



# 한국자료분석학회

사무국: (우)06650 서울특별시 서초구 반포대로 24길 76, 6층 602호(서초동)  
전화: (02)6673-0709 / 회장 : 한상태 / 총무이사 : 이성건

문서번호 : KDAS-2402000

시행일자 : 2024. 6. 28.

(경유)

수 신 : 전 회원

참 조 :

선결			지	
	일자		시	
접 수	시간		결 재	
	번호			
담당자				

제목: 한국자료분석학회 2024년 하계 학술논문발표대회

- 귀 회원님의 무궁한 발전을 기원하며, 본 학회의 발전을 위한 회원 여러분의 관심과 노고에 깊이 감사를 드립니다.
- 본 학회 2024년도 하계 학술논문발표대회를 다음과 같이 개최함을 알려 드리오니 많은 참여 바랍니다.
  - 개최일자 : 2024년 6월 27일(목) ~ 28일(금)
  - 장 소 : 서울시립대학교(서울특별시 동대문구 서울시립대로 163)

한국자료분석학회장 한상태



PROCEEDINGS OF THE KOREAN  
DATA ANALYSIS  
SOCIETY

---

June 27-28, 2024

at

Department of Statistics,  
University of Seoul, Seoul, Korea.

(<https://www.uos.ac.kr>)

---

THE KOREAN DATA ANALYSIS SOCIETY Founded 1998

---

PROCEEDINGS OF THE KOREAN DATA ANALYSIS SOCIETY June 28, 2024



## 한국자료분석학회 2024년도 하계 학술논문발표대회 준비위원

### 조직위원회

위원장	한 상 태	호서대학교 빅데이터AI학부
부위원장	이 학 배	연세대학교 응용통계학과
	김 기 환	고려대학교 빅데이터사이언스학부
	김 용 대	서울대학교 통계학과
	이 영 섭	동국대학교 통계학과
	윤 성 민	부산대학교 경제학부
	박 찬 근	한국해양대학교 데이터사이언스전공
	박 승 열	(주) 케이스탯리서치 회장
	조 형 준	고려대학교 통계학과
	홍 재 범	부경대학교 경영학부
위원	이 성 건	성신여자대학교 수리통계데이터사이언스학부
	최 호 식	서울시립대학교 도시빅데이터융합학과
	신 승 준	고려대학교 통계학과
	진 서 훈	고려대학교 빅데이터사이언스학부
	허 태 영	충북대학교 정보통계학과
	이 동 희	경기대학교 경영학부
편집위원장	강 현 철	호서대학교 빅데이터AI학부
자문위원	최 종 후	고려대학교 응용통계학과
	조 완 현	전남대학교 통계학과
	박 희 창	창원대학교 통계학과
	강 창 완	동의대학교 산업경영·빅데이터공학과

### 운영위원회

공동위원장	최 호 식	서울시립대학교 도시빅데이터융합학과
	신 승 준	고려대학교 통계학과
부위원장	이 성 건	성신여자대학교 수리통계데이터사이언스학부
	김 지 수	가천대학교 간호학과
	제 상 영	고려대학교 경제통계학과
	조 장 식	경성대학교 정보통계학과
	안 수 한	서울시립대학교 통계학과
위원	김 지 희	강원대학교 응급구조학과
	김 태 훈	경성대학교 경제금융물류학부
	김 성 환	건국대학교 응용통계학과
	김 양 진	숙명여자대학교 통계학과
	곽 일 업	중앙대학교 통계학과
	박 만 식	성신여자대학교 수리통계데이터사이언스학부
	김 은 석	(주) GDS 컨설팅그룹 대표이사
	김 지 연	(주) 케이스탯리서치 대표이사

# 한국자료분석학회

## 2024년도 하계 학술논문발표대회

- 일시 : 2024년 6월 27일(목) ~ 6월 28일(금)
- 장소 : 서울시립대학교
- 공동주최 : 한국자료분석학회,  
서울시립대학교(통계학과, 데이터사이언스 융합인재양성사업단)
- 주관 : 서울시립대학교 통계학과
- 후원 : 한국연구재단, (주)케이스탯리서치, 트랜드리서치, (주)GDS컨설팅,  
데이터솔루션, 나이스평가정보, 서울시립대학교

### ● 6월 27일 목요일

17:00 - 19:00 편집위원회

17:00 - 19:00 이사회

### ● 6월 28일 금요일

09:00 - 18:00 100주년기념관 / 포스터전시(9:00 - 16:30, 국제회의장)  
포스터발표(15:10 - 16:30, 국제회의장)

09:30 - 10:50 논문발표 A1 - A4(308, 307, 나501, 305)

11:10 - 12:10 튜토리얼(국제회의장)  
좌장: 최호식 교수(학술이사, 서울시립대학교)  
제목 : 딥러닝 방법론의 이해 - CNN, RNN, Transformer를 중심으로  
강연자 : 이성건 교수(성신여자대학교 수리통계데이터사이언스학부)

12:10 - 12:50 총회(국제회의장)  
진행: 이성건 교수(총무이사, 성신여자대학교)

12:50 - 14:00 중식시간(자연과학관 식당)

14:00 - 15:00 학회장 초청 강연(국제회의장)  
좌장: 한상태 교수(학회장, 호서대학교)  
제목 : 2024년 한국 청년의 삶, 마음, 행복  
강연자 : 김석호 교수(서울대학교 사회학과)

15:10 - 16:30 논문발표 및 포스터발표 B1 - B5(308, 307, 나501, 305, 국제회의장)

16:40 - 18:00 논문발표 C1 - C5(308, 307, 나501, 305, 국제회의장)

**논문발표 A: 09:30 - 10:50**

**논문발표 A1(308호)**

**Data Science PLUS 기획세션 1**

좌장 : 신승준(고려대학교)

09:30   10:50	<p>1. The R package psvmSDR: A Unified Algorithm for Sufficient Dimension Reduction via Principal Machines ..... 1 신정민, 신승준*(고려대학교 통계학과)</p> <p>2. Group rank for encrypted data ..... 3 최호식*(서울시립대학교 도시빅데이터융합학과)</p> <p>3. Proformer: a Hybrid Macaron Transformer Model Predicts Expression Values from Promoter Sequences ..... 5 곽일엽*(중앙대학교 응용통계학과), 김병찬, 강태인(중앙대학교 통계데이터사이언스학과), 이주현(중앙대학교 프랑수어문학부), Daniel J. Garry, Wuming Gong(Department of Medicine, Lillehei Heart Institute, University of Minnesota), Jianyi Zhang(Department of Biomedical Engineering, University of Alabama at Birmingham)</p> <p>4. Studies of generative models for synthetic tabular data ..... 7 전종준*(서울시립대학교 통계학과)</p>
---------------------	---

**논문발표 A2(307호)**

통계

좌장 : 조형준(고려대학교)

09:30   10:50	<p>1. A Deep Learning Approach for Dimension Reduction on High-dimensional Data ..... 9 문영호*, 이영섭(동국대학교 통계학과)</p> <p>2. Change point test in the slope of the linear trend in time series: using adjusted range for self-normalization ..... 11 이두형*, 김경희(고려대학교 통계학과)</p> <p>3. Variable Selection in Semiparametric Models with Categorical Predictors ..... 13 황유진*, 송준(고려대학교 통계학과)</p> <p>4. ECLayr: Fast and Robust Topological Layer based on Differentiable Euler Characteristic Curve ..... 15 이하진*, 김광호(고려대학교 통계학과), 김지수(서울대학교 통계학과)</p>
---------------------	--

**논문발표 A3(나501호)**

**다양한 측면에서 본  
공공데이터의 활용**

좌장 : 한상태(호서대학교)

---

09:30	1. 국회의원 선거 결과에 대한 민감도 분석 .....	17
10:50	2. 서울시 통계혁신 방안에 관한 연구 .....	19
	한상태*(호서대학교 빅데이터AI학부), 서우석(서울시립대학교 도시사회학과), 강기훈(한국외국어대학교 통계학과)	
	3. 서울시 전세보증금 미반환(역전세) 위험에 관한 실증연구 .....	21
	김진우*, 최호식(서울시립대학교 도시빅데이터융합학과)	
	4. 설명 가능한 기계 학습 모형과 다변량 모형을 활용한 지역 경제와 부동산 시장 지표가 출산율에 미치는 영향의 비교 분석 .....	25
	최인수*(KAIST 산업및시스템공학과), 안수아(동국대학교 산업시스템공학과), 심은(서강대학교 컴퓨터공학과), 최연아(동국대학교 산업시스템공학과), 임소영 (경희대학교 산업경영공학과), 김우창 (KAIST 산업및시스템공학과)	

**논문발표 A4(305호)**

**언어모형응용**

좌장 : 김성용(호서대학교)

---

09:30	1. 비교 기술어 생성을 통한 시각적 분류 정확도 향상 .....	27
10:50	2. 거대언어모델 튜닝 기법에 따른 영어학습자료 문항 유형 다중 분류 성능 비교 .....	29
	김태경*, 정재경, 최호식(서울시립대학교, 도시빅데이터융합학과)	
	3. EconFinKoSolar-10.7B: 경제 및 금융 도메인에 특화된 한국어 거대언어모형 개발 .....	31
	홍소은*, 김수연, 김재영, 김효원, 남재윤, 박대민, 박종현, 윤수민, 이정민(서강대학교 일반대학원 AI 금융 전공 석사과정), 정재식(서강대학교 경제학)	
	4. 생활환경안전정보시스템 데이터베이스 설계 및 QA 파이프라인 구축 .....	33
	고재현*, 이다운, 정재경, 최호식(서울시립대학교 도시빅데이터융합학과)	



**논문발표 B3(나501호)**

**의료/간호/보건**

좌장 : 전은미(배재대학교)

---

15:10   16:30	1. 노인의 성별에 따른 골관절염 유병 관련 요인 ..... 53 박공주(김해대학교 간호학과), 김민영(경남정보대학교 간호학과), 전은미*(배재대학교 간호학과) 2. 국내 의료서비스 취약지역 대상 의료공급 개선방안 연구: 통계청 및 국민건강보험 공단 자료를 활용하여 ..... 55 정수진*(경희대학교 경영학과), 양성병(경희대학교 경영학과/빅데이터응용학과) 3. Eye movement data for Mild Cognitive Impairment diagnosis using functional data analysis ..... 57 김학준*, 이근범, 이상인(충남대학교 통계데이터사이언스학과), 구본초, 김중일, 김재욱(한국한의학연구원 디지털임상연구부) 4. Privacy-Preserving Mortality Predictive Model through Homomorphic Encryption-Enhanced Federated Learning ..... 59 이가람*, 박수영, 어수행, 김설기(크립토크 헬스케어AI실), 이상욱(서울아산병원 마취통증의학과), 서준교(서울아산병원 비뇨의학과)
---------------------	---

**논문발표 B4(305호)**

**경제/경영**

좌장 : 김희영(고려대학교)

---

15:10   16:30	1. Factors affecting the adoption of adaptation measures to climate change. A case study in Huong Phong Commune, Huong Tra Town, Thua Thien Hue Province, Vietnam ..... 61 Thanh Hoa Bui*, Keunjae Lee(Pusan National University, Department of Economics) 2. Select priority tourist sites in Vietnam's coastal tourist areas that satisfy green criteria: A research using Fuzzy AHP and Fuzzy TOPSIS ..... 63 Nguyen Tuan Dat*, Keunjae Lee(Department of Economics, Pusan National University) 3. A Consumption-Based Identification of Global Economic Uncertainty ..... 65 이은희*(경상국립대학교 경제학부) 4. Application of the Gaussian mixture model - hidden Markov model for financial time series ..... 67 기석주*, 조용진, 김희영(고려대학교 빅데이터 사이언스)
---------------------	---

**논문발표 C: 16:40 - 18:00**

**논문발표 C1(308호)**

경제/경영/딥러닝 응용

좌장 : 홍성철(서울시립대학교)

16:40   18:00	<p>1. Revisit Forward Rate Unbiasedness Hypothesis: Smooth Time-Varying Parameter Approach ..... 69  <span style="float: right;">Xiaoyu Kang*(중앙대학교)</span></p> <p>2. Dynamic dependence and risk spillover between oil price shocks and nonferrous metals. .... 71  <span style="float: right;">Meiling Jin*, Sung Y. Park(중앙대학교, 경제학과)</span></p> <p>3. Cryptocurrency Price Forecasting using Variational AutoEncoder with Versatile Quantile Modeling ..... 73  <span style="float: right;">홍성철*, 진종준(서울시립대학교 통계학과), 안승환(서울시립대학교 통계데이터사이언스학과)</span></p> <p>4. Deep fake detection system for ADD 2023 challenge ..... 75  <span style="float: right;">Soyul Han(Seoul National University), Taemin Kang(Chung-Ang University), Sunmook Choi(Korea University), Jaejin Seo(Chung-Ang University), Sanghyeok Chung(Korea University), Sumi Lee(Chung-Ang University), Seungsang Oh(Korea University), Il-Youp Kwak(Chung-Ang University)</span></p>
---------------------	--

**논문발표 C2(307호)**

비정형데이터분석

좌장 : 광일엽(중앙대학교)

16:40   18:00	<p>1. 벨 모양 필터를 이용한 오디오 데이터 증강 기법 연구 ..... 77  <span style="float: right;">서재진*(중앙대학교 통계데이터사이언스학과), 광일엽(중앙대학교 응용통계학과)</span></p> <p>2. 기계학습 기반 학교 배치 형태에 따른 실내외 소음도의 예측에 관한 연구 ..... 79  <span style="float: right;">김송옥*, 장서일(서울시립대 환경공학부)</span></p> <p>3. 국내 열차 차종 분류 신경망 모형의 입력데이터 구성에 대한 연구 ..... 81  <span style="float: right;">김민중*, 장서일(서울시립대 환경공학부)</span></p> <p>4. SEER 데이터를 이용한 다양한 딥러닝 생존분석 모형들의 비교 연구 ..... 83  <span style="float: right;">박근호*(가톨릭대학교 의생명건강과학과 의학통계전공), 민은정(가톨릭대학교 의과대학 의생명과학교실)</span></p>
---------------------	---

**논문발표 C3(나501호)**

경제/경영

좌장 : 오주현(가천대학교)

---

16:40	1. 엔트로피 지수를 통한 문화재정지출의 불균등도 분석 .....	85
18:00	오주현*(가천대학교 경영학부)	
	2. 주요 에너지 선물가격과 국내 업종별 주가지수 간의 연결성 분석 .....	87
	임지훈*, 박성용(중앙대학교 경제학과)	
	3. 거시경제지표를 이용한 이산화탄소 배출량 예측 .....	93
	김민성*, 박성용(중앙대학교 경제학부)	
	4. 실시간 해운 항만 데이터와 동적요인모형을 이용한 GDP Nowcasting .....	95
	강경태*, 조용복(동아대학교 경영정보학과)	

**논문발표 C4(305호)**

물리/통계

좌장 : 최호식(서울시립대학교)

---

16:40	1. Aging study on the Triple-GEM GE2/1 M2 detector .....	97
	정용호*, 박인규, 이상훈, 강다영(시립대학교 물리학과), 김태정, 김동현(한양대학교	
18:00	물리학과), 양운기(서울대학교 물리학과)	
	2. CMS iRPC 업그레이드 완료 및 빠른 타이밍 가스 검출기 위한	
	신규 연구개발* .....	99
	강민호*, 조영민(한양대학교), 이경세(고려대학교 극한 핵물질 연구센터)	
	3. SRGM을 사용한 딥러닝 소프트웨어 신뢰성 모형에 관한 연구 .....	101
	공영선*(조선대학교 전산통계학과), 장인홍, 송광윤, 김윤수, 이다혜(조선대학교	
	컴퓨터통계학과)	

---

16:40   18:00	<p>1. A Case Study of Statistical Analysis on Autonomous Vehicles Accidents Data ..... 103          심기영, 이인규*, 양정아(호서대학교 데이터사이언스학과), 연구필(호서대학교 빅데이터AI학과)</p> <p>2. 자율주행차량 사고 원인 분석을 통한 안정성 제고 방안 ..... 105          김근아*, 이은학(호서대학교 데이터사이언스학과), 연구필(호서대학교 빅데이터AI학부)</p> <p>3. 주식투자 관련 유튜브 개인방송과 TV방송의 허위정보 판단을 위한 비교 및 분석: 텍스트, 음성, 이미지 데이터를 활용하여 ..... 107          조민지*(경희대학교 일반대학원 빅데이터응용학과), 양성병(경희대학교 경영대학 경영학과/빅데이터응용학과)</p> <p>4. 초분광영상을 활용한 대면적 문화유산 내 아까시나무 탐지 ..... 109          권산*, 정재경(서울시립대학교 도시빅데이터융합학과), 어수행((주)크립토탭), 유세종((주)지오스토리) 김충식, 이재용(한국전통문화대학교 전통조경학과), 최호식(서울시립대학교 도시빅데이터융합학과)</p>
---------------------	--

**포스터 발표(국제회의장)**

15:10-16:30

좌장 : 최호식(서울시립대학교)

