. Park (✉)
Department of Multimedia Science,
Sookmyung Women's University, Seoul, Korea
e-mail: yhpark@sm.ac.kr

retrieval [3, 4]. To this end, sophisticated similarity search algorithms for various sequences have been proposed. However, existing search algorithms have several problems. The database should be loaded in main memory or the search speed reduced in proportion to the size of a database. In addition, the word length affects the accuracy of results. To solve these problems, many studies use indexing methods with a suffix tree [5–7]. The suffix tree is efficient when the frequency of common prefix is high. It is also widely used in sequence similarity search for DNA sequences, MIDI sequences of music, documents, and limited patterns [8–12]. We will explain the related works in detail in Sect. 2.

In this paper, we propose the space-efficient disk-based suffix tree construction method, called NST (Non-merging suffix tree). In contrast to the existing method [13], we do not need the merging step, which incurs a large amount of memory, disk space, and disk I/Os.

The NST algorithm consists of three steps. First, we partition the sequence. In this step, with full scanning of the sequence, we create partitions for the suffixes that have common prefixes. When the number of the suffixes that have the same prefix is higher than a given threshold, we expand the prefix and repartition the suffixes. Second, we build suffix sub-trees. We construct suffix sub-trees for the partitions in which the number of the suffixes is lower than the threshold. Third, we process the query. We construct a tree, which has the suffix as the prefix of the partition built with sub-trees, in main memory. It can search in the suffix tree by integrating super-trees of the sub-trees. In this paper, we propose the NST algorithm to remove a merging step for the organization of all sub-trees that have a unique prefix. Then, we compare the performance of the NST algorithm with that of the TRELLIS algorithm [13], which is the state-of-the-art method. The NST algorithm has an advantage of being able to reduce the usage of main memory and disk space. For large sequence data, it is important to construct index with limited memory.

2 Related work

The existing methods for sequence similarity search are Phased Tree [14, 15], TDD (top-down disk-based) [16, 17], STTD64 (suffix tree, top down, 64 bits) [18], and TRELLIS (External Suffix Tree with Suffix Links for Indexing Genome-scaLe Sequence) algorithms [13]. We briefly review each method in this section.

A lot of existing algorithms for sequence similarity search have a problem that the search speed reduces in proportion to the size of sequences. Thus, indexing is required to improve speed and accuracy in the sequence similarity search. Many studies use a suffix tree to index sequences.

The Phased Tree algorithm [14, 15], which was proposed by Ela Hunt, is the method to solve a memory problem. It constructs a large-scale suffix tree through the disk partition. This approach removes the memory bottleneck [19] and improves the locality of references as removal of the suffix link [14]. However, the Phased Tree algorithm has the problems that random access of disk frequently occurs while constructing a tree, since it inserts all suffixes including the same prefix into the sub-tree of the disk based on PJama (Java-based object store).

The TDD algorithm [16, 17] solves the problem of Phased Tree algorithm. The TDD algorithm also does not use suffix link and constructs a suffix tree based on the Wotdeager algorithm [20]. The TDD algorithm constructs all sub-trees and includes the strategy for efficient cache usage when the first access occurs. However, as the size of a full sequence is bigger, an additional cost of preprocessing occurs. Moreover, a data skew [21] occurs in some cases of a partial partition.

The STTD64 [18] indexes large sequence data and can handle the entire human genome on typical desktops. It also reduces disk I/Os by storing the depth values for the leaf nodes. The STTD64 has a search efficiency and better scalability for larger sequences. However, STTD64 has increased construction time and storage space compared to TDD [18].

To solve the skew problem, the TRELLIS algorithm [13] builds the suffix tree based on a partitioning method via variable-length prefixes and a suffix sub-tree merging algorithm. However, more memory and disk space are required when constructing a tree, and a large amount of disk I/Os occur in order to merge sub-trees that has a common prefix. Thus, we propose the NST algorithm to solve the above-mentioned problems and to reduce the memory and disk space overhead.