1. 언어모델을 활용한 미국 대선 영향 사건까지의 시차 예측 ..... 111  
백우빈\*, 손형진(경희대학교 산업경영공학과), 최인수, 김우창(KAIST  
산업및시스템공학과), 김장호(고려대학교 기술경영전문대학원)
2. 프롬프트 엔지니어링을 활용한 LLM 속성 기반 감성 분석에 대한 연구 ..... 113  
손민녕\*(고려대학교 SW·AI융합대학원 빅데이터융합학과), 임희석(고려대학교 정보대학  
컴퓨터학과)
3. 경제뉴스 감성분석과 딥러닝 모델을 활용한 판매량 예측 연구 ..... 115  
박시은\*(고려대학교 SWAI 융합대학원 빅데이터융합학과)
4. A Comprehensive Analysis of Multi-hop Question Answering Datasets: Insights  
and Implications ..... 117  
신현서\*, 황원석(서울시립대학교 인공지능학과)
5. NASA 그랜트 데이터를 이용한 항공우주 연구토픽 분석 ..... 119  
노경란\*, 한혁(한국과학기술정보연구원 R&D투자전략연구센터)
6. 대전시 공공자전거의 효율적 운영을 위한 수요예측 ..... 121  
박정연\*(충남대학교 통계데이터사이언스학과), 이상인(충남대학교 정보통계학과)
7. 고등학교 학업성취에 영향을 미치는 요인 분석 및 예측 지표 개발  
: Propensity Score Matching과 인과추론 접근 ..... 123  
이창현\*(고려대학교 빅데이터융합학과)
8. 2단계 집락 조건부 Quatember 확률화응답모형 ..... 125  
이기성\*(우석대학교 아동사회복지학부), 홍기학(동신대학교 컴퓨터학과), 손창균(동국대학교  
빅데이터·응용통계학과), 박근화(한국문화관광연구원), 홍성준(마사회 말산업연구소)
9. 주가 예측을 위한 LSTM, MAMBA 모델의 하이퍼파라미터 최적화 예비 연구 ..... 129  
김희연\*, 양하늘, 정민서(성신여자대학교, AI융합학과)
10. ARIMAX를 이용한 COVID-19 시기의 파생 변수와국내 섹터 ETF의 가격 변동성 예측에  
대한 연구 ..... 133  
이태경\*, 박성수, 신경수, 이수인(경희대학교 산업경영공학과), 최인수, 김우창(KAIST  
산업및시스템공학과), 김장호(고려대학교 기술경영전문대학원)
11. 다층 퍼셉트론을 활용한 화학 독성 데이터 다중 레이블 분류 ..... 135  
오승현\*, 노영민(서울시립대학교 인공지능학과)
12. Association between Oxidative Balance Score and Health-related Quality of Life Instrument  
with 8 Items (HINT-8) in Adults Based on Data from the 2021 Korea National Health and  
Nutrition Examination Survey ..... 137  
성수빈\*, 김진흠(수원대학교 데이터과학부)

13. Analysis of the risk factors towards metabolic syndrome for Korean adults over 40 years old using long-term follow-up cohort data .....	139
김명준*, 김진흠(수원대학교 데이터과학부)	
14. 결합된 제2종 점진적 복합 중도절단 상황에서 지수분포에 대한 정확한 가능성도 추론 .....	141
황영은*, 김지현(국립금오공과대학교 수학과), 이경준(국립금오공과대학교 수리빅데이터학과)	
15. 신장 및 심혈관질환 예측 및 위험평가 모델 개발 .....	143
정상현*(국립금오공과대학교 수학과), 이경준(국립금오공과대학교 수리빅데이터학과)	
16. 점진적 중도절단 상황에서 적합도 통계량 개발 .....	145
조영찬*, 박성희(국립금오공과대학교 수학과), 이경준(국립금오공과대학교 수리빅데이터학과)	
17. 리그오브레전드(League of Legend) 게임유저 데이터를 이용한 승패예측 .....	147
조혜령*, 이도현, 손재우, 송국영(한서대학교 항공컴퓨터학과)	
18. 음원데이터를 활용한 주간 차트 특징 분석 및 예측 .....	151
이민주(한서대학교 공항행정학과), 이하은*(한서대학교 항공컴퓨터학과), 박준범(한서대학교 항공컴퓨터학과)	
19. 라만 분광법을 이용한 10 $\mu$ m 이하 미세플라스틱의 신속한 검출 및 분류를 위한 단일 선형 네트워크 분석 및 제안 .....	157
이동현*, 노영민(서울시립대학교 인공지능학과)	
20. A New Large-Scale Knowledge Graph for Link Prediction in Chemicals inducing Diseases .....	159
우경동*, 조소영, 신원형, 김기원(서울시립대학교 통계데이터사이언스학과), 김동현, 최진희(서울시립대학교 환경공학부), 진종준(서울시립대학교 통계학과)	
21. Direct Optimization Approach for Query Performance Prediction .....	161
정재환*, 진종준(서울시립대학교 통계학과)	
22. Clustering sleep patterns and exploring their correlation with hs-CRP Levels .....	163
임현정*, 김희진, 이은지(충남대학교 통계데이터사이언스학과), 양보람(충남대학교 약학과)	
23. 전기 노인과 후기 노인의 신체활동 제한 사유와 악력 정도의 비교 .....	165
노효련*(강원대학교 물리치료학과)	
24. 우리나라 여성의 골다공증 영향요인-50세 이상을 대상으로- .....	167
홍희정*(동원과학기술대학교 간호학과)	
25. 코로나19예방접종 영향 요인 분석 .....	171
김은경(군산대학교)	

- 26. 발목 관절의 통증과 가동범위 제한에 대한 붕대형 파라핀 적용 단일사례연구 ..... 173  
 노효련\*, 이경봉, 정우경, 정은지, 최희원, 김혜미, 남채립, 이나린, 이승현, 이유진, 길우진,  
 이수빈(강원대학교 물리치료학과)
- 27. 성별에 따른 직무유연성의 차이가 비자발적퇴직후 고용에 미치는 영향 ..... 177  
 정예은\*, 홍지훈(부산대학교 경제통상대학 경제학과)
- 28. Option-Ownership Alignment and Downside Risks of Multinationality ..... 179  
 Sangcheol Song\*(Department of Management, Saint Joseph's University)
- 29. NASA SBIR 지원과제의 Topic Code에 따른 메타정보를 활용한 토픽모델링 연구 ..... 183  
 한혁\*, 노경란(한국과학기술정보연구원)
- 30. Visual Analysis of Arctic Sea Ice Imagery ..... 185  
 박재성, 조윤서\*(시립대학교 통계데이터사이언스학과), 홍성철, 전종준(시립대학교 통계학과)
- 31. 강수량에 의한 미세먼지 변화 ..... 187  
 김도현\*, 김준현(한서대학교 항공컴퓨터전공), 문우형(한서대학교 문화재보존학과),  
 박준범(한서대학교 항공컴퓨터학과)
- 32. 머신러닝과 딥러닝 모형을 이용한 '신고' 배나무의 수액 흐름 예측에 관한 연구 ..... 193  
 최원기\*, 김민수(전남대학교 통계학과), 김동언(전남대학교 수학/통계학과)
- 33. 생육기간 기상 및 생산량, 수출량을 활용한 배추 가격 예측 ..... 195  
 김혜진\*(전남대학교 수학/통계학과), 김민수(전남대학교 통계학과)
- 34. 기계학습기반 모형의 토양수분 예측성능 비교평가 ..... 197  
 임호진\*(전남대학교 수학/통계학과), 김민수(전남대학교 통계학과)
- 35. Large Scale Road-Traffic Noise Mapping with AI Technology ..... 199  
 최성수\*(서울시립대학교 통계데이터사이언스학과), 박태호(한국환경연구원),  
 전종준(서울시립대학교 통계학과)
- 36. Bibliometric Analysis of Depression Research Among Young Adults in Korea: A Preliminary  
 Analysis ..... 201  
 김아름\*, 이하임, 정한울(용인세브란스병원 정신건강의학과), 김유경(이화여자대학교  
 심리학과), 안병훈(맥길대학교 의학 및 건강 과학과)

37. 일반고 학생의 학교참여 유형 탐색 .....	203
조은수*(서울대학교 사회복지학과)	
38. 문화 이벤트 관람과 만족도 및 향후 참여 의향 분석: 인천광역시를 중심으로 .....	205
김하니*(강남대학교 인공지능융합공학부)	
39. 중국유학생의 한국유학과 한국대학 선택에 대한 영향요인분석 .....	209
도몽이*, 양정호(성균관대학교 교육학과)	
40. 중국 공립 중학교 교사승진제도분석-중국 산동성 청도, 동영, 빈주 중심으로 .....	211
조개열*, 양정호(성균관대학교 교육학과)	
41. 중국의 온라인 교육 플랫폼 비교분석:신동방 온라인, 후지앙왕샤오, 요우다오 프리미엄 코스 중심으로 .....	213
손정옥*, 양정호(성균관대학교 교육학과)	
42. 중국 ‘쌍일류’ 대학의 세계 대학 순위 변화와 영향요인 분석 .....	215
단시기*, 양정호(성균관대학교 교육학과)	
43. 중국인 유학생에 대한 한국학생의 인식 및 태도에 관한 연구 .....	217
비시월*, 양정호(성균관대학교 교육학과)	
44. 중국 대학생의 페미니즘 의식에 미치는 영향에 대한 연구 .....	219
유예첸*, 양정호(성균관대학교 교육학과)	
45. 기술수용모델 적용을 위한 새로운 패러다임 .....	221
고서정*(한국항공대학교 경영학과)	

# 튜토리얼

(1) 발표자 : 이성건 (성신여자대학교 수리통계데이터사이언스학부)

(2) 제목 : 딥러닝 방법론의 이해 - CNN, RNN, Transformer를 중심으로

최근 인공지능(artificial intelligence)은 연구자들 뿐만 아니라 사회전반에 걸쳐 매우 뜨거운 주제이다. 인공지능의 개념을 간단하게 설명하면, 사람의 지능과 비슷한 역할을 수행하는 소프트웨어라고 할 수 있다. 기계학습(machine learning)으로 통칭되는 “데이터 기반”의 접근방식과 마찬가지로 인공지능 방법론에서 흔히 사용되는 딥러닝(deep learning) 또한 직접 데이터를 분석하고 학습하여 숨어 있는 규칙이나 패턴을 찾아 문제를 해결한다.

딥러닝은 동물의 신경세포를 흉내 낸 퍼셉트론은 단위로 하여 구성된 신경망(neural network) 구조를 이용해 학습을 수행하고 문제를 해결한다. 인공지능 초창기부터 신경망을 이용한 연구가 진행되었지만 오랜 기간 변방에 머물러 주목받지 못하였으나 2000년대에 들어서 여러 돌파구들이 만들어지고 2010년 무렵부터 성공적인 결과를 얻어내고 있다. 딥러닝은 이제 인공지능 분야의 대표 주자로서 활약하고 있으며, 본 튜토리얼에서는 이미지와 텍스트 데이터 분석에서 널리 활용되는 Convolutional Neural Network, Recurrent Neural Network, Transformer의 기본 방법론을 이해하는 데에 목적이 있다.

## 1. Deep Learning

역전파(back propagation), 경사하강법(gradient decent), 활성화함수(activation function), 학습율(learning rate), 모멘텀(momentaum), 가중치감퇴(weight decay), 드롭아웃(drop out) 등

## 2. Convolutional Neural Network

이미지(image) 데이터의 특성을 고려한 합성곱신경망 네트워크, Convolution의 개념, 필터/커널(filter/kernel), 풀링(pooling), Resnet, 전이학습(transfer learning)

## 3. Recurrent Neural Network

텍스트 데이터의 특성을 고려한 순환신경망 네트워크, 자연어처리(Natural Language Processing), Word Embedding, RNN 구조, LSTM 등

## 4. Transformer

Encoder/Decoder, Attention, Self Attention, Query/Key/Value, BERT, GPT 등

## 초청강연

(1) 발표자 : 김석호 교수 (서울대학교 사회학과)

(2) 제목 : 2024년 한국 청년의 삶, 마음, 행복

-강연자료는 학회 당일 지급함.

## The R package psvmSDR: A Unified Algorithm for Sufficient Dimension Reduction via Principal Machines

*Jungmin Shin*<sup>1</sup>, *Seung Jun Shir*<sup>2</sup>

### Abstract

Dimension reduction is essential in statistics. Among many others, the sufficient dimension reduction that seeks a lower-dimensional subspace of predictors that contains regression/classification information has been popular in machine learning communities. In this work, we developed a psvmSDR R package that implements what we call the principal machine that covers a wide variety of SDR estimators stemming from the principal support vector machine (Li et al, 2011). The psvmSDR package implements the well-known gradient descent algorithm for the principal machines to efficiently compute the SDR estimators in various situations. We believe this easy-to-use package becomes an attractive alternative to the dr package, the only available SDR package in the communities.

*Keywords* : Gradient Descent Algorithm, Principal Support Vector Machine, Sufficient Dimension Reduction.

---

<sup>1</sup>02841 서울특별시 성북구 안암로 145 고려대학교 통계학과 박사수료 E-mail: @korea.ac.kr

<sup>2</sup>02841 서울특별시 성북구 안암로 145 고려대학교 통계학과 부교수. E-mail: sjshin@korea.ac.kr

## Group rank for encrypted data<sup>\*</sup>

*Hosik Choi<sup>†</sup>*

### Abstract

Recently, there has been an increasing demand for privacy-preserving techniques in numerous machine learning algorithms, elevating it to a critical concern. One promising solution involves the application of homomorphic encryption (HE). This study focuses on obtaining statistics based on the ranks of HE-encrypted data as a vital tool for robust data analysis. However, computing ranks in HE comes with significant computational costs due to the necessity of comparison operations, and there is currently no efficient method available. To address this gap, we propose an approximate rank method that exploits pairwise comparisons of data to derive ranks for encrypted information. This method effectively measures the association between two-dimensional ranks. Specifically, by utilizing approximate ranks of two variables, we estimate Spearman rank correlation without relying on perfect sorting and introduce a technique to reduce the number of required comparisons. Numerical experiments have been conducted to validate our approach, demonstrating that the disparity in values between rank correlation and approximate rank correlation is not substantial. Notably, the processing of one block comprising 32,768 ciphertexts took approximately one minute, exhibiting observed linear complexity dependent on the number of blocks.

*Keywords* : Privacy enhancing technology, homomorphic encryption, rank, quantile,

---

<sup>\*</sup>Hosik Choi was supported by the Basic Science Research Program through the NRF funded by the Ministry of Education (2017R1D1A1B05028565).

<sup>†</sup>Professor, Graduate School, Department of Urban Big Data Convergence, University of Seoul, Seoul, S. Korea. E-mail: choi.hosik@uos.ac.kr

## Proformer: a Hybrid Macaron Transformer Model Predicts Expression Values from Promoter Sequences

*Il Youp Kwak<sup>1</sup>, Byeong Chan Kim<sup>2</sup>, Juhyun Lee<sup>3</sup>, Taerin Kang<sup>4</sup>,  
Daniel J. Garry<sup>5</sup>, Jianyi Zhang<sup>6</sup>, Wuming Gong<sup>7</sup>*

### Abstract

The breakthrough high-throughput measurement of the cis-regulatory activity of millions of randomly generated promoters provides an unprecedented opportunity to systematically decode the cis-regulatory logic that determines the expression values. We developed an end-to-end transformer encoder architecture named Proformer to predict the expression values from DNA sequences. Proformer used a Macaronlike Transformer encoder architecture, where two half-step feed forward (FFN) layers were placed at the beginning and the end of each encoder block, and a separable 1D convolution layer was inserted after the first FFN layer and in front of the multihead attention layer. The sliding k-mers from one-hot encoded sequences were mapped onto a continuous embedding, combined with the learned positional embedding and strand embedding (forward strand vs. reverse complemented strand) as the sequence input. Moreover, Proformer introduced multiple expression heads with mask filling to prevent the transformer models from collapsing when training on relatively small amount of data. We empirically determined that this design had significantly better performance than the conventional design such as using the global pooling layer as the output layer for the regression task. These analyses support the notion that Proformer provides a novel method of learning and enhances our understanding of how cis-regulatory sequences determine the expression values.

*Keywords:* Transformer, Expression prediction, Sequence model, Deep Learning.

---

<sup>1</sup>(Corresponding Author) Associate Professor, Department of Applied Statistics, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea. Email: ikwak2@cau.ac.kr

<sup>2</sup>Graduate Student, Department of Statistics and Data Science, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea.

<sup>3</sup>Undergraduate Intern, Department of Statistics and Data Science, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea.

<sup>4</sup>Graduate Student, Department of Statistics and Data Science, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea.

<sup>5</sup>Professor, Cardiovascular Division, Department of Medicine, Lillehei Heart Institute, University of Minnesota, 2231 6th St SE, Minneapolis, MN 55455, USA.

<sup>6</sup>Professor, Department of Biomedical Engineering, The University of Alabama at Birmingham, Birmingham, AL 35233, USA.

<sup>7</sup>Assistant Professor, Cardiovascular Division, Department of Medicine, Lillehei Heart Institute, University of Minnesota, 2231 6th St SE, Minneapolis, MN 55455, USA.

## Studies of generative models for synthetic tabular data

*Jong-June Jeon*<sup>1</sup>

### AbstractAbstract

The Gaussian assumption has been consistently criticized as a main limitation of the Variational Autoencoder (VAE) despite its efficiency in computational modeling. In this paper, we propose a new approach that expands the model capacity (i.e., expressive power of distributional family) without sacrificing the computational advantages of the VAE framework. Our VAE model's decoder is composed of an infinite mixture of asymmetric Laplace distribution, which possesses general distribution fitting capabilities for continuous variables. Our model is represented by a special form of a nonparametric M-estimator for estimating general quantile functions, and we theoretically establish the relevance between the proposed model and quantile estimation. We apply the proposed model to synthetic data generation, and particularly, our model demonstrates superiority in easily adjusting the level of data privacy.

*Keywords* : Variational Autoencoder, Quantile Estimation, Synthetic Data Generation.

---

<sup>1</sup>02504 Professor, Department of Statistics, University of Seoul, Seoulsiripdaero 163, Korea.  
E-mail: jj.jeon@uos.ac.kr

## A Deep Learning Approach for Dimension Reduction on High-dimensional Data

*Youngho Moon<sup>1</sup>, Yung-Seop Lee<sup>2</sup>*

### Abstract

High-dimensional data analysis presents a significant challenge in data science due to its complex statistical properties. In high-dimensional spaces, the curse of dimension poses a problem, where data points become increasingly sparse and evenly distributed making Euclidean distance and nearest neighbor methods less effective. Thus, classical statistical methods are often unsuitable for analyzing high-dimensional data. To address this challenge, numerous dimension reduction(DR) techniques have been proposed. However, most DR techniques often fail to show good class separability in the low-dimensional spaces. We propose DR methods(both supervised and unsupervised autoencoders) to improve class separability. These methods aim to maximize the variance between different classes while minimizing the variance within each class. This leads to features that are better separated and easier to distinguish. Extensive experiments demonstrate that our proposed approach outperforms state-of-the-art existing DR methods.

*Keywords* : dimension reduction, manifold learning, feature extraction, autoencoder

---

<sup>1</sup>Department of Statistics, Dongguk University, 30, Phildong-ro 1-gil, Jung-gu, Seoul, 04620, Korea.

<sup>2</sup>Department of Statistics, Dongguk University, 30, Phildong-ro 1-gil, Jung-gu, Seoul, 04620, Korea.

## Change Point Test in the Slope of the Linear Trend in Time Series: Using Adjusted Range for Self-normalization<sup>\*</sup>

*Duhyeong Lee<sup>1</sup>, Kyoung Hee Kim<sup>2</sup>*

### Abstract

To test for a change-point of a linear trend in a trend stationary process, a self-normalization (SN) method is used to eliminate the nuisance parameters such as long-run variance in testing statistics. A well-known SN method in time series data uses the variance of a partial sum as a self-normalizer (see e.g. Jiang et al., 2023). Assuming that the error term follows an AR(1) model, the testing power tends to decrease when the autocorrelation term increases to 1. To handle this problem, we suggest a new SN method based on the adjusted range of a partial sum. The proposed SN method uses only one trimming parameter, which provides analytical benefits as well as practical advantages compared to existing methods that require two trimming parameters. Some asymptotic properties of our method are presented, and then through Monte Carlo simulation, we confirm that the power of the suggested testing maintains as autocorrelation approaches 1.

*Keywords:* Change point, Self-normalization, CUSUM process, Trend-stationary process

---

<sup>\*</sup>This work was in part supported by the National Research Foundation of Korea (NRF) funded by the Korea government (MSIT) (No. 2023R1A2C1003730 and No. RS-2023-00219212).

<sup>1</sup>Graduate student, Department of Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea. E-mail: dhlee1125@korea.ac.kr

<sup>2</sup>Corresponding author: Associate professor, Department of Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea. E-mail: arlenet@korea.ac.kr

## Variable Selection in Semiparametric Models with Categorical Predictors

*Yujin Hwang<sup>1</sup>, Jun Song<sup>2</sup>*

### Abstract

This research addresses the complexities of high-dimensional datasets that integrate both numerical and categorical variables by developing theories and methods for variable selection in semiparametric models with categorical predictors. Our methodology employs b-splines to effectively capture and handle potential nonlinearity between the numerical feature variables and the response variable. We use adaptive group lasso for variable selection in both the parametric and nonparametric components. The methodology's robustness is enhanced by non-asymptotic error bounds, ensuring its validity in finite sample scenarios. Extensive simulations and real data analyses demonstrate the efficacy of our approach. This work provides a robust tool for uncovering intricate relationships in semiparametric models, enhancing the approach to variable selection in high-dimensional environments.

*Keywords* : adaptive group lasso, high-dimensional analysis, semiparametric approach, variable selection

---

<sup>1</sup>Graduate student, Department of Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: yjh32@korea.ac.kr

<sup>2</sup>(Corresponding Author) Associate Professor, Department of Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: junsong@korea.ac.kr

## ECLayr: Fast and Robust Topological Layer based on Differentiable Euler Characteristic Curve

*Hajin Lee<sup>1</sup>, Jisu Kim<sup>2</sup>, Kwangho Kim<sup>3</sup>*

### Abstract

In the realm of Topological Data Analysis, persistent homology has traditionally served as a primary tool for extracting topological features. However, approaches relying on persistent homology often encounter challenges due to their high computational costs. To address this issue, we propose a computationally efficient novel topological layer tailored for general deep learning architectures, leveraging the Euler Characteristic Curve (ECC). Unlike methods based on persistent homology, ECC offers computational advantages by circumventing the need for persistent homology calculation, while still allowing access to crucial information about the underlying topological structure. The proposed layer can readily adapt to diverse data modalities by allowing appropriate filtration according to the user's preference, enabling its application across various learning problems without data preprocessing. We show that the proposed layer ensures stable backpropagation, facilitating its seamless integration into a variety of deep learning models. We go on to present stability analysis, showing that the proposed layer is robust against noise and outliers. We apply our method to topological autoencoders, showing that the standard loss function can effectively regularize topological structures of the latent space. Through classification experiments across various datasets, we illustrate the benefits of our approach in mitigating information loss under conditions of data scarcity or data contamination.

*Keywords* : Topological Data Analysis, Euler Characteristic Curve, Deep Learning, Differentiability

---

<sup>1</sup>02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 석사과정. E-mail: hlee2745@korea.ac.kr

<sup>2</sup>08826 서울시 관악구 관악로1, 서울대학교 통계학과 조교수. E-mail: jkim82133@snu.ac.kr

<sup>3</sup>(교신저자) 02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 조교수.

E-mail: kwanghk@korea.ac.kr

## 국회의원 선거 결과에 대한 민감도 분석\*

임요한<sup>1</sup>, 박민규<sup>2</sup>

### 요약

출구조사를 이용한 정당의 예상 의석수 산출 시, 조사 결과에서 발생하고 있는 체계적인 오차를 고려하는 방안에 대한 연구가 필요하다. 민감도 분석(sensitivity analysis)은 검증할 수 없는 가정이 어느 정도로 위배될 경우 결론이 어떻게 달라지는지를 분석하는 방법으로, 본 연구는 예상 의석수 산출 시 민감도 분석을 활용하는 방안을 모색하기 위한 기초 연구로서 진행되었다. 본 연구에서는 제20-22대 국회의원 선거에서 출구조사 결과와 실제득표 결과의 차이를 분석하였고, 특히 제22대 선거에서는 55개 지역구를 대상으로 사전투표에 대한 전화조사가 추가적으로 실시됨에 따라 사전투표와 본투표로 나누어서 결과를 비교하였다. 또한, 주변 민감도 모형(marginal sensitivity model)을 응용한 측도를 사용하여 출구조사 결과와 실제 결과 간의 오차가 어느 정도로 발생하였는지 살펴보았다. 분석 결과 제22대 국회의원 선거에서 열세자 침묵 효과가 존재한 것을 발견하였으며, 또한 사전투표에서는 더불어민주당이, 본투표에서는 국민의힘이 명확하게 우세한 경향이 있는 것을 확인하였다.

주요 용어 : 국회의원 선거, 출구조사, 민감도 분석, 열세자 침묵, 사전투표

<sup>1</sup>02841 서울특별시 성북구 안암로 145 고려대학교 정경관 402호, 고려대학교 통계학과 석박사통합과정.

E-mail: lyh2744@korea.ac.kr

<sup>2</sup>(교신저자) 02841 서울특별시 성북구 안암로 145 고려대학교 정경관 402호, 고려대학교 통계학과 교수.

E-mail: mpark2@korea.ac.kr

## 서울시 통계혁신 방안에 관한 연구<sup>2</sup>

한상태<sup>1</sup>, 서우석<sup>2</sup>, 강기훈<sup>3</sup>

### 요약

사회의 변화에 따라 시민들의 삶과 욕구가 분화되고, 이에 따라 시민들이 서울시에 요구하는 행정수요도 복잡·다단해지고 있다. 한편, 모바일 사용 인구의 증가와 사물인터넷 등 인터넷 기술의 발달로 과거에는 없었던 수많은 자료가 수집되고 있으며 이를 활용해 시민에게 더 많은 편익을 제공할 수 있는 기술적 토대가 마련되고 있다. 이러한 변화 속에서 서울시는 첫째, 지방자치체도의 정착에 따른 다양한 지역 정책 수요는 계속 증가하고 있으나, 정작 지역맞춤형 정책 수립과 이 과정에서 과학적인 의사 결정을 위한 지역 통계는 부족한 상황이고, 둘째, 그동안 통계 시스템 보급, 지역 통계의 확충 등 외형적인 인프라를 지속해 확충해 왔으나, 시민들의 생활을 직접 개선하고, 시민이 체감할 수 있는 생활통계, 소지역 단위 통계 등 내용적 콘텐츠는 여전히 부족한 상황이다. 셋째, 서울시는 내부적으로 그동안 통계조직의 축소('08년 13명->'19년 4명), 규모에 비해 적은 예산('19년 44억, 동경 '13년 기준 21.3억엔(245억 원), '13년 기준 베를린 4,560만 유로(673억 원) 등으로 전체적인 통계관리 체계가 부재한 상황이며, 넷째, 외부적으로는 4차 산업혁명 시대를 맞아 데이터 및 통계의 중요성이 더욱 두드러지고, 지역 통계에 관한 관심 증대, 지자체의 지역 통계 작성이 활성화 되어야 하는 상황에 직면했다. 본 연구는 이런 배경하에 서울시 정책과 연계한 통계 수요를 체계적으로 발굴하여 시정통계를 확충하고 시민 체감형 통계 개발로 통계활용도를 제고하고, 시정의 신뢰 향상 및 투명성 제고를 위해 서울시 통계가 직면한 환경을 정확하게 진단하고, 서울시 통계혁신 방안을 마련하여 서울시 정책의 과학화 및 선진화에 이바지하고자 한다.

주요 용어: 지역 통계, 생활통계, 통계 관리체계, 서울시 통계혁신 방안.

\*이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2022M3J6A1063595).

<sup>1</sup>31499 충청남도 아산시 배방읍 호서로79번길 20, 호서대학교 빅데이터AI학부 교수

<sup>2</sup>02504 서울시 동대문구 시립대로 163(전농동), 서울시립대학교 도시사회학과 교수

<sup>3</sup>17035 경기도 용인시 처인구 모현읍 외대로 81, 한국외국어대학교 통계학과 교수

## 서울시 전세보증금 미반환(역전세) 위험에 관한 실증연구\*

김진우<sup>1</sup>, 최호식<sup>2</sup>

### 요약

급작스러운 주택 전세/매매가격 하락은 주거유형을 2년 계약의 전세로 거주하고 있는 임차인 으로서는 전세보증금 반환에 대한 불안감을 가질 수 밖에 없다. 주택매매가격 하락정보와 함께 지역별, 유형별, 신축정도 등의 특성을 반영한 주택전세 실거래 데이터를 기반으로 전세금 반환 가능 여부의 기준인 전세가율에 대한 다양한 연구를 진행하였다. 첫째, 주택가격이 하락하게 되면 아파트와 연립다세대에서 전세가율이 하락하게 되어 민감하게 반응함을 확인하였다. 둘째, 신축 일수록 그리고 주택유형 중 연립다세대 및 오피스텔에서 전세금 미반환위험이 증가함을 알 수 있었다. 셋째, 서울시 내 행정동을 기준 한 지역별 분석에서는 아파트는 상대적으로 양호한 반면, 연립다세대 및 오피스텔은 상대적으로 위험도가 높고, 지역성을 띄는 것으로 확인되었다. 사회초년생 및 사회적 약자가 주로 거주하는 것으로 예상되는 연립다세대 및 오피스텔이 주택가격 하락에 대한 전세금 미반환위험이 높아짐이 확인됨에 따라 이에 대한 다양한 시스템과 정책이 필요함을 강조하고자 한다.

주요 용어: 전세보증금, 역전세, 전세 미반환 위험, 전세가격, 가격 변동.

### 1. 서론

최근 주택 전세가격과 매매가격이 하락함에 따라 주거유형을 전세로 거주하고 있는 임차가구의 경우 전세보증금 미반환 위험 현상이 나타나고 있다. 심각한 경우 전세계약 시 전달된 전세보증금 보다 주택 매매가격이 더 떨어지는 경우도 발생하는 데, 이를 ‘역전세’현상으로 말한다. 실제, 2022년 초까지 지속적으로 상승했던 주택가격은 이후 매매가격이 하락하면서, 다수의 지역에서 전세가격이 매매가격(시세)보다 높은 역전세현상이 발생하였다고 알려져 있다(국토연구원 2023).

이에 금리인상 및 경기침체 등이 함께 동반되면서 주택 전세가격은 장기적으로 위험을 동반하고 있다. 주택가격 최고점에 계약한 전세 계약은 보증금 미반환 위험에 노출되어 있으며, 뉴스 등 매스컴에서도 이와 같은 보도가 이어지고 있어, 전세보증금 미반환 위험에 대한 불안감은 높아지고 있다. 하지만 전세보증금 반환에 대한 위험성을 줄이기 위한 정보는 매우 부족한 것이 현실이다. 보증금 미반환에 리스크를 예방하기 위해서는 정부에서 매매거래에 대한 공시가격을 제시하는 것처럼 전세거래에 있어서도 적정 전세가격을 제시하여 계약자 스스로 확인 및 대응할 수 있는 계약 전 전세가격 적정 정보가 필요하다.

\*이 논문은 2023년 2차년도 서울시립대학교 대학혁신지원사업 참여 연구임.

<sup>1</sup>02504 서울시 동대문구 서울시립대로 163, 서울시립대학교 도시빅데이터융합학과 박사과정.

E-mail: oldjinu4@naver.com

<sup>2</sup>02504 서울시 동대문구 서울시립대로 163, 서울시립대학교 도시빅데이터융합학과 교수.

E-mail: choi.hosik@uos.ac.kr

## 2. 분석방법

이 연구에서 활용되는 자료는 국토교통부 실거래가 데이터를 기반 하였다. 투자자 또는 임대인 관점에서 제시되고 있는 전세가율을 전세계약기간 동안 거주하고 있는 임차인의 관점에서 수정하였다.

$$\text{전세가율 위험정도} = \frac{\text{전세가격}_{\text{계약시점}} \times \text{최고점 보정}}{\text{매매가격}_{\text{현재시점}} \times \text{최저점 보정}}$$

분석의 시간적 범위는 2021.12~2022.09에 전세계약 자료와, 일반적인 전세계약 기간인 2년 후를 반영하여 2023.01~2023.10의 매매계약 자료를 기준으로 하였다. 공간적 범위는 서울특별시를 기준으로 하였으며, 분석단위는 행정동을 기준으로 하였다. 분석방법은 paired-t-test, one-way anova를 활용하였다.

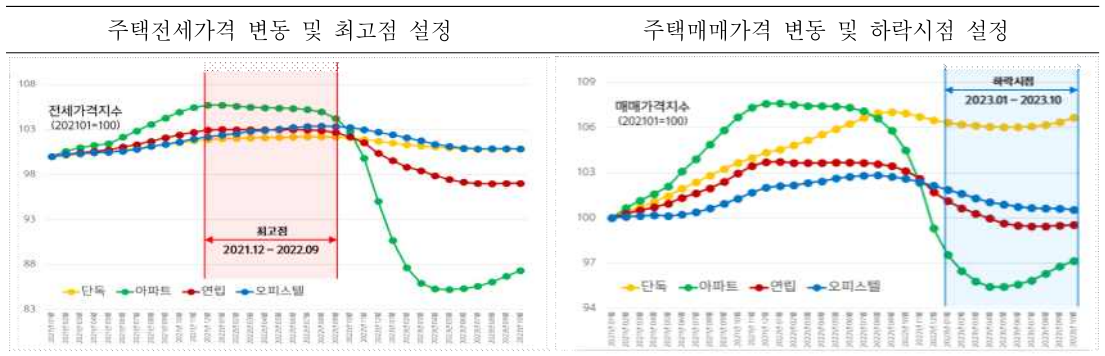


Table 1. basic statistics

구분	단독다가구		아파트		연립다세대		오피스텔	
	시점_가	가	시점_가	가	시점_가	가	시점_가	가
시점_가	최고점_전세가격	하락시점_매매가격	최고점_전세가격	하락시점_매매가격	최고점_전세가격	하락시점_매매가격	최고점_전세가격	하락시점_매매가격
N	338	177	320	309	321	290	246	233
최소값	100.300	212.200	192.200	290.740	255.300	323.600	259.550	318.290
최대값	936.200	2264.100	1502.700	3421.195	1444.600	3872.100	1869.472	2798.620
평균	395.963	658.021	734.919	1282.661	657.982	1022.762	789.046	832.094
표준오차	5.580	28.014	12.702	28.676	11.866	34.378	14.509	21.561
표준편차	102.592	372.700	227.223	504.085	212.593	585.431	227.561	329.111

## 3. 주택가격 하락에 따른 전세보증금 미반환 위험 분석

높은 추세의 2년전 주택 전세가격 기준 주택매매가격의 하락은 임차인에게 『전세보증금 미반환 위험』에 노출되어 있음을 예상할 수 있다. 아파트와 연립다세대의 전세가율은 전세/매매가격 하락에 대해 민감하게 반응함을 알 수 있다.

주택 신축 정도와 유형별 전세보증금 미반환(역전세) 위험 분석에서는 주택의 경과연수가 짧을 수록 전세가율이 증가함으로 전세보증금 미반환 위험은 증가하는 것으로 확인된다. 아파트는 타

주택유형에 비해 전세보증금 미반환위험은 낮으나, 연립다세대와 오피스텔의 전세보증금 미반환 위험은 상대적으로 높게 확인된다.

지역별 전세보증금 미반환(역전세)위험 정도 및 순위(상위10개)을 분석하였다. 서울 전 지역에서 아파트는 상대적으로 전세보증금 미반환 위험에 대해서 안정적인 것으로 확인되며, 높은 위험도를 보이는 것은 서남지역에 밀집되어 있는 연립다세대와 서울 전지역에 분포되어 있는 오피스텔로 확인된다.