3 NST algorithm

In this section, we propose the NST algorithm to search sequences similar to the query sequence. We use a DNA sequence, as a target input data example, consisting of 4 letters in the paper. We explain the variable-length suffix tree construction in Sect. 3.1, the non-merging partitioning algorithm in Sect. 3.2, and the suffix node storage in Sect. 3.3.

3.1 Variable-length tree construction

Our algorithm uses the following approaches.

First, we remove suffix links. The reason to remove a suffix link is as follows. The suffix link exists in each internal node, and it points the tree node to index the suffix

A Design and Implementation of A Real-time Recruiting System on Android Platform

Sun-Young Ihm and Young-Ho Park*

Department of Multimedia Science, Sookmyung Women's University
Cheongpa-ro 47-gil 100, Yongsan-gu, Seoul, 140-742, Korea
{sunnyihm, yhpark}@sookmyung.ac.kr
*Corresponding Author

Recently, interests for recruiting and the rates of using smart phones are growing fast. This paper proposes a recruiting system, which is a smart phone application based on android platform, that helps real-time communication matching between users. We provide service for users which want to hire applicants and high credibility service for users. The paper presents design and implementation of a real-time recruiting system and shows the snapshots of the presented system.

Key Words: Real-time Recruiting System, Recruit, Android

1. INTRODUCTION

Recently, smartphones or mobile phones serve as the central computing and communication device in people's lives [2]. Such a trend is inevitable in the real world due to the tremendous growth in recent years of new smart devices, such as iPhones, iPads, and Android-based platform, which is expected to continue [1]. Also, people always want to get a job and hire good job seekers. There are some applications which help finding jobs, however, people who want to hire a job seeker cannot use these services. In addition, they cannot use in real-time for recruiting. This paper propose a real-time recruiting system, which is a smart phone application based on android platform [3,4,6]. Android is an open source framework designed for mobile devices [7]. In this paper, we called a person who wants to get a work as '*Recruitee*' and a person who wants to hire an applicant as '*Recruiter*'. Our system has advantages. First, we provide services for Recruiter. Recruiter can look through the resumes or search suitable applicants themselves. Second, we provide a real-time communication service between Recruitee and Recruiter. Third, the users give a reliability score of the information in our application for high credibility. Our system is designed in Korean, because of the convenience of implementation. The contributions we make in this paper are as follows.

(1) We provide 'A Recruiter Member Service' to solve the problem that only applicants can use exist services. Recruiter is looking through the resumes and communicating in real-time with Recruitee. Also, Recruiter can search suitable Recruitee directly. (2) We adopt 'A Reliability Evaluating Function' to maintain information in order that people can trust. All of members can evaluate information such as resume or recruiting notice.

The rest of this paper is organized as follows. Section 2 describes existing services related to this paper. Section 3 introduces the proposed system. Section 4 summarizes and concludes the paper.

2. RELATED SERVICES

In this section, we explain related services. We can find a job using ‘Alba-Chunkuk’[1] which is a smart phone application based on android platform. ‘Alba-Chunkuk’ has some advantages when we want to get a job. First, we can find a job with various options such as area, type of occupation and pay. Second, we can clipping the job information that we are interested in. Third, job seekers can apply using this application on smart phone. Another application, ‘Indeed’[5], is a web page for smart phone. This is different from other applications, because they operate similar with search engine. We can find lots of job information with ‘Indeed’ at a time. Moreover, they provide 53 languages. However, there is a significant problem exist. All of these recruiting applications services are only for job seekers. People who want to hire an applicant cannot use these applications.

3. A REAL-TIME RECRUITING SYSTEM

In this section, we introduce a real-time recruiting system which we proposed. Our system provides several services different from other existing systems. We have two main services: (1) A Recruiter Member Service and (2) A Reliability Evaluating Function. We introduce these two services in Section 3.1 and Section 3.2 and we analysis proposed system in Section 3.3.