Table 2. Sensitivity Results

구분 (전세가을)	평균	표준편차	평균의 표준오차	차이의 95% 신뢰구간		t	자유도	유의확률 (양쪽)
				하한	상한			
단독다가구	-.033	.272	.022	-.077	.011	-1.477	147	.142
아파트	.013	.104	.006	.001	.026	2.158	280	.032
연립다세대	.051	.230	.014	.024	.079	3.669	270	.000
오피스텔	-.017	.227	.016	-.048	.014	-1.080	206	.281

Table 3. Analysis of difference in jeonse rate according to construction progress

구분	N	평균	표준 편차	표준 오차	평균에 대한 95% 신뢰구간	
					하한값	상한값
~5년이하	476	.822	.230	.011	.801	.842
~10년이하	434	.792	.207	.010	.773	.812
~20년이하	589	.699	.211	.009	.682	.716
20년초과	676	.609	.221	.008	.592	.626
합계	2175	.717	.233	.005	.707	.726

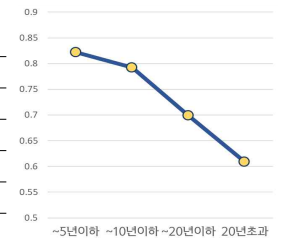


Table 4. Analysis of difference in jeonse rate by housing type

구분	N	평균	표준 편차	표준 오차	평균에 대한 95% 신뢰구간	
					하한값	상한값
단독다가구	227	.669	.256	.017	.635	.702
아파트	728	.634	.186	.007	.620	.647
연립다세대	805	.705	.225	.008	.689	.720
오피스텔	415	.911	.199	.010	.892	.930
합계	2175	.717	.233	.005	.707	.726

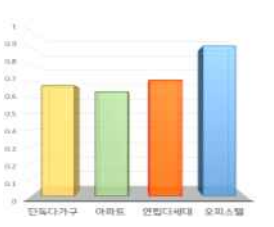


Table 5. Risk area for non-return of rental deposit in Seoul

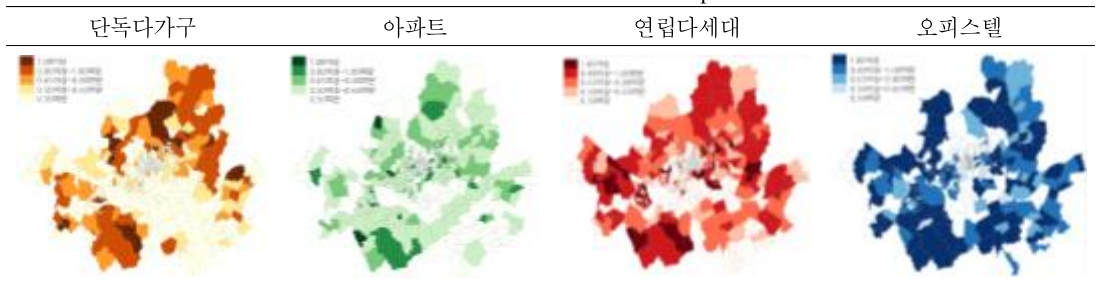


Table 6. Ranking of risk areas for non-return of rental deposit in Seoul

상 위	단독다가구			아파트			연립다세대			오피스텔		
	행정동	전세가을	행정구	행정동	전세가을	행정구	행정동	전세가을	행정구	행정동	전세가을	
1	영등포구 당산동1 가	1.405	은평구	갈현동	1.325	중구	남산동2 가	3.217	구로구	개봉동	2.428	
2	성북구	종암동	1.345	성북구	동선동3 가	1.236	성북구	보문동7 가	1.818	구로구	천왕동	2.278
3	성북구	상월곡동	1.331	영등포구	영등포동 3가	1.234	영등포구	영등포동 4가	1.564	서대문구	홍은동	2.155
4	강동구	암사동	1.254	영등포구	영등포동 4가	1.208	성북구	동소문동 7가	1.415	강북구	변동	1.598
5	구로구	오류동	1.239	동대문구	신설동	1.170	성북구	보문동6 가	1.398	서초구	양재동	1.593
6	종로구	명륜1가	1.178	영등포구	당산동6 가	1.143	종로구	신교동	1.341	도봉구	방학동	1.579
7	성북구	삼선동3 가	1.172	종로구	충신동	1.142	강서구	염창동	1.326	동대문구	신설동	1.548
8	영등포구	당산동2 가	1.149	중구	홍인동	1.099	영등포구	당산동3 가	1.265	은평구	녹번동	1.537
9	성북구	동소문동 6가	1.142	동작구	상도1동	1.046	성북구	동소문동 6가	1.237	영등포구	당산동4 가	1.527
10	영등포구	영등포동	1.136	서대문구	창천동	1.032	강서구	가양동	1.231	마포구	노고산동	1.527

급작스러운 주택가격 하락은 전세보증금 미반환위험을 가중시킴을 확인되었으며, 사회적 약자 또는 사회초년생(등)이 주로 거주하고 있는 연립다세대와 오피스텔에 집중되고 있어, 미반환에 대한 사회적 비용은 커질 수 밖에 없다. 보증금 미반환 리스크를 예방하기 위해서는 정부에서 매매 거래에 대한 공시가격을 제시하는 것처럼 전세거래에 있어서도 적정 전세가격을 제시하여 임차인 스스로 확인 및 대응할 수 있는 계약 전 전세가격 적정 정보와 이에 따른 정책과 시스템이 필요하다. 이런 사회적 문제를 해결하기 위해 AI기술을 반영한 혁신적인 제안이 대두된다.

대표적 예로 ML기술을 활용한 전세추정가<sup>Automated Valuation Model</sup>, DL기술의 전세 이상가격 탐지기술<sup>Anomaly Detection</sup>, 전세계약 이상거래탐지기술<sup>Fraud Detection System</sup> 등의 도입을 통해 전세보증금 미반환위험을 감소시킬 수 있도록 하여야 한다. 이를 위해서는 AI기술을 접목한데이터 정보 활성화 정책이 필요하며, 이를 위한 기술-법-제도-금융서비스(등)의 사회 전반의 합의가 필요해 보인다.

## References

- Min, B. C. (2021), Measuring the Risk of Jeonse Deposit Loss Using Jeonse Deposit Fluctuation Distribution, *KREAA*, 27(2), 63-75. (in Korea)
- Kim, S. W., Chung, K. S(2023), A Study on the Jeonse Deposit Loss in Rental Housing Market Using Panel Logit Model, *Korean Comparative Government Review*, 27(1), 119-132. (in Korea)
- Ahn, Y. B. (2022), What Makes Rental Housing Market Worse? Evidence from South Korea's Jeonse Price, *Journal of Market Economy*, 51(2), 1-18.(in Korea)
- NKIS (2019). A Study on the Phenomenon of Inverse Lease on a Deposit Basis and Consolidation of Protection for Lessees: Focused on Guarantee for a Refund of Jeonse Deposit, 107. (in Korean).

## 설명 가능한 기계 학습 모형과 다변량 모형을 활용한 지역 경제와 부동산 시장 지표가 출산율에 미치는 영향의 비교 분석

최인수<sup>1</sup>, 안수아<sup>2</sup>, 심은<sup>3</sup>, 임소영<sup>4</sup>, 최연아<sup>2</sup>, 김우창<sup>1</sup>

### 요 약

본 연구는 지역 경제와 부동산 시장 지표들이 출산율에 어떠한 영향을 미치는지 다차원 분석을 통해 조사하였다. 우리는 상관관계 분석과 회귀분석을 수행하여 출산율 예측에 중요한 변수들을 식별하고, SARIMAX(Seasonal Autoregressive Integrated Moving Average with eXternal variable)와 다양한 기계 학습 모형을 활용하여 예측 모델을 구축하였다. 모델 성능의 비교를 통해 랜덤 포레스트 모형이 본 연구의 데이터를 기준으로 가장 우수한 성능을 보이는 것으로 확인되었으며, 해당 모형을 바탕으로 특성 중요도를 추출하여 지역별 출산율에 영향을 미치는 주요 요인들을 분석하고 SARIMAX의 분석 결과와 비교하였다.

주요 용어: 지역 경제 변수, 부동산 시장 지표, 출산율, SARIMAX, 랜덤 포레스트, 설명 가능한 인공지능 기법.

---

<sup>1</sup>한국과학기술원 산업및시스템공학과

<sup>2</sup>동국대학교 산업시스템공학과

<sup>3</sup>서강대학교 컴퓨터공학과

<sup>4</sup>경희대학교 산업경영공학과

## 비교 기술어 생성을 통한 시각적 분류 정확도 향상

이한결<sup>1</sup>, 서가원<sup>2</sup>, 정지영<sup>3</sup>

### 요약

최근 연구에 따르면, 비전-언어 모델의 제로샷 분류 작업에서 클래스명만을 사용하는 방식보다 대형 언어 모델을 활용하여 생성한 추가적인 정보, 상위 개념, 심지어 무작위 문자를 포함한 기술어를 사용하는 것이 더 나은 성능을 보인다. 기존 연구의 결과를 분석하는 과정에서, 우리는 대부분의 오분류가 일부 유사한 클래스 사이에서 발생하는 현상을 발견했다. 이는 모델이 미세한 차이가 있는 클래스들을 구별하는 데 어려움을 겪으며, 종종 공통된 특성을 가진 클래스들을 혼동한다는 것을 의미한다. 이 문제를 해결하기 위해 우리는 '비교 기술어'를 생성하는 방법을 제안한다. 기존 방식이 특정 클래스에 대한 설명을 직접 생성하도록 언어 모델에게 지시하는 반면, 우리는 유사한 클래스를 식별한 후 이들과 목표 클래스와의 비교를 통해 서로 구별되는 특징을 강조하는 설명을 생성하도록 지시한다. 우리의 방식으로 생성한 비교 기술어를 사용하면, 동음이의어로 인한 모호성과 비전-언어 모델이 이해하지 못하는 비시각적 정보를 포함한 기술어를 생성하는 문제가 감소한다. 또한 유사한 클래스 간의 오분류가 감소하여 궁극적으로 시각적 분류의 정확도가 향상된다.

주요 용어 : 비교 기술어, 유사한 클래스, 비전-언어 모델, 대형 언어 모델.

---

<sup>1</sup>02504 대한민국 서울특별시 서울시립대로 163, 서울시립대학교 석사과정. E-mail: leehk@uos.ac.kr

<sup>2</sup>02504 대한민국 서울특별시 서울시립대로 163, 서울시립대학교 학부재학생. E-mail: gawon1224@uos.ac.kr

<sup>3</sup>(교신저자)02504 대한민국 서울특별시 서울시립대로 163 창공관 424호, 서울시립대학교 부교수.

E-mail: jyjung@uos.ac.kr

## 거대언어모델 튜닝 기법에 따른 영어학습자료 문항 유형 다중 분류 성능 비교\*

김태경<sup>1</sup>, 정재경<sup>2</sup>, 최호식<sup>3</sup>

### 요 약

고등학교 영어교육과정에서 요구하는 성취기준 달성정도는 문항풀이를 위한 주요개념과 원리에 맞게 세분화되어 평가된다. 따라서 영어학습자료의 문항유형을 분류하고 유형별 능력을 측정하여 학생들에게 제공하는 것은 학습 효과를 향상시키는 데 도움이 될 수 있다. 거대언어모델은 광범위한 언어 데이터와 문맥이해능력을 바탕으로 분류 문제를 해결할 수 있으며 튜닝 방법에 따라 성능도 다양하다. 본 연구에서는 거대언어모델의 튜닝 방법에 따른 문항 유형의 분류 성능을 비교해 보고자 한다. 분석에 활용한 데이터 셋은 문항을 풀기 위해 요구되는 스킬과 방법이 라벨링 된 다중 레이블 자료이다. 영어학습자료에는 영어와 한국어가 함께 등장하므로 다국어 거대언어모델을 활용하였다. 각 문항은 질문, 본문, 조건, 선지, 정답 등의 요소로 구성되어 있으며, 분류에 도움이 되는 정보는 질문 요소에 충분히 내재되어 있음을 가정하고 질문 요소만을 입력 데이터로 활용하였다. 일반적으로 사전 학습된 거대언어모델의 파라미터를 필요한 태스크에 맞추어 모두 업데이트하는 파인튜닝은 우수한 성능 대비 높은 계산량을 요구하고, 사전 학습된 파라미터는 고정된 채로 입력에 추가되는 프롬프트 텍스트의 파라미터를 업데이트하는 프롬프트 튜닝은 계산량이 매우 적다. 따라서 사전 학습된 거대언어모델의 학습 파라미터 개수 대비 성능을 비교하여 경량화된 튜닝 기법의 서비스 활용 가능성을 평가해 보고자 한다.

주요 용어 : 거대언어모델, 프롬프트 튜닝, 다중분류, 영어교육학습자료.

---

\*이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임  
(NO. 2022M3J6A1084845).

<sup>1</sup>서울특별시 동대문구 서울시립대로 163 건설공학관 3-237, 서울시립대학교 도시빅데이터 융합학과 석사과정.  
E-mail: tkdl0324@uos.ac.kr

<sup>2</sup>서울특별시 동대문구 서울시립대로 163 건설공학관 3-237, 서울시립대학교 도시빅데이터 융합학과 박사과정.  
E-mail: jaekyeong@uos.ac.kr

<sup>3</sup>(교신저자) 서울특별시 동대문구 서울시립대로 163 법학관 606호, 서울시립대학교 도시빅데이터 융합학과 교수.  
E-mail: choi.hosik@uos.ac.kr

## EconFinKoSolar-10.7B: 경제 및 금융 도메인에 특화된 한국어 거대언어모형 개발

홍소은<sup>1</sup>, 김수연<sup>2</sup>, 김재영<sup>3</sup>, 김효원<sup>4</sup>, 남재윤<sup>5</sup>, 박대민<sup>6</sup>, 박종현<sup>7</sup>, 윤수민<sup>8</sup>, 이정민<sup>9</sup>, 정재식<sup>10</sup>

### 요 약

본격적으로 AI 시대가 시작됨에 따라 ‘경제 분야에서 거대언어모형의 중요성은 시대적 요청’이라는 표현은 지남침이 없다. 본 연구는 경제 및 금융 분야에 최적화된 한국어 거대언어모형을 개발 과정에 대한 결과물을 기술하고 있다. 개발된 모형의 목적은 경제 및 금융 분야의 세부 도메인에 적용이 가능하도록 하고, 일반인들이 경제와 금융을 쉽게 접근하여 이해도를 높일 수 있는 확장 가능한 도구를 제공하는 데 있다. 한국어 사전학습 모형인 yanolja/KoSOLAR-10.7B-v0.2을 활용하였고, 학습 데이터로는 경제 및 금융에 특화된 공개 데이터를 사용하였다. 학습에는 SFT(Supervised Fine-Tuning), DPO(Direct Preference Optimization), Merge 방법이 적용되었다. 모형의 성능은 전통적인 평가지표인 BLEU, ROUGE, Accuracy, F1 score, Recall과 GPT-4를 프롬프팅(Prompting)하여 평가하였다. 그 결과, 본 연구에서 개발된 모형이 번역뿐만 아니라 경제학적 개념과 원리를 정확하게 이해하고 적용하는 능력에서 뛰어난 성능을 보임으로써 경제 및 금융 교육에 있어 한국어 거대언어모형의 잠재적 기회를 탐색함과 동시에, 비영어권 언어 처리의 고질적 문제인 번역 오류와 전문 용어의 정확한 문맥 이해에 대한 해결책을 제시하여 후속 연구의 방향성을 제시한다.

주요 용어 : 경제·금융 특화, 거대언어모형, SFT, DPO, GPT-4.

<sup>1</sup>제1저자, 서강대학교 일반대학원 AI 금융 전공 석사과정. E-mail: wendy0770@u.sogang.ac.kr

<sup>2</sup>서강대학교 일반대학원 AI 금융 전공 석사과정. E-mail: syosue@u.sogang.ac.kr

<sup>3</sup>서강대학교 일반대학원 AI 금융 전공 석사과정. E-mail: bluekim1234@u.sogang.ac.kr

<sup>4</sup>서강대학교 일반대학원 AI 금융 전공 석사과정. E-mail: gy1503@u.sogang.ac.kr

<sup>5</sup>서강대학교 일반대학원 AI 금융 전공 석사과정. E-mail: skawodbs03@u.sogang.ac.kr

<sup>6</sup>서강대학교 일반대학원 AI 금융 전공 석사과정. E-mail: qjtrogoa@u.sogang.ac.kr

<sup>7</sup>서강대학교 일반대학원 AI 금융 전공 석사과정. E-mail: hyuni1338@u.sogang.ac.kr

<sup>8</sup>서강대학교 일반대학원 AI 금융 전공 석사과정. E-mail: yoonsomin@u.sogang.ac.kr

<sup>9</sup>서강대학교 일반대학원 AI 금융 전공 석사과정. E-mail: jm980215@u.sogang.ac.kr

<sup>10</sup>(교신저자) 서강대학교 경제학부 교수. E-mail: cschung@sogang.ac.kr

## 생활환경안전정보시스템 데이터베이스 설계 및 QA 파이프라인 구축\*

고재현<sup>1</sup>, 이다운<sup>2</sup>, 정재경<sup>3</sup>, 최호식<sup>4</sup>

### 요약

생활환경안전정보시스템(이하 초록누리)은 일상생활에서 사용하는 제품들에 포함된 화학물질에 대한 종합적인 정보를 제공하지만, 제품 단위로 제공되기 때문에 개별 화학물질 분석에 어려움이 있다. 이에 본 연구에서는 사용자의 입력에 따라 정보를 유연하게 출력할 수 있는 관계형 데이터베이스 파이프라인을 구축하여 초록누리 정보를 효과적으로 탐색할 수 있는 시스템을 제안한다. 사용자가 검색할 수 있는 입력 형태는 단어와 문장 두 가지로, 정규 표현식을 통해 구분하였다. 단어 형태로 검색할 경우, SQL 쿼리 자동화 함수를 사용하여 카스 정보(CASRN, CAS-name, Synonyms), 생활화학제품 리스트(클래스별), 연관규칙 분석 결과를 출력한다. 문장 형태로 검색할 경우, 거대언어모델(LLM)을 활용한 Automated SQL Query Generation Feature를 통해 중요 키워드를 식별하고, 데이터베이스에서 정보를 검색하여 출력한다. 거대언어모델은 문장 내 특정 질문 내용을 이해하고, 이를 처리하는 SQL Query를 생성할 수 있기 때문에 평균값이나 개수 등과 같은 통계량 정보를 출력하는 것도 가능하다. 이 시스템은 사용자에게 더욱 직관적이고 유연한 데이터 접근 방식을 제공함으로써 화학물질 및 제품 분석에 필요한 정보를 손쉽게 얻을 수 있게 하며, 자체적으로 구축한 동의어와 연관규칙 분석 결과를 포함함으로써 더욱 포괄적이고 깊이 있는 분석을 지원한다.

주요 용어 : Relational database, Large language model, Text-to-sql, Chemical substance

\*이 논문은 환경부 생활화학제품 안전관리 개발사업(R&D)의 지원을 받아 진행된 논문입니다 (RS-2023-00215309).

<sup>1</sup>서울특별시 동대문구 서울시립대로 163 건설공학관 3-237, 서울시립대학교 도시빅데이터 융합학과 석사과정.  
E-mail: ko9805@uos.ac.kr

<sup>2</sup>서울특별시 동대문구 서울시립대로 163 건설공학관 3-237, 서울시립대학교 도시빅데이터 융합학과 석사과정.  
E-mail: dlekdns07@uos.ac.kr

<sup>3</sup>서울특별시 동대문구 서울시립대로 163 건설공학관 3-237, 서울시립대학교 도시빅데이터 융합학과 박사과정.  
E-mail: jaekyeong@uos.ac.kr

<sup>4</sup>(교신저자)서울특별시 동대문구 서울시립대로 163 법학관 606호, 서울시립대학교 도시빅데이터 융합학과 교수.  
E-mail: choi.hosik@uos.ac.kr

# Masked Language Modeling Becomes Conditional Density Estimation for Tabular Data Synthesis

*Seunghwan An<sup>1</sup>, Gyeongdong Woo<sup>1</sup>, Jaesung Lim<sup>1</sup>,  
ChangHyun Kim<sup>1</sup>, Sungchul Hong<sup>2</sup>, and Jong-June Jeon<sup>3</sup>*

## Abstract

In this paper, our goal is to generate synthetic data for heterogeneous (mixed-type) tabular datasets with high machine learning utility (MLu). Given that the MLu performance relies on accurately approximating the conditional distributions, we focus on devising a synthetic data generation method based on conditional distribution estimation. We propose a novel synthetic data generation method, MaCoDE, by redefining the multi-class classification task of Masked Language Modeling (MLM) as histogram-based non-parametric conditional density estimation. Our proposed method enables estimating conditional densities across arbitrary combinations of target and conditional variables. Furthermore, we demonstrate that our proposed method bridges the theoretical gap between distributional learning and MLM. To validate the effectiveness of our proposed model, we conduct synthetic data generation experiments on 10 real-world datasets. Given the analogy between predicting masked input tokens in MLM and missing data imputation, we also evaluate the performance of multiple imputations on incomplete datasets with various missing data mechanisms. Moreover, our proposed model offers the advantage of enabling adjustments to data privacy levels without requiring re-training.

*Keywords* : Masked Language Modeling, Synthetic Data Generation, Tabular Data, Distributional Learning.

---

<sup>1</sup>Department of the Statistical Data Science, University of Seoul 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Korea.

<sup>2</sup>Department of the Statistics, University of Seoul 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Korea.

<sup>3</sup>(Corresponding Author) Department of the Statistics, University of Seoul 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Korea. E-mail: jj.jeon@uos.ac.kr

# Computationally Efficient Sparse Sufficient Dimension Reduction via Least Squares Support Vector Machine and its Extensions

*Jungmin Shin<sup>1</sup>, Seung Jun Shin<sup>2</sup>*

## Abstract

Sufficient dimension reduction (SDR) is a class of supervised dimension reduction techniques targeting a feature extraction in a regression or classification context without losing information between the response and the predictors. However, sufficient predictor is a linear combination of all the original predictors, making it difficult to interpret the resulting estimates. Meanwhile, the sparse SDR has been introduced to enhance the interpretability of the central subspace, but it is often computationally intensive. To address these drawbacks, in this paper, we propose a computationally efficient sparse SDR method based on least square support vector machines (LSSVM), and we extend it under the principal logistic regression framework. The proposed method can be easily implemented by applying the group coordinate descent (GCD) algorithm and enjoys the oracle property in the variable selection context. Extensive simulation studies and real data analysis are carried out to illustrate the promising performance of the proposed method in terms of estimation accuracy and computational efficiency compared to existing methods in both regression and classification.

*Keywords* : sufficient dimension reduction, group coordinate descent, non-convex penalty, convex optimization.

---

<sup>1</sup>Ph.D Candidate, Department of the Statistics, Korea University, 145 Anam-ro, Seoungbuk-Gu, 02841, Seoul, Korea. E-mail: jungminshin@korea.ac.kr

<sup>2</sup>Associate Professor, Department of the Statistics, Korea University, 145 Anam-ro, Seoungbuk-Gu, 02841, Seoul, Korea. E-mail: sjshin@korea.ac.kr

## An Adaptive Importance Sampling for Locally Stable Point Processes<sup>\*</sup>

*Hee-Geon Kang*<sup>1</sup>, *Sunggon Kim*<sup>2</sup>

### Abstract

**Summary:** The problem of finding the expected value of a statistic of a locally stable point process in a bounded region is addressed. We propose an adaptive importance sampling for solving the problem. In our proposal, we restrict the importance point process to the family of homogeneous Poisson point processes, which enables us to generate quickly independent samples of the importance point process. The optimal intensity of the importance point process is found by applying the cross-entropy minimization method. In the proposed scheme, the expected value of the function and the optimal intensity are iteratively estimated in an adaptive manner. We show that the proposed estimator converges to the target value almost surely, and prove the asymptotic normality of it. We explain how to apply the proposed scheme to the estimation of the intensity of a stationary pairwise interaction point process. The performance of the proposed scheme is compared numerically with the Markov chain Monte Carlo simulation and the perfect sampling.

**Keywords:** locally stable point process, pairwise interaction point process, Monte Carlo method, adaptive importance sampling, cross entropy minimization.

---

<sup>\*</sup>This work was supported by the National Research Foundation of Korea(NRF) grant funded by the korea government(MSIT) (NO.2022M3J6A1084845).

<sup>1</sup>Department of statistics, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Korea.  
E-mail: hgkang28@uos.ac.kr

<sup>2</sup>Department of statistics, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Korea.  
E-mail: sgkim@uos.ac.kr

## ADMM for Broken Adaptive Ridge Penalized Quantile Regression

*Minji Sung*<sup>1</sup>, *Seohyeon Park*<sup>2</sup>, *Sangbum Chor*<sup>3</sup>

### Abstract

Sparse penalized quantile regression, while widely adopted for variable selection and robust estimation, often fails to account for group structure among highly correlated variables. To address this limitation, we introduce the  $L_0$ -based broken adaptive ridge (BAR) framework within quantile regression, which iteratively applies reweighted  $L_2$ -penalization. This approach approximates the advantages of  $L_0$ -penalization while avoiding its computational challenges, such as being NP-hard and not scalable. The BAR penalty enhances sparsity and identifies grouping information, resulting in a more stable and accurate model compared to  $L_1$ -type methods like LASSO. The optimization challenge posed by the nonsmoothness of the check loss function in quantile regression is addressed using the alternating direction method of multipliers (ADMM) for efficient convergence. Simulation studies demonstrate that BAR-ADMM performs better or on par with existing methods in terms of variable selection accuracy, parameter estimation, and the grouping effect. Additionally, its application to real data confirms the practical utility of our method.

*Keywords* : ADMM, Quantile Regression, Broken Adaptive Ridge, Grouping Effect, Penalized Feature Selection.

---

<sup>1</sup>Graduate Student, Department of Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: minji913@korea.ac.kr

<sup>2</sup>Graduate Student, Department of Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: seohyeon20@korea.ac.kr

<sup>3</sup>(Corresponding Author) Professor, Department of the Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: choisang@korea.ac.kr

## 데이터 불균형 문제를 고려한 변수 선택 기법 연구

박진수<sup>1</sup>, 조용복<sup>2</sup>

### 요약

변수 선택은 불필요하거나 관련성이 낮은 독립변수를 제거하는 데이터 전처리 과정으로, 분류 모형 구축 시 모형의 과적합 방지 및 해석 가능성 제고를 위해 반드시 선행해야 하는 절차이다. 효과적인 변수 선택을 위한 다양한 기법이 제안되고 있으나, 종속변수의 분포가 균형을 이루고 있는 상황을 가정하고 있는 경우가 많기 때문에 불균형 데이터를 다룰 경우 이는 최적의 변수 선택 기법이 아닐 수 있다. 본 연구에서는 데이터 불균형 문제가 존재할 때, 랜덤 포레스트 알고리즘을 통한 지니 불순도 기반 변수 선택 기법의 한계점을 지적한 뒤  $F_1$  Score 기반 변수 선택 기법을 활용할 것을 제안한다. 실증 분석 결과  $F_1$  Score 기반의 변수 선택 기법을 활용할 경우, 지니 불순도 기반의 변수 선택 기법과는 다른 변수가 선택되었으며 분류 모형의 성능 또한 개선되는 경향이 있음을 확인하였다. 이를 통해 변수 선택 기법을 활용할 때, 데이터 불균형 문제를 반드시 고려할 필요가 있음을 확인하였다.

주요 용어: 데이터 불균형, 변수 선택 기법, 랜덤 포레스트, 지니 불순도,  $F_1$  Score.

---

<sup>1</sup>49236 부산광역시 서구 구덕로 225, 동아대학교 경영정보학과 석사과정. E-mail: 2373458@donga.ac.kr

<sup>2</sup>(교신저자) 49236 부산광역시 서구 구덕로 225, 동아대학교 경영정보학과 조교수. E-mail: ybcho@dau.ac.kr

## 인접한 표본을 고려하기 위한 순차적 단계 분류\*

박소연<sup>1</sup>, 김민수<sup>2</sup>

### 요약

배나무 결과지 전정 작업은 고품질 과수를 수확하기 위해 매년 과수원에서 진행된다. 기존 결과지 전정 작업은 농민 및 원예 재배자의 주관적인 판단을 통해 주로 이루어졌다. 본 연구에서는 수집된 농가 데이터를 이용하여 전정된 결과지와 전정되지 않은 결과지의 특성을 가지고 새로운 전정 판별 모델 절차를 제안하고자 한다. 결과지 데이터의 주요 분포에서 전정 작업이 시행된 결과지의 평균 간격이 더 좁고, 밀집하여 분포하는 비율이 높음을 확인하였다. 이를 반영하기 위해 개별 결과지 특성을 이용한 Random Forest 1차 분류와 해당 결과지 및 근접 결과지의 1차 분류 점수를 활용한 Naive Bayes 2차 분류를 설정하였다. 결과지는 주지에 종속적으로 발생하므로, 주지별 상황과 전정 작업이 독립적이라고 보기 어려워 2차 분류를 통해 실제 환경에 부합한 분류 절차를 적용하고자 하였다. 분류 예측 결과 2차 분류에서 1차 분류 대비 정량적 성능(정확도, 정밀도, F1 점수, AUC(Area Under the Curve))이 향상되었다. 2차 분류에서 전정을 시행하지 않는 결과지의 예측 정확도가 향상되어 분류 모델이 개선되었다. 2차 분류에서 전반적으로 가장 영향력이 높았던 변수는 1차 분류 점수이다. 1차 분류 점수 0.45~0.55 구간에서는 2차 분류 시 전·후 결과지의 높은 1차 분류 점수와  $\pm 20\text{cm}$  구간 내 다수의 1차 제거 판정 결과지가 전정 판단에 주요하게 작용하였다. 향후 본 연구를 통해 주지 내 결과지 분포 특성을 고려한 연속적인 절차의 전정 분류가 배 과수 재배자의 실제 작업 조건에 효과적으로 활용될 수 있을 것으로 기대된다.

주요 용어 : 결과지 전정, Random Forest, Naive Bayes.

\*본 논문은 농촌진흥청 공동연구사업(과제번호: PJ015648) 지원에 의하여 연구되었음.

<sup>1</sup>61186 광주광역시 북구 용봉로 77, 전남대학교 수학과/통계학과 석·박사통합과정.

E-mail: syeonon1512@gmail.com

<sup>2</sup>(교신저자) 61186 광주광역시 북구 용봉로 77, 전남대학교 통계학과 교수. E-mail: kimms@chonnam.ac.kr

## 장르별 이질성을 고려한 공연예술 빅데이터 관람수요분포예측: 조건부 분위수회귀모형을 중심으로\*

박영희<sup>1</sup>, 박성용<sup>2</sup>

### 요약

본 논문은 KOPIS 빅데이터를 사용하여 문화예술경영 분야에서 필요한 공연예술 관람수요의 분포를 예측하는 새로운 방법을 제시한다. KOPIS 빅데이터는 공연예술관람 수요자의 예매 및 결제정보 뿐만 아니라 각 공연장의 특성과 공연의 정보를 포함한 광범위한 자료이기에 공연예술 분야 장르별 수요자료 식별을 가능하게 해주어 수요분포예측에 아주 유용한 정보를 제공한다. 기존 연구들이 공연예술 관람수요의 시계열 자료만을 이용하여 수요예측을 기계적으로 실시한 것과는 다르게 본 분석에서는 분야별 공연의 횡단면 자료를 사용하여 조건부 관람수요분포를 예측한다. 본 연구에서는 이러한 관람수요분포를 예측하기 위해 조건부 분위수회귀모형을 사용하는 방법을 제시한다. 조건부 분포예측은 추정된 조건부 분위수함수를 이용하여 비모수적(non-parametric)방법을 사용해 추정될 수 있는데 본 연구에서는 적응적 비모수커널추정(adaptive nonparametric kernel estimation)방법을 사용하여 조건부 분포를 추정하였다. 이러한 분위수 회귀방법을 이용하여 공연예술 수요의 분포를 예측함으로써 정보가 필요한 기획사나 공연예술 관계자 및 정부 차원의 정책구상에 유용한 정보를 제공할 수 있다.

주요 용어: 공연예술수요, 수요예측, 분위수회귀, 기계학습, KOPIS 빅데이터.

### 1. 서론

국민소득이 증가하고 국민의 생활 수준이 향상됨에 따라 공연예술경영 분야가 점점 활성화되고 산업의 중요한 한 분야로 자리 잡고 있다. 그러나 공연예술분야의 빠른 성장에 비해 공연예술 관람수요에 대한 예측분석은 활발히 이루어지지 않고 있다. 이는 공연예술관람수요예측이 다른 경제변수나 관광변수 예측보다 어렵다는 점에 기인한다. 공연예술 관람수요는 소득이나 가격과 같은 경제적 변수에도 영향을 받지만, 많은 경우 경제행위를 하는 의사결정자의 개인 선호와 취향에 의해 결정되는 경향이 크기 때문이다. 그럼에도 불구하고 공연관람에 대한 수요예측은 여러 가지 면에서 가장 기초가 되고 중요한 작업임은 틀림없으며 수요예측이 없는 공연예술 분야 활성화 정책은 실행 가능성이 없거나 실행된다고 해도 그 정책효과가 불확실할 수밖에 없다고 하겠다. 또한 각기 다른 공연예술 분야에서 관람수요의 성격은 장르에 따라 매우 다르다. 따라서 실효성이 있는 수요예측을 위해서는 각기 다른 장르에 대한 개별적인 수요예측이 필요하다. 또한, 기존연구들과

\*본 연구는 2023년 제3회 KOPIS 빅데이터 공모전 서비스 개발 부문 수상작으로 지원을 받아 수행되었음.

<sup>1</sup>06974 서울시 동작구 흑석로 84, 중앙대학교 일반대학원 문화예술경영학과 박사과정. E-mail: indigo@cau.ac.kr

<sup>2</sup>(교신저자) 06974 서울시 동작구 흑석로 84, 중앙대학교 경제학부 교수. E-mail: sungpark@cau.ac.kr

같이 서로 특성이 연결되지 않는 여러 단기 공연들을 결합한 관람수요 시계열 자료를 구축하여 분석하는 것에는 분명 한계가 있기에 같은 장르의 공연들을 같이 고려하여 분석하는 것이 절대적으로 필요하다.

기존 공연예술 관람수요예측 연구들은 고전적인 시계열 분석론에서 다루는 ARIMA모형류를 사용한 분석이 주를 이룬다. 머신러닝 기법을 사용하여 수요예측을 하는 연구들도 있지만 대부분 몇 개 모형을 사용하여 수요예측을 하거나 각 모형에서 연구자가 정해야만 하는 튜닝모수(tuning parameter)를 명확한 기준 없이 몇몇 값을 대입하여 예측하는 연구들이 주를 이루고 있다. 이 보다 더 큰 문제는 시계열 자료를 생성할 때 공연들의 단순 합산으로 계산되어 공연들의 이질성을 전혀 고려하지 못하는데 있다.

이러한 이유로 본 연구에서는 기존의 시계열 분석을 이용한 분석이 아닌 각 공연특성을 이용한 횡단면 분석을 고려한다. 같은 장르의 예술공연이라고 해도 각 공연마다 많은 이질성이 존재한다. 즉, 관람객이 많은 공연은 관람객이 적은 공연과는 다른 결정요인을 가질 수 있다는 것이다. 이러한 사실을 고려하기 위해 본 연구는 분위수회귀모형(Quantile regression model)을 이용하여 관람객 수요의 분포를 예측한다. 본 연구의 모형은 점 예측값(point predicted value)뿐만 아니라 예측값의 분포(distribution)를 제공한다는 것에서 기존의 연구와 매우 큰 차이가 있다. 물론 관련 기획사나 공연예술 관계자는 명확한 점 예측값을 선호할 수 있으나 예측의 특성상 예측 오차가 항상 존재하기 마련이고 이를 명확히 나타내주는 정보가 분포예측이므로 이는 유용하게 사용될 것이라 사료된다.

## 2. 분석모형

고려되는 문화공연의 이질적인 성격을 잘 고려할 수 있는 모형으로 분위수회귀모형(quantile regression model)이 관람객 수요에 대한 분석의 예측모형으로 적합하다고 판단된다. 분위수회귀모형은 Koenker and Basset (1978)이 제시한 모형으로 종속변수인 예술공연예술 관람수요의 크기에 따라 여러 설명변수가 미치는 영향이 다른 것을 모형화시킬 수 있는 유용한 모형이며 각 분위수에 대한 조건부 분위수가 추정됨으로 조건부 분포를 추정할 수 있는데 이것이 공연예술 관람수요의 분포예측에 사용될 수 있다. 일반적인 선형 공연수요 회귀방정식은 다음과 같이 주어진다.