3.1 A Recruiter Member Service

There is no application that services for Recruiter, however, we have ‘A Recruiter Member Service’. Recruitee can apply by our application, then, resume will send to Recruiter. Recruiter is looking through resumes and if there are some interesting Recruittees, Recruiter and Recruitee can communicate in real-time. In addition, Recruiter can search Recruitee directly and see the information of Recruitee. If they find the suitable Recruitee, then, press the ‘select’ button. Also, Recruiter registers the job post immediately when they need Recruitee by smart phone.

Recruiter can handle resumes easily by using our application and find suitable Recruitee quickly by real-time communicating with Recruitee. Moreover, Recruiter can register the job post himself and find the suitable Recruitee at once. Figure 1 shows the process of registering the job post by Recruiter.

3.2 A Reliability Evaluating Function

As many people using this real-time recruiting system, there is a possibility that the information in application is unreliable. Thus, we adopt ‘A Reliability Evaluating Function’ that people can trust all of information in our application. When users look through our application, if they think the information is credible, they press the ‘reliability’ button. Then the reliability score of the information increases or decreases 1 point. As the members who use this application give the reliability score themselves, we can trust the information in our application by the reliability score.

자연과학연구소 규정 (준칙)

제 1 장 총 칙

- 제1조 (명칭 및 위치) 이 연구소는 숙명여자대학교 자연과학연구소 (이하 ‘연구소’라 한다)라 하며 숙명여자대학교 내에 둔다.
- 제2조 (목적) 이 규정은 연구소의 조직과 운영에 관한 사항을 규정함을 목적으로 한다.
- 제3조 (설립목적) 연구소는 자연과학의 기초 및 응용분야에 관한 연구를 수행함으로써 자연과학의 발전에 기여함을 목적으로 한다.
- 제4조 (사업) 연구소는 제3조의 목적을 달성하기 위하여 다음의 사업을 한다.
1. 자연과학의 기초 및 응용분야의 연구, 개발 및 조사
 2. 학술 강연회와 연구발표회 및 공개강좌 개최
 3. 연구자료의 확보 및 연구 논문집 발간
 4. 위탁된 연구조사 및 감정 분석 실험
 5. 기타 연구소 설립목적에 부합되는 사업
- 제5조 (연구센터) 연구소는 제4조의 사업을 원활히 수행하기 위하여 연구센터를 둘 수 있다.

제 2 장 조 직

- 제6조 (소장)
- ① 연구소에 소장을 두며, 소장은 연구소를 대표하고 연구소 운영에 관한 제반업무를 통할한다.
 - ② 소장의 임명은 본 대학교 직제규정이 정하는 바에 의한다.
- 제7조 (연구센터장)
- ① 연구센터에 연구센터장을 두며, 연구센터장은 연구센터의 운영에 관한 제반 업무를 통할한다.
 - ② 연구센터장은 제9조에 규정된 연구위원 중에서 총장이 임명한다.
 - ③ 연구센터장의 임기는 2년으로 하되 연임할 수 있다.
- 제8조 (간사)
- ① 소장의 업무를 보좌하기 위하여 연구소에 간사를 둘 수 있다.
 - ② 간사는 제9조에 규정된 연구위원 중에서 소장의 제청으로 총장이 임명한다.
 - ③ 간사의 임기는 2년으로 하되 연임할 수 있다.

제9조 (연구위원)

- ① 연구소의 제반 연구활동을 수행하기 위하여 연구위원을 둔다.
- ② 연구소의 설립목적과 사업에 동의하는 본 대학교 소속 교원은 본인의 동의가 있는 경우 소장의 동의에 의해 연구위원으로 위촉할 수 있다.
- ③ 연구위원은 본 대학교에 재직하는 동안 연구위원의 지위를 유지할 수 있다.

제10조 (연구원)

- ① 연구소에 전임 및 비전임연구원을 둘 수 있으며 책임연구원, 선임연구원, 연구원으로 구성할 수 있다.
- ② 책임연구원은 박사학위를 소지한 자로서 소장의 제청으로 총장이 임명한다.
- ③ 선임연구원은 박사과정 또는 수료한 자로서 소장의 제청으로 총장이 임명한다.
- ④ 연구원은 석사학위과정 또는 소지자로서 소장의 제청으로 총장이 임명한다.
- ⑤ 연구원의 임기는 2년으로 하되 연임할 수 있다.