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \epsilon_i, \quad (1)$$

여기서  $\epsilon_i$ 는 확률오차항이며  $y_i$ 는 예술공연  $i$ 의 수요를 나타내며  $x_i$ 는 수요를 설명하는 설명변수로서 평균좌석수, 평균예매금액, 평균장당금액, 평균할인금액, 수상여부, 아동공연, 축제여부, 내한공연, 오픈런, 레스토랑 유무, 수유실 유무, 카페 유무, 편의점 유무, 놀이방 유무, 장애인 주차장 유무, 장애인 화장실 유무, 장애인 경사로 유무, 자체주차, 공연주차, 지역, 공연년도, 공연월 등등을 나타낸다. 추정된 조건부 분위수(conditional quantile) 함수는 다음과 같이 나타낼 수 있다.

$$\widehat{Q}_{y_i}(\tau|X_i) = \widehat{\beta}_0(\tau) + \widehat{\beta}_1(\tau)x_{1i} + \dots + \widehat{\beta}_k(\tau)x_{ki}. \quad (2)$$

여기서  $\tau$ 는 0과 1 사이에 있는 분위수 수준을 나타낸다. 예를 들어,  $\tau=0.5$ 일 때 위의 목적함수를 극소화하는 추정량은 조건부 중위값이 된다.  $\widehat{\beta}(\tau)$ 는 회귀분위수(regression quantile) 추정량이며 분

위수 수준( $\tau$ )에 의존하게 되는데  $\tau$ 가 0.9와 0.1일 경우는 위의 설명변수들이 공연예술 관람수요의 상위 90%와 하위 10%에 어떻게 영향을 미치는지를 보여주게 된다. 따라서 분위수회귀는  $\tau$ 를 변경함으로써 공연예술 관람수요를 각기 다른 분위수 수준에서의 결정요인을 분석할 수 있는 방법을 제공한다. 추정방법과 검정방법에 대해서는 Koenker (2005) 및 Galvao, Montes-Rojas and Park (2013)을 참고할 수 있다. 현재까지 분위수회귀모형을 사용하여 공연예술 관람수요에 대해서 분석하는 연구 문헌은 거의 없으며 수요예측에 사용하는 연구는 없다.

위의 조건부 분위수함수가 추정되면 조건부 분포는 다음의 식을 사용하여 추정될 수 있다.

$$\hat{f}_{y_i}(\tau|X_i) = (\tau_i - \tau_{i-1}) / (\hat{Q}_{y_i}(\tau_i|X_i) - \hat{Q}_{y_i}(\tau_{i-1}|X_i)). \quad (3)$$

적절히 선택된  $\tau$ 의 수열( $\tau = (\tau_1, \tau_2, \dots, \tau_N)$ )을 고려하여 조건부 분포를 추정할 수 있게 된다. 직관적으로 각  $\tau$ 에 대해 조건부분위수 함수는 분위수회귀방법에 의해 일치적으로 추정될 수 있으며 이렇게 추정된 조건부 분위수 함수를 이용하여 조건부 분포는 비모수적(non-parametric) 방법을 사용해 추정될 수 있는데 본 연구에서는 적응적비모수커널추정(adaptive nonparametric kernel estimation) 방법을 사용하여 조건부 분포를 추정하였다.

### 3. 추정결과 (서양음악의 경우)

<그림 1>은 서양음악 관람수요의 크기에 따라 설명변수의 한계효과가 어떻게 변화하는지를 나타낸다. 예를 들어, 평균가격효과(avg\_price)는 수요가 낮을 때는 수요를 촉진시키지 않지만 높은 수요에서는 높은 가격이 더 큰 효과를 가짐을 알 수 있다. 수상유무는 저점 또는 중간수준의 수요에는 영향을 미치지 않지만 수요가 아주 높을 때는 수상이 큰 영향을 주지 못하고 있음을 알 수 있다. 이와 같이 분위수회귀방법을 이용하면 변수 간의 관계가 이질성을 가질 때 그 효과를 효율적으로 추정하고 있음을 알 수 있다. 또한, 수요 수준에 따라 각 결정요인이 다름을 알 수 있어 관람수요를 예측할 때 수요의 이질성이 중요함을 알 수 있다.

<그림 2>는 2개의 시나리오에서 서양음악 관람수요의 분포예측값을 보여주고 있다. 이 분포예측은 식 (3)을 이용하여 추정할 수 있는데 이때 설명변수들의 값이 필요하다. 실선으로 주어진 수요의 예측분포는 지역은 서울이고 시점은 2022년 1월, 그리고 모든 설명변수에 사용된 변수들의 평균값을 가정한 시나리오 1의 분포예측값이며 점선으로 나타낸 수요의 예측분포는 지역은 서울이고 시점은 2022년 12월, 설명변수는 이전 시나리오 1과 같은 평균값이지만 수상 여부의 값이 1인, 수상을 한 적이 있는 공연 시나리오 2에 대한 분포예측값이다. <그림 2>에서 볼 수 있듯이 수상 여부는 관객 수요에 양의 영향을 미치게 되어 분포가 중위수준에서 상승하는 것을 볼 수 있다. 조건부 관객이 큰 경우에는 점선이 실선보다 작아지는데 이는 수요가 클 때 수상 여부는 수요에 크게 영향을 미치지 않기 때문이다.

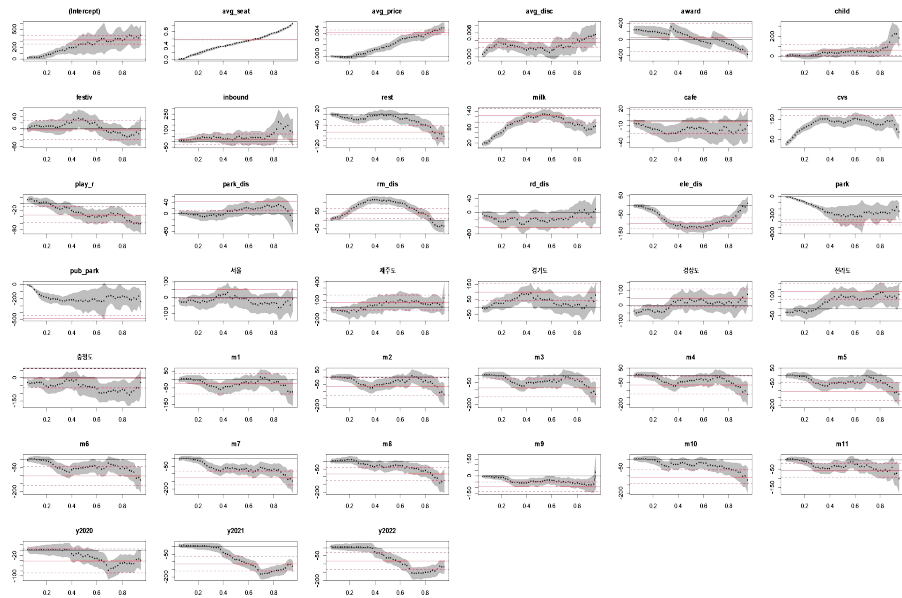


Figure 1. Regression quantile estimates (Classical music)

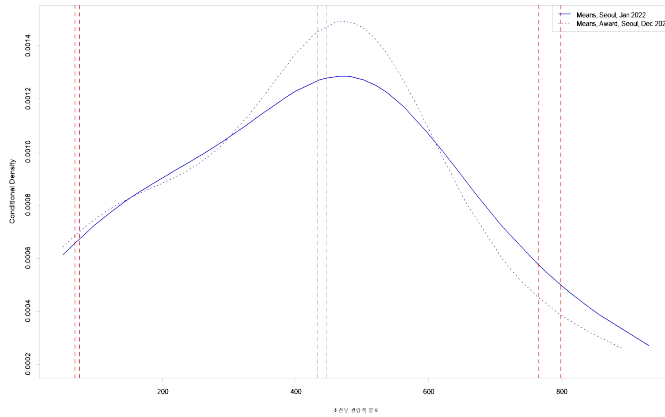


Figure 2. Conditional density estimates

### References

Galvao Jr, A. F., Montes-Rojas, G. and Park, S. Y. (2013). “Quantile Autoregressive Distributed Lag Model with an Application to House Price Returns”. *Oxford Bulletin of Economics and Statistics*, 75(2): 307-321.

Koenker, R. (2005) *Quantile Regression*, Cambridge.

Koenker R., Bassett G. (1978). “Regression Quantile”. *Econometrica*. 46(1): 33-50.

## Spatiotemporal Forecasting in Image Data via Convolutional LSTM Networks

*Sung Jae Kim<sup>1</sup>, Yongbok Cho<sup>2</sup>*

### Abstract

Harmful algal blooms (HABs) have emerged as a growing global concern, with South Korea being no exception. The National Institute of Environmental Research has been diligently monitoring and documenting concentrations of HABs at 939 specific locations, although this coverage does not encompass the entirety of the country. Spanning from 2006 to 2024, the collected data includes spatial information denoted by X and Y coordinates. To address areas lacking direct measurements, the technique of universal kriging is employed to interpolate HABs concentrations. Leveraging the capabilities of QGIS, a geographic information system software, a series of images depicting the interpolated HABs data were generated. These 217 images were then utilized in the ConvLSTM, a convolutional LSTM network, to forecast future occurrences of HABs owing to its ability to capture spatiotemporal dynamics. The ConvLSTM model effectively predicts patterns of HABs utilizing these images, thereby showcasing its potential for application in satellite and video image predictions within similar contexts.

*Keywords* : Universal Kriging, Spatiotemporal, ConvLSTM, QGIS.

---

<sup>1</sup>Research Professor, Department of Management Information Systems College of Business Administration, Dong-A University, Busan 49236, Korea. E-mail: sung0704@dau.ac.kr

<sup>2</sup>Assistant Professor, Department of Management Information Systems College of Business Administration, Dong-A University, Busan 49236, Korea. E-mail: ybcho@dau.ac.kr

## 노인의 성별에 따른 골관절염 유병 관련 요인

박공주<sup>1</sup>, 김민영<sup>2</sup>, 전은미<sup>3</sup>

### 요약

본 연구는 65세 이상 노인들의 골관절염 유병 관련 영향요인을 확인하고 그에 따른 골관절염 관리를 위한 중재방안 개발의 기초자료를 제공하고자 한다. 본 연구대상자는 2016~2020년 국민건강영양조사 자료에서 65세 이상 노인 8,175명(남성 3,521명, 여성 4,654명)으로 SPSS/WIN 25 프로그램을 이용하여 복합표본에서의 분석방법을 적용하였다. 실수와 백분율, 평균, 표준편차, 성별 간 비교는 교차분석을 이용한 카이제곱 검정을 하였고 골관절염 유병여부에 영향을 미치는 요인을 파악하기 위해서 로지스틱회귀분석을 실시하였다. 연구결과 남녀노인의 인구학적 및 건강관련 특성에서 연령, 교육수준, 배우자 동거여부, 주택소유여부, 흡연, 음주, 유산소신체활동, 걷기, BMI, 우울경험, 관절염유무에서 통계적인 차이가 있었다. 본 연구의 표본으로 추정된 한국 65세 이상 노인의 골관절염 유병률은 25.0%(남성 10.3%, 여성 35.9%)였다. 남녀 노인의 인구사회적 특성 및 건강관련 특성에 따른 골관절염 유병률 차이를 검증한 결과 남성노인의 경우 교육수준, 음주, 걷기, 비만여부에서 골관절염 유병률에 통계적 차이를 보이고 있고 여성노인의 경우는 교육수준, 배우자유무, 걷기, 비만여부, 우울여부에서 골관절염 유병률에 유의한 차이를 보이는 것으로 나타났다. 또한 남성노인의 골관절염 유병률에 영향을 미치는 요인은 교육수준이었고, 반면 여성 노인의 경우는 교육수준, 거주지역, BMI로 확인되었다. 이상의 결과를 바탕으로 다각적이고 성별에 따른 개별적인 간호중재프로그램 개발이 필요하다.

주요 용어: 골관절염, 남녀노인, BMI, 로지스틱회귀분석.

<sup>1</sup>50811 경남 김해시 삼안로 112번지 198, 김해대학교 간호학과 부교수. E-mail: conti52@hanmail.net

<sup>2</sup>47011 부산광역시 주례로 45, 경남정보대학교 간호학과 조교수. E-mail: kimmy@eagle.kit.ac.kr

<sup>3</sup>(교신저자) 35345 대전광역시 서구 배재로 155-40(도마동), 배재대학교 간호학과 교수.

E-mail: charminggold@pcu.ac.kr

## 국내 의료서비스 취약지역 대상 의료공급 개선방안 연구: 통계청 및 국민건강보험공단 자료를 활용하여

정수진<sup>1</sup>, 양성병<sup>2</sup>

### 요 약

국내 의료공급은 다른 공공재원 증가에 비해 지속적으로 감소하는 추이를 보이며, 공급과 재정이 수직적(1, 2, 3차 의료기관) 및 수평적(지역적 분포)으로 불균형한 독특한 특징을 나타낸다. 이는 의료전달체계가 공공이 아닌 민간의료 중심으로 형성된 결과로, 필수요료를 포함한 지역 간 의료공급 및 서비스의 양적 및 질적 격차를 초래한다. 이러한 비효율적인 의료전달 체계는 건강보험의 지속가능성에도 문제를 제기하므로, 건강보험 보장성 강화와 의료전달 체계 개선은 단순히 의료서비스의 양적 확대만으로는 해결될 수 없다. 따라서, 민간이 아닌 공공의료기관 중심의 의료전달체계를 구축하기 위해 권역별로 공공의료기관을 확충할 필요가 있다. 이에, 본 연구에서는 통계청 e-지방지표와 국민건강보험공단 표본코호트 DB를 활용하여 2009-2019년 기간 동안 권역별 의료기관 수, 병상 수, 의료인 수 등의 의료서비스 추이를 일반화추정방정식 (Generalized Estimating Equations: GEE)을 통해 분석하였다. 또한, 관내 의료기관 이용 대비 관외 의료기관 이용 비율로 산정된 의료이용 오즈비(Odds Ratio)를 이용하여, 지역별 관외 의료기관 이용에 영향을 미치는 의료형태(예: 입원, 응급, 수술 등), 진료과(예: 신경과, 외과, 정형외과, 산부인과 등), 환자 특성(예: 성별, 연령, 보험등급, 장애등급 등) 등에 대한 효과를 라플라스 혼합 효과로지스틱회귀모형(Laplace Mixed-Effects Logistic Regression Model: MLM)을 통해 추정하였다. 분석 결과를 바탕으로 자체 의료 이용률이 취약한 지역을 중심으로 구체적인 의료공급 개선방안을 도출, 제안한다.

**주요 용어:** 의료공급 불균형, 의료서비스 개선방안, 의료이용 오즈비, 일반화추정방정식, 혼합 효과로지스틱회귀모형, 통계청 e-지방지표, 국민건강보험공단 표본코호트 DB.

<sup>1</sup>02447 서울특별시 동대문구 경희대로 26 경희대학교 경영대학원 경영학과 AI비즈니스전공 석사과정.

E-mail: nxzero@naver.com

<sup>2</sup>(교신저자) 02447 서울특별시 동대문구 경희대로 26 경희대학교 경영대학 경영학과/빅데이터응용학과 교수.

E-mail: sbyang@khu.ac.kr

## Eye movement data for Mild Cognitive Impairment diagnosis using functional data analysis<sup>\*</sup>

*Hakjun Kim<sup>1</sup>, Geunbum Lee<sup>2</sup>, Boncho Ku<sup>3</sup>, Joongil Kim<sup>4</sup>, Jaek Kim<sup>5</sup>, Sangin Lee<sup>6</sup>*

### Abstract

Mild Cognitive Impairment(MCI) is an intermediate stage between normal cognitive decline due to aging and Alzheimer's disease(AD). Researchers are exploring various cost-effective methods to diagnose MCI promptly and prevent its progression to AD. In this study, we investigate the association between eye movement changes and cognitive impairment. Eye movement data were obtained during an anti-saccade experiment, where participants performed either pro-saccades (automatically looking at peripheral stimuli) or anti-saccades (voluntarily inhibiting automatic responses). Our study proposes a novel approach to MCI diagnosis using functional data analysis (FDA) based on pupil dynamics.

*Keywords* : Alzheimer's Disease, Mild Cognitive Impairment, Eye movement changes, Functional Data Analysis, Anti-saccade.

---

<sup>\*</sup>This work was supported by a grant (KSN1824130) from the Korea Institute of Oriental Medicine (KIOM), funded by the Korean government and was supported by the National Research Foundation (NRF) grant funded by the Korea government (NRF-2022M3J6A1084843).

<sup>1</sup>34134 Graduate student, Department of Statistics and Data Science, Chungnam National University, Daejeon, Korea. E-mail: hjhjhj9465@naver.com

<sup>2</sup>34134 Graduate student, Department of Statistics and Data Science, Chungnam National University, Daejeon, Korea. E-mail: veing9856@naver.com

<sup>3</sup>34054 Digital Health Research Division, Korea Institute of Oriental Medicine, Daejeon, Korea. E-mail: secondmoon@kiom.re.kr

<sup>4</sup>34054 Digital Health Research Division, Korea Institute of Oriental Medicine, Daejeon, Korea. E-mail: ikim@kiom.re.kr

<sup>5</sup>34054 Digital Health Research Division, Korea Institute of Oriental Medicine, Daejeon, Korea. E-mail: jaekkim@kiom.re.kr

<sup>6</sup>(Corresponding author) 34134 Associate professor, Department of Statistics and Data Science, Chungnam National University, Daejeon, Korea. E-mail: sanginlee44@gmail.com

## Privacy-Preserving Mortality Predictive Model through Homomorphic Encryption-Enhanced Federated Learning

*Garam Lee<sup>1</sup>, Suyeong Park<sup>1</sup>, Sang-Wook Lee<sup>2</sup>, Jungyo Suh<sup>2</sup>, Sulgi Kim<sup>1</sup>*

### Abstract

In the realm of medical research, leveraging large datasets for big data analysis has become increasingly prevalent. However, integrating data from multiple healthcare institutions poses significant challenges due to concerns regarding patient privacy and data security. The application of Privacy Enhancing Technologies (PETs), such as Federated Learning (FL) and Homomorphic Encryption (HE), has been identified as an effective solution to these challenges, allowing for the secure analysis of sensitive medical data across institutions. This study explores the integration of FL and HE to address the gradient leakage issue of FL in predicting postoperative mortality from multi-institutional data, comprising 425,160 non-cardiac surgery patients across three medical institutions. Employing logistic regression as our predictive model, we compared the performance of models trained on vanilla FL against models trained on datasets using HE-based FL. Our findings indicate that the incorporation of HE into FL does not compromise predictive performance, as evidenced by similar AUROC metrics between standard FL and FL enhanced with HE. Furthermore, our analysis revealed that HE could be implemented without significant overhead in terms of encrypted data size and computational efficiency. This research underscores the viability of homomorphic encryption as a solution to enhance the security of FL enabling the collaborative use of multi-institutional medical data for improved predictive analytics.

*Keywords:* Homomorphic Encryption, Federated Learning, Gradient Leakage, Privacy-Enhancing Technology (PET), Multi-institutional Data.

---

<sup>1</sup>CryptoLab Inc., Seoul.

<sup>2</sup>Asan Medical Center, University of Ulsan College of Medicine, Seoul

Factors affecting the adoption of adaptation measures to climate change.  
A case study in Huong Phong commune, Huong Tra town,  
Thua Thien Hue province, Vietnam

*Thanh Hoa Bui<sup>1</sup>, Keunjae Lee<sup>2</sup>*

Abstract

In recent decades, changes in climate have caused impacts on natural and human systems on all continents and across the oceans. Vietnam generally and the coastal area in Thua Thien Hue province particularly is vulnerable to climate change and some extreme climate events. Adaptation is considered as one of the best long-term strategies to a community to better face with local extreme conditions and associated climate change. This study used Logistic regression model to determine factors influencing farmers' decisions to adopt climate change adaptation measures. The results indicated that that age, years of schooling, years of farming experience of the household head, household size, ratio of number of farm labors to number of consumers, farmer's access to extension services and adaptation measures, and the place where farmer lives factors significantly influence adoption decisions. From the results, some recommendations were derived to help farmers in the coastal area of Thua Thien Hue province adapt to climate change.

*Keywords* : climate change adaptation, extreme climate events, Logistic regression, perception, Thua Thien Hue province.

---

<sup>1</sup>46241 Busan Geumjeong-gu, Busandaehak-ro 63beon-gil 2, Pusan National University, Department of Economics, Graduate student. E-mail: buithanhhoa@pusan.ac.kr

<sup>2</sup>46241 Busan Geumjeong-gu, Busandaehak-ro 63beon-gil 2, Pusan National University, Department of Economics, Professor. E-mail: kjlee@pusan.ac.kr

## Select priority tourist sites in Vietnam's coastal tourist areas that satisfy green criteria: A research using Fuzzy AHP and Fuzzy TOPSIS

*Nguyen Tuan Dat*<sup>1</sup>, *Keunjae Lee*<sup>2</sup>

### Abstract

Vietnam's breathtaking coastline is a major tourist attraction, boasting pristine beaches, diverse ecosystems, and rich cultural heritage. However, this sector also faces the challenge of balancing economic growth with environmental protection. The concept of a green economy offers a solution by promoting sustainable tourism practices that minimize environmental impact while fostering social equity and economic benefits. This study proposes a framework for prioritizing tourist attractions along Vietnam's coast based on green economy principles. This framework utilizes two multicriteria decision-making (MCDM) techniques: Fuzzy Analytic Hierarchy Process (AHP) and Fuzzy Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). By prioritizing tourist attractions that align with green economic principles, we can ensure the long-term sustainability of its coastal tourism industry while preserving its unique natural environment and rich cultural heritage.

*Keyword:* green economy; fuzzy AHP; fuzzy TOPSIS; coastal areas; Vietnam.

---

<sup>1</sup>(Corresponding Autor) Ph.D.Student, Department of Economics, Pusan National University, Busan, Korea.  
E-mail: nguyendat.tuan@pusan.ac.kr

<sup>2</sup>Professor, Department of Economics, Pusan National University, Busan, Korea.

## A Consumption-Based Identification of Global Economic Uncertainty

*Hwagyun Kim<sup>1</sup>, Eunhee Lee<sup>2</sup>, Joon Y. Park<sup>3</sup>*

### Abstract

This paper identifies a global uncertainty factor by estimating an international asset pricing model featuring macroeconomic uncertainty with long-run risk factors. The global factor captures the time-varying fluctuations of common stochastic volatilities of consumption and dividend growths for countries, and reflects uncertainty in that it generates the highest volatility of volatility in transition period. The model quantitatively explains key asset pricing moments, and the estimated factor sharply increases during major international adverse events. Shocks to our global economic uncertainty factor significantly account for the likelihood of key economic and financial events, and outperforms existing measures of economic and financial uncertainties.

*Keywords:* Global Macroeconomic Uncertainty, Long-Run Uncertainty, Stochastic Volatility, Volatility of Volatility, Bayesian Method, Generalized Method of Moments.

---

<sup>1</sup>Department of Finance, Mays Business School, Texas A&M University. E-mail: hagenkim@tamu.edu

<sup>2</sup>Department of Economics, Gyeongsang National University. E-mail: eunheelee@gnu.ac.kr

<sup>3</sup>Department of Economics, Indiana University and Sungkyunkwan University, joon@indiana.educational  
Korea Maritime and Ocean University, Republic of Korea. E-mail: ngotpham@g.kmou.ac.kr

## Application of the Gaussian mixture model - hidden Markov model for financial time series<sup>\*</sup>

*Seokju Ki<sup>1</sup>, Yongjin Cho<sup>2</sup>, Hee-Young Kim<sup>3</sup>*

### Abstract

The state space model has become a powerful tool for time series modeling and forecasting. Such models, in conjunction with the Kalman filter, have been used in a wide range of applications. Hidden Markov models (HMMs) keeps the hierarchical structure of the Gaussian linear state space, but removes the limitations of linearity and Gaussianity. HMMs is a discrete time process  $\{(X_t, Y_t), t \in N\}$ , where  $\{X_t, t \in N\}$  is a Markov chain and, conditional on  $\{X_t, t \in N\}$ ,  $\{Y_t, t \in N\}$  is a sequence of independent random variables s.t. the conditional distribution of  $Y_t$  only depends on  $X_t$ . There have been numerous models based on HMMs, including stochastic volatility models. Most recently, Gopinathan et al. (2024) propose new model, GMM-HMM, and the unique characteristic of this method is the modelling of hidden states of HMMs as Gaussian mixture model. In this article, to forecast the next day's closing price for a given day, Samsung Electronics and SK Hynix, we most adopt the method of Gopinathan et al. (2024), and compare GMM-HMM to ARIMA models.

*Keywords* : Hidden Markov model; Gaussian mixture model; time series.

---

<sup>\*</sup>This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07045707, NRF-2021R1F1A1048309) and Natural Science Foundation of Korea university (No. K2021811).

<sup>1</sup>Student, Department of Bigdata Science, Korea University, Sejong-Street 2511 Sejong-City 339-700.

<sup>2</sup>Student, Department of Bigdata Science, Korea University, Sejong-Street 2511 Sejong-City 339-700.

<sup>3</sup>(Corresponding Author) Professor, Bigdata Science, Korea University, Sejong-Street 2511 Sejong-City 339-700, E-mail : starkim@korea.ac.kr

## Revisit Forward Rate Unbiasedness Hypothesis: Smooth Time-Varying Parameter Approach

*Xiaoyu Kang*<sup>1</sup>, *Sung Y. Park*<sup>2</sup>

### Abstract

In order to examine the forward rate unbiasedness hypothesis (FRUH), this paper develops the cointegration and difference regression models to analyze the relationship between spot and forward exchange rates. Specifically, we consider sixteen foreign exchange markets of advanced and emerging countries. Many previous studies analyze a fixed relationship between these two rates to test FRUH. However, this assumption is quite strong since there are many factors such as macroeconomic shocks that can change the fixed relationship over the time horizon. Consequently, we propose a time-varying cointegrating regression model in which the functional parameters are specified by a well-known orthonormal series. It is also well-known that exchange rate returns exhibit volatility clustering effects. To address this, we consider a generalized autoregressive conditional heteroscedasticity (GARCH) type specification in the proposed model. The results demonstrate that the relationships are indeed time-varying, and that the FRUH is maintained over specific periods. Moreover, the results differ between countries. Our research findings also have implications for policy makers.

*Keywords* : Spot and Forward Exchange Rate; Forward Rate Unbiasedness Hypothesis; Time-varying Cointegration Regression; GARCH-type Errors.

---

<sup>1</sup>Doctoral Program, School of Economics, Chung-Ang University, 84 Heukseok-Ro Dongjak-Gu, Seoul, 06974, Republic of Korea. E-mail: kxy8166@cau.ac.kr

<sup>2</sup>(Corresponding Author) Professor, School of Economics, Chung-Ang University, 84 Heukseok-Ro Dongjak-Gu, Seoul, 06974, Republic of Korea. E-mail: sungpark@cau.ac.kr

## Dynamic dependence and risk spillover between oil price shocks and nonferrous metals.

*Meiling Jin<sup>1</sup> Sung Y. Park<sup>2</sup>*

### Abstract

This study examines the dynamic interdependencies and risk propagation between six base metals and four categories of oil shocks using the CoVaR methodology. Our findings suggest that the interrelationship between base metals and oil shocks is subject to fluctuations and changes across different shocks. Aluminum and copper demonstrate both positive and asymmetric upper tail dependence on economic activity shocks, while zinc and lead exhibit upper and lower tail independence. In the majority of cases, oil demand shocks indicate the absence of positive lower tail dependence or tail dependence in metals. There exist both positive and negative dependencies on oil inventory demand. This study also emphasizes the heterogeneity of the impact of oil shocks on non-ferrous metal futures prices, as it does not solely depend on oil prices like previous research. It reveals that there is asymmetric risk transmission in extreme market conditions, while there is no extreme tail dependence for some shocks. Consequently, the findings of this research provide additional information for those working in the non-ferrous metal industry, financial authorities, and commodity investors.

---

<sup>1</sup>School of Economics, Chung-Ang university, 84 Heukseok-Ro, Dongjak-Gu, Seoul, South Korea.  
Email: meiling1206@cau.ac.kr.

<sup>2</sup>(Corresponding Author) School of Economics, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul, South Korea. E-mail: sunpark@cau.ac.kr.

# Cryptocurrency Price Forecasting using Variational AutoEncoder with Versatile Quantile Modeling

*Sungchul Hong<sup>1</sup>, Seunghwan An<sup>2</sup>, and Jong-June Jeon<sup>3</sup>*

## Abstract

In recent years, there has been a growing interest in probabilistic forecasting methods that offer more comprehensive insights by considering prediction uncertainties rather than point estimates. This paper introduces a novel variational autoencoder learning framework for multivariate distributional forecasting. Our approach employs distributional learning to directly estimate the cumulative distribution function of future time series conditional distributions using the continuous ranked probability score. By incorporating a temporal structure within the latent space and utilizing versatile quantile models, such as the generalized lambda distribution, we enable distributional forecasting by generating synthetic time series data for future time points. To assess the effectiveness of our method, we conduct experiments using a multivariate dataset of real cryptocurrency prices, demonstrating its superiority in forecasting high-volatility scenarios.

*Keywords* : Cryptocurrency Forecasting, Time Series Forecasting, Variational AutoEncoder, Distributional Learning.

---

<sup>1</sup>(co-author) Postdoctoral researcher, Department of Statistics, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul, 02504, Republic of Korea.

<sup>2</sup>(co-author) PhD candidate, Department of Statistical Data Science, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul, 02504, Republic of Korea.

<sup>3</sup>Professor, Department of Statistics, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul, 02504, Republic of Korea.

## Deep fake detection system for ADD 2023 challenge

*Soyul Han<sup>1</sup>, Taein Kang<sup>2</sup>, Sunmook Chor<sup>3</sup>, Jaejin Seo<sup>2</sup>, Sanghyeok Chung<sup>3</sup>,  
Sumi Lee<sup>2</sup>, Seungsang Oh<sup>4</sup> and Il-Youp Kwak<sup>5</sup>*

### Abstract

The paper presents the participation of the CAU\_KU team in the ADD 2023 Challenge, specifically in track 1.2 (audio fake game - detection track) and track 3 (deepfake algorithm recognition track). Various deep learning models were explored using features from the pretrained wav2vec2 network, as well as CQT, mel-spectrogram, etc. We modified the representation extraction component of the AASIST model to incorporate 2D spectrograms (wav2vec2 or CQT) and attempted different deep learning models, with model ensembling employed to create the final model. For track 1.2, our submitted ensemble model for round 1 utilized the CQT-LCNN and CQT-AASIST models. For round 2, our model used the CQT-LCNN, CQT-AASIST, and W2V2-GMM models. For track 3, we ensembled the CQT-LCNN, CQT-OFD and AASIST models. Additionally, we applied the openmax algorithm to detect unknown deepfake attacks. Our best submission achieved 23.44% and 21.26% on round 1 and 2 of track 1.2, respectively, and ranked 3rd in track 1.2.

*Keywords:* audio deep synthesis, audio deepfake detection, deep learning, deepfake algorithm recognition.

---

<sup>1</sup>Postdoctoral researcher, Institute for Data Innovation in Science, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul, 08826, Republic of Korea. E-mail: soyul5458@snu.ac.kr

<sup>2</sup>Department of Statistics, Chung-Ang University, Seoul, 84, Heukseok-ro, Dongjak-gu, Seoul, 06974, Republic of Korea.

<sup>3</sup>Department of Mathematics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Republic of Korea.

<sup>4</sup>(Co-corresponding author) Department of Mathematics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Republic of Korea. E-mail: seungsang@korea.ac.kr

<sup>5</sup>(Co-corresponding author) Department of Applied Statistics, College of Business & Economics, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul, 06974, Republic of Korea. E-mail: ikwak2@cau.ac.kr

## 벨 모양 필터를 이용한 오디오 데이터 증강 기법 연구

서재진<sup>1</sup>, 광일엽<sup>2</sup>

### 요약

딥러닝은 음성 인식, 위조 음성 탐지, 음향 이벤트 감지, 화자 인식 등 다양한 오디오 분야에서 중요한 역할을 한다. 딥러닝 모델을 효과적으로 학습시키기 위해서는 충분한 양의 데이터가 필요하다. 그러나 실제 데이터를 수집하는 것은 비용과 시간이 많이 드는 일이므로, 데이터 증강 기법이 중요하게 사용된다. 이러한 증강 기법을 통해 모델이 다양한 환경에서 더 강건하게 작동할 수 있고, 성능을 향상시킬 수 있다. 최근에는 오디오 데이터에 적용되는 증강 기법들이 많이 연구되고 있으며, 그 중에서도 스펙트로그램에 마스킹을 적용하는 SpecAugment 방법이나, 필터를 생성하여 스펙트로그램의 주파수 대역에 적용하여 스펙트로그램을 변형시키는 FilterAugment 등이 있다. 이러한 증강 기법에 더하여 본 연구에서는 벨 모양 필터 마스킹 기법을 제안한다. 이 새로운 기법은 마스킹 영역에 벨 모양 필터를 적용하여 부드러운 마스킹 효과를 제공함으로써 기존의 마스킹 기법에 비해 스펙트로그램의 형태를 유지하면서 오디오 데이터를 증강할 수 있다. 음성 위조 탐지 작업에 해당 기법을 적용한 실험 결과, 벨 모양 필터 마스킹 기법이 다른 증강 기법에 비해 낮은 EER(Equal Error Rate)을 달성하여 효과적임을 확인하였다. 이러한 분석을 통해 벨 모양 필터 마스킹 기법이 음성 위조 탐지뿐만 아니라 다양한 오디오 처리 응용 분야에서도 성능 향상을 기대할 수 있음을 시사한다.

주요 용어: 데이터 증강, 딥러닝, 스펙트로그램, 오디오 데이터 등.

<sup>1</sup>06974 서울시 동작구 흑석로 84, 중앙대학교 통계데이터사이언스학과 석사과정. E-mail: seojaejin@cau.ac.kr

<sup>2</sup>(교신저자) 06974 서울시 동작구 흑석로 84, 중앙대학교 응용통계학과 부교수. E-mail: ikwak2@cau.ac.kr

## 기계학습 기반 학교 배치 형태에 따른 실내외 소음도의 예측에 관한 연구\*

김송옥<sup>1</sup>, 장서일<sup>2</sup>

### 요 약

일상생활에서 소음에 장기간 노출될 경우, 인체는 WHO에서 정의하는것과 같이 신체적, 정신적 그리고 사회적 기능에 일시적 또는 장기적으로여러 부정적 영향을 받게 되는데 교육환경 내에서는 학생들의 학습 능력, 교사의 강의 전달력 등 전반적인 학습 경험에도 소음이 많은 영향을 끼치게 된다. 특정 지역에서 소음 수준을 파악하기 위해 수행되는 소음 예측 시뮬레이션의 경우 방대한 양의 자료 수집이 필요하며 소음지도를 작성하기까지 많은 시간이 소요된다는 단점이 있다. 또한, 학교 신설 시에 소음도 평가가 진행되는 시점에서 시설 배치가 완전히 이루어지지 않은 경우, 실제 환경에 적합하지 않은 방음 시설이 설계될 우려가 있다. 따라서, 소음을 예측하는 데 편리성을 도모하고 소음을 관리하는 합리적인 방안을 제안하고자 소음도 예측을 위한 인공지능경망 모델을 제안하였다. 연구에 사용된 자료는 2000년 이후 시공된 초등학교, 중학교 총 32개교를 대상으로 선정하여 30분간 수집한 교통량, 속도, 그에 따른 등가 소음도와 개별 학교의 구조, 배치 형태, GIS 데이터 등을 사용하였다. 실외 소음도를 예측하는 데에 사용하는 입력 변수는 대상 학교의 실제 환경을 충분히 설명할 수 있도록 구성하였으며, 실내 소음도 예측에 사용한 입력 변수 또한 소음이 외부로부터 내부로 유입되는 환경과 대상 건물에 대한 정보를 포함하고 있다. 본 연구에서는 현장 측정으로 수집된 데이터와 소음 예측 프로그램인 Sound PLAN에서 검증된 모델을 통해 생성한 데이터를 사용하였다. 이를 통해, 시뮬레이션 데이터가 모델 학습에 활용될 수 가능성을 확인하고 모델의 전반적인 예측 성능을 평가하고자 하였다. 또한, 순열중요도 분석을 통해 모델 성능에 큰 영향을 미치는 변수들을 파악하고자 하였으며, 건물과 소음원의 세 가지 배치 유형(수평형, 수직형, 혼합형)이 중요한 독립변수로 작용하는 것을 확인하였다. 본 모델은 전통적인 시뮬레이션 소음 예측 방식을 보조하는 역할을 수행하거나 교육시설 입지를 고려할 때 배치 유형에 따라 예측되는 소음 수준을 참고하여 교사 배치를 결정하는 도구로 활용할 수 있을 것으로 기대된다.