제11조 (연구조교)

- ① 연구소에 연구조교를 둘 수 있으며 제반 업무에 종사한다.
- ② 연구조교는 학사학위 이상학력의 소지자로서 소장의 제청으로 총장이 임명한다.

제12조 (객원연구원)

- ① 전문분야에 연구경력이 탁월한 해외 및 국내 학자들로서 본교 연구소의 연구활동에 참여하고자 하는 자는 제15조에 규정된 연구소 운영위원회의 심의를 거친 후 총장의 결재를 받아 객원연구원으로 소장이 위촉할 수 있다.
- ② 객원연구원의 임기는 1년으로 하되 연임할 수 있다.

제13조 (박사학위 소지자 연구과정 : Post-Doc. course) 연구소는 국내외에서 박사학위를 취득한 소장연구자들에게 연구의 기회를 제공하기 위하여 박사 후 연구과정을 운영할 수 있으며, 제15조에 규정된 연구소 운영위원회의 심의를 거친 후 총장의 결재를 받아 박사학위소지자 연구과정요원으로 소장이 위촉할 수 있다.

제14조 (감사)

- ① 연구소에 감사를 두며, 감사는 연구소의 회계결산에 대한 내부감사를 수행한다.
- ② 감사는 간사가 아닌 연구위원 중에서 총장이 위촉하며, 임기는 1년으로 하되 연임할 수 있다.
- ③ 감사는 감사종료 후 10일 이내에 감사보고서를 작성하여 총장과 제15조에 규정된 연구소 운영위원회에 제출하여야 한다.

제 3 장 운영위원회

제15조 (운영위원회)

- ① 연구소의 운영에 관한 중요사항을 심의하기 위하여 연구소 운영위원회(이하 “운영위원회”라 한다)를 둔다.
- ② 운영위원회는 소장과 10명 이내의 연구위원으로 구성하고 소장이 위원장이 된다.

제16조 (기능)

1. 사업계획에 관한 사항
2. 예산 및 결산에 관한 사항
3. 연구소 관련규정의 제정 및 폐지에 관한 사항
4. 객원연구위원의 위촉에 관한 사항
5. 기타 연구소 운영에 관한 중요사항

제17조 (회의)

- ① 운영위원회 회의는 위원장이 소집하고 그 의장이 된다.
- ② 운영위원회는 재적위원 2/3이상의 출석과 출석위원 과반수의 찬성으로 의결한다. 단, 가부동수인 때에는 의장이 결정권을 갖는다.

제 4 장 재 정

제18조 (재정)

1. 대학, 정부기관, 관련단체의 보조금
2. 기업 및 개인의 찬조금
3. 연구위원의 기여금
4. 기타 수입

제19조 (회계연도) 연구소의 회계연도는 본 대학교의 회계연도와 같다.

제20조 (장부 및 물품관리)

- ① 소장은 연구소의 재정에 관한 회계장부를 기록, 유지하여야 한다.
- ② 소장은 연구소내의 제반 시설과 물품을 성실히 관리하여야 한다.

제 5 장 보 칙

제21조 (재산의 귀속) 연구소가 해산할 경우 그 재산은 본 대학교에 귀속한다.

제22조 (세칙) 이 규정 시행에 필요한 세부사항은 소장이 운영위원회의 심의와 총장의 승인을 얻어 세칙으로 정한다.

부 칙

- (1) (시행일) 이 준칙은 1996년 4월 29일부터 시행한다.
- (2) (경과조치) 연구기관의 장은 이 준칙이 시행되는 날로부터 1개월 이내에 연구기관의 규정을 이 준칙에 저촉되지 않도록 개정해야 한다.

부 칙

- (1) (시행일) 이 준칙은 1996년 10월 14일부터 시행한다.

부 칙

- (1) (시행일) 이 준칙은 2004년 4월 9일부터 시행한다.