주요 용어 : 기계학습, 심층신경망, 도로교통소음 예측, 교육환경.

\*이 본 결과물은 한국토지주택공사의 지원을 받아 수행된 연구임(No. 2022659)

<sup>1</sup>서울시립대학교 도시빅데이터융합학과 박사과정. E-mail: kso@uos.ac.kr

<sup>2</sup>(교신저자) 서울시립대학교 환경공학부 교수. E-mail: schang@uos.ac.kr

## 국내 열차 차종 분류 신경망 모형의 입력데이터 구성에 대한 연구\*

김민중<sup>1</sup>, 장서일<sup>2</sup>

### 요약

현재 철도의 급격한 발전은 사람과 물류의 이동 편의성을 가져옴과 동시에 소음에 의한 피해를 가져왔다. 철도 주변 지역의 소음 피해를 파악하기 위해 수행되는 소음 예측 시뮬레이션에는 철도 소음 영향 인자에 대한 수집과 입력이 필요하다. 다양한 영향인자 중에서 차종별 교통량은 필수적으로 수집되어야 하지만 기존의 교통량 수집 방식은 오차를 줄이기 위해 수작업으로 진행되는 것이 일반적이고 시간적, 비용적 문제가 있었다. 이러한 문제를 해결하고자 철도에서 측정된 열차 통과 소음에서 얻어지는 정보를 이용하여 차종을 구분하는 신경망 모형에 대한 연구를 수행한다. 통과 소음을 신경망학습에 이용함에 있어 원시 데이터에서 특징을 추출하는 것은 학습을 위해 가장 기본적으로 수행되어야 하는 과정이다. 특징 추출 방법에 따라 신경망 모형의 분류 성능이 달라짐에도 불구하고 철도소음에 적합한 특징 추출 방법이 명확히 제시되지는 못했다. 따라서 본 논문에서는 특징 추출 방법이 분류 성능에 미치는 영향에 대하여 연구한다. 또한 차종 분류 모형을 실제 도로에서 교통량 수집과 소음도 예측에 이용하기 위해 분류 모형과 차속의 관계, 분류 모형과 소음도 예측의 관계에 대해 연구한다. 이번 연구의 분류 대상 차종은 소음 예측에 사용되는 KTX와 도시철도 그리고 화물차이다. 외부 소음의 영향이 없는 2차선 직선철도에서 평균 속도 70km/h의 차종별 통과 소음을 직접 수집하여 연구에 활용하였다. 수집한 차종별 통과 소음에서 추출할 특징 데이터는 관련 연구들에서 높은 성능을 보여주었던 특징들을 위주로 선정하였다. 소음도와 1/3옥타브 밴드로 구성된 기본 소음 특징, 통과 소음의 평균 MFCC 특징, 주파수 스펙트럼 이미지 특징으로 데이터셋을 만들어 학습에 이용하였다. 이번 연구처럼 소음 정보만을 이용하는 연구는 이미지를 이용한 연구에서 어려운 점으로 꼽히는 이미지 품질에 대한 문제를 해소해 줄 수 있을 것으로 기대된다. 트킹 밤시간 영상 수집, 장애물에 의한 렌즈 왜곡 등의 문제가 있다. 이러한 이미지 사용의 어려움을 이번 연구에서는 소음 신호로 해결할 수 있는 가능성을 보여주었다. 또한 통과 소음을 현재는 차종 분류에만 이용하고 있지만 지속적 소음 수집과 학습을 통해 통고 속도 예측 등에 활용할 수 있을 것으로 예상되며 이렇게 수집되는 정보는 최종적으로 소음도 예측의 정확도를 높여줄 것으로 기대하고 있다.

주요 용어: 합성곱신경망, 특징추출, 철도소음 예측.

\*이 본 결과물은 과학기술정보통신부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임  
(No. 2022M3J6A1084845)

<sup>1</sup>서울시립대학교 환경공학과 석사과정. E-mail: book049500@uos.ac.kr

<sup>2</sup>(교신저자) 서울시립대학교 환경공학부 교수. E-mail: schang@uos.ac.kr

## SEER 데이터를 이용한 다양한 딥러닝 생존분석 모형들의 비교연구

박근호<sup>1</sup>, 민은정<sup>2</sup>

### 요약

생존분석은 관심있는 사건이 발생할 때까지의 시간을 분석하는 방법이며, 다양한 분야에 적용이 가능하다. 최근 딥러닝이 관심을 많이 받게 되면서 딥러닝 모형을 활용하여 생존 데이터를 분석하려는 시도 또한 각광 받고 있다. 이에 기존에 생존분석에 사용하는 통계분석과 함께 딥러닝 모형을 활용하여 모형의 성능을 비교하고자 한다. 사용하는 모형은 생존분석에서 대표적으로 사용하는 Cox-regression, AFT 모형, 기계학습 모형인 Random Survival Forest, 딥러닝 모형인 DeepSurv, DeepHit 모형을 사용한다. 모의실험은 절단비율 등 다양한 상황을 고려하는 모의실험을 만들었고, 실제 데이터 분석은 SEER 데이터 중 갑상선 암을 선택하여 분석하였다. 모형 성능 비교 방법으로는 Concordance index와 Integrated brier score를 사용하였고, 모의실험과 실제 데이터를 분석하여 비교한 결과 딥러닝 모형에서 성능이 가장 좋았다. 하지만 딥러닝 모형이 기존에 모델과 비교해서 큰 차이를 보이지 못했는데, 그 이유 중 하나가 Hyperparameter tuning에 있다.

주요 용어: 생존분석, 딥러닝, 비교, Concordance index, Integrated brier score.

<sup>1</sup>가톨릭대학교 의생명건강과학과 의학통계전공

<sup>2</sup>가톨릭대학교 의과대학 의생명과학교실

## 엔트로피 지수를 통한 문화재정지출의 불균등도 분석

오주현<sup>1</sup>

### 요약

문화격차(cultural disparity)는 문화 활동에 대한 접근과 향유 기회가 개인마다 다르게 작용하는 문화 불평등 현상을 의미한다. 문화격차 현상이 가속화될수록 특정 계층에 있어 문화소외현상이 발생할 수 있으며, 문화격차는 경제적 격차로 이어져 사회적 불평등으로까지 확대될 수 있다. 이러한 문화격차를 해소하기 위해 정부의 문화·예술 분야에 대한 재정지출의 규모도 지속적으로 증가하였지만, 여전히 문화격차는 해소되고 있지 않는 실정이다. 따라서 본 연구는 문화재정지출의 적절성을 파악하기 위해 국가별 및 한국의 지역별 문화재정지출 규모를 비교하고자 한다. 기존 연구들에서는 문화재정지출 규모를 국가 및 지역별로 비교분석 하였으나 국가별 및 한국의 지역별 문화재정지출의 격차와 불균등도를 종합적으로 비교 및 평가하지 못한 한계가 있다. 이에 본 연구는 엔트로피 접근법(entropy approach)을 이용하여 국가별 및 지역별 문화재정지출의 규모와 불균등도를 분석하고자 한다. 엔트로피 지수를 통해 국가별 및 지역별 격차와 시점 간 불균등도를 비교 및 분석함으로써 문화재정지출의 적절성에 대한 시사점을 제공하고자 한다.

주요 용어: 문화격차, 문화재정지출, 엔트로피 접근법, 불균등도.

## 주요 에너지 선물가격과 국내 업종별 주가지수 간 연결성 분석

임지훈<sup>1</sup>, 박성용<sup>2</sup>

### 요 약

코로나19 이후, 러시아-우크라이나 전쟁과 이스라엘-하마스 전쟁 등의 불안정한 국제 정세가 지속되며 에너지 안보의 중요성이 대두되었다. 각국은 불안정성 하에서 국가 내 산업 생태계의 지속을 위한 에너지 채널을 유지 및 모니터링하는 것에 대한 필요가 이전보다 늘어나게 되었다. 하지만 대부분 에너지와 산업 간 관계성을 다룬 연구들은 에너지 가격 충격이 산업에 미치는 연구가 주를 이루고 있으며, 반대 방향의 해석을 포함하는 연구는 충분히 이루어지지 않았다. 에너지 가격과 산업의 상황은 상호작용하기에 양방향적 분석이 더 적절할 것으로 사료된다. 따라서 양방향의 Spillover 연결 수준을 나타내는 Diebold와 Yilmaz의 Connectedness Network framework를 TVP-VAR 방식으로 개선한 Antonakakis, Chatziantoniou, Gabauer의 방법론을 채용하여, 원유, 석탄, 천연가스 선물가격과 KRX 업종별 주가지수 간의 연결성 분석을 시행하였다. 정태적 네트워크 분석 결과, 에너지 선물가격들은 네트워크에서 Spill-over의 수용자였고, 업종별 주가지수는 업종별로 역할이 달랐다. 이 중 경기소비재 지수가 네트워크 상에서 가장 큰 Spill-over 전달자였으며, JKM(Japan-Korea Marker) 선물가격이 가장 큰 Spill-over 수용자임을 확인하였다.

주요 용어: Network Connectedness(네트워크 연결성), Spillover effect, TVP-VAR, 분산분해.

### 1. 서론

2019년도 이후부터 코로나19, 러시아-우크라이나 전쟁, 이스라엘-하마스 전쟁과 미중 갈등의 지속, 등의 일련의 사건들과 함께 국제적으로 불안정성이 높아진 형국이다. 이러한 세계정세 하, 미국, 중국, 러시아와 같이 자국에서 에너지 자원 생산이 가능한 국가들의 경우, 경제 활성화에 있어서 에너지 수급이 국제적 불안정성에 노출되는 정도는 그렇지 못한 국가들에 비해서 낮은 편이다. 반대로 자국 내 에너지 자원 생산이 불가능한 국가들은 에너지 수급 과정에서 국제적 불안정성에 온전히 노출된다(Zhu, Deng, Hu, 2023). 이는 해당 국가 경제주체에 있어 경제적 활동에 투입되는 비용을 상승, 소비 및 투자를 위축시킬 여력이 있기에 정부는 이를 고려한 전략을 수립해야 한다.

때문에, 많은 기존의 연구들은 유가를 대표로, 에너지 가격 충격이 국가 경제의 다양한 부문에 미치는 영향에 대해 분석해오고 있지만, 에너지 가격은 경제와 서로 영향을 주고받기 때문에 이는 충분치 않다. 특히, 산업 부문에 있어, 에너지 가격은 산업 내 개별 기업의 성과에 영향을 주는(Lee, Yoon, 2022) 동시에, 에너지 가격은 에너지 공급 대비 산업의 에너지 수요에 영향을 받는다(Bahmani-Oskooee, Ghodsi, Hadzic, 2019).

<sup>1</sup>06974 대한민국 서울시 동작구 흑석로 84, 중앙대학교 경제학부 석사과정. E-mail: jihoon700@cau.ac.kr

<sup>2</sup>(교신저자) 06974 대한민국 서울시 동작구 흑석로 84, 중앙대학교 경제학부 교수. E-mail: sungpark@cau.ac.kr

에너지 수급과 산업 경제의 양방향 관계성을 다룬 기존의 연구는 에너지 생산이 가능한 국가를 중심으로 이루어져 있는 경향이 있다(Wang, Liu, Fan, 2023). 때문에, 정작 연구가 필요한 국가의 에너지 수급과 산업 경제의 관계성을 다룬 연구는 에너지 생산 국가 대상 연구에 비해 충분히 이루어지지 않았다. 따라서 자국 내 에너지 생산이 불가하지만 발달한 산업 수준을 지닌 한국의 사례를 중심으로 에너지 가격과 산업 간 상호작용을 분석하는 것은 에너지와 산업 간 관계성 연구의 확장으로서 의의를 가진다.

## 2. 방법론

### 2.1. TVP-VAR(Time-Varying Parameter Vector Autoregression)

Pesaran, Shin(1998)에서 제시된 Generalized impulse response functions(GIRF)와 Generalized Forecast Error Variance Decomposition(GFEVD)을 이용한 Diebold, Yilmaz(2014)의 네트워크 연결성 분석을 통해 금융 시계열 데이터 네트워크 분석을 가능케 하였다. 하지만 Antonakakis, Chatziantoniou, Gabauer(2020)에 의하면, Diebold, Yilmaz(2014)의 동태적 분석 방식은 rolling-window 방식으로, 해당 모형이 임의의 window 크기를 설정해야 하고, 이에 따른 관측치 손실이 존재하는 단점을 가지고 있다고 언급하였다. 더해서 이상치에 대해 민감하다고 주장하였으며, TVP-VAR을 적용하여 해당 단점을 보완할 수 있음을 보였다.

TVP-VAR 방법론은 Koop, Korobilis(2014)에서 제시된 방법론으로, stochastic volatility 칼만 필터에 의해 계수와 분산이 고정되지 않은 방법론이다. TVP-VAR 모형은 다음과 같다.

$$\begin{cases} Y_t = \beta_t Y_{t-1} + \epsilon_t & \epsilon_t | F_{t-1} \sim N(0, S_t) \\ \beta_t = \beta_{t-1} + v_t & v_t | F_{t-1} \sim N(0, R_t) \end{cases}$$

$Y_t$ 는  $N \times 1$  조건부 변동성 벡터,  $Y_{t-1}$ 는  $Np \times 1$  lagged 조건부 벡터,  $\beta_t$ 는  $N \times Np$  시변환 계수 행렬,  $\epsilon_t$ 는  $N \times 1$  오차 disturbance 벡터,  $S_t$ 는  $N \times N$  시변환 분산-공분산 행렬이다.  $\beta_t$ 는  $\beta_{t-1}$ 에 의존(depend)하고  $N \times Np$  오차 행렬과  $Np \times Np$  분산-공분산 행렬을 가진다.

저자들은 Wold representation theorem을 이용하여 VAR을 Vector Moving Average(VMA) representation으로 변환하였다.

$$Y_t = A_t \epsilon_t, \quad A_{t,t} = \beta_{1,t} A_{t-1,t} + \dots + \beta_{p,t} A_{t-p,t}$$

$\beta_t = [\beta_{1,t}, \beta_{2,t}, \dots, \beta_{p,t}]'$ 이고,  $A_t = [A_{1,t}, A_{2,t}, \dots, A_{p,t}]'$ 인  $N \times N$  모수 행렬들이다.

GIRF는  $i$  번째 변수 충격에 따른 다른 변수들의 반응으로 표현된다. 저자들은 구조적 모형이 없기에, J-step-ahead forecast 간의 difference들을 구하였다. difference는 다음과 같다.

$$\begin{cases} GIR_t(J, \delta_{j,t}, F_{t-1}) = E(Y_{t+J} | \epsilon_{j,t} = \delta_{j,t}, F_{t-1}) - E(Y_{t+J} | F_{t-1}) \\ \Psi_{j,t} = \frac{A_{j,t} S_t \epsilon_{j,t}}{\sqrt{S_{jj,t}}} \frac{\delta_{j,t}}{\sqrt{S_{jj,t}}} & \delta_{j,t} = \sqrt{S_{jj,t}} \end{cases}$$

$J$ 는 forecast horizon,  $\delta_{j,t}$ 는  $j$ 번째 selection 벡터,  $F_{t-1}$ 은  $t-1$ 기까지의 정보 집합이다.

다음으로, GFEVD는 다른 변수들에 대한 특정 변수 분산이 가지는 점유율로 해석 가능하며, 이를 일반화한다면, 다음과 같이 한 행의 합이 1이, 행렬 전체의 합이  $N$ 이 된다.

$$\tilde{\phi}_{ij,t}(J) = \frac{\sum_{t=1}^{J-1} \psi_{ij,t}^2}{\sum_{j=1}^N \sum_{t=1}^{J-1} \psi_{ij,t}^2}, \quad \sum_{j=1}^N \tilde{\phi}_{ij,t}(J) = 1, \quad \sum_{i,j=1}^N \tilde{\phi}_{ij,t}(J) = N$$

### 3. 실증분석 결과

한국의 에너지 수입 구조를 고려하여 5개 에너지 선물 가격(Newcastle FOB, WTI, Dubai, JCC, JKM futures)과, 17개 KRX 업종별 주가지수의 지난 5개 연도 기간(2019-04-22~2022-04-26)의 데이터를 사용하여 네트워크 연결성 분석을 진행하였다.

분석 결과, 전달자 부문은 건설, 경기소비재, 필수소비재, 에너지화학, IT, 기계장비, 증권, 철강, 운송으로 총 9개 부문이며, 가장 큰 Spillover 전달자는 경기소비재 지수이다. 수용자 부문은 자동차, 은행, 헬스, 보험, 엔터테인먼트, 통신, 반도체, 유틸리티, WTI, 두바이, JKM, JCC, 석탄(선물)으로 총 13개 부문이며, 가장 큰 수용자는 JKM 선물이다.

Table 1. First Connectedness result Table :

	Coal	Dubai	JCC	JKM	WTI	자동차	은행	건설	경기 소비재	필수 소비재	에너지 화학
Coal	75.83	3.17	0.71	0.95	1.86	0.88	0.84	0.92	0.62	0.99	1.32
Dubai	3	63.71	0.75	0.95	17.05	0.77	0.72	0.9	0.6	0.65	1.13
JCC	1.15	0.27	81.2	0.47	0.44	1.28	1.09	0.85	1.27	1.5	0.98
JKM	1.49	1.13	0.47	73.75	1.31	1.44	1.21	1.53	1.23	1.15	1.26
WTI	1.22	16.6	0.26	0.93	61.41	0.92	1.04	1.49	1.01	0.98	1.65
자동차	0.39	0.3	0.4	0.45	0.76	20.68	4.65	5.59	6.9	4.89	5.24
은행	0.31	0.32	0.25	0.32	0.96	4.34	19.44	5.39	5.92	4.41	4.33
건설	0.4	0.25	0.2	0.35	0.79	4.41	4.35	16.07	5.88	4.09	7.49
경기 소비재	0.2	0.14	0.33	0.25	0.76	5.19	4.77	5.73	15.14	7.64	4.8
필수 소비재	0.47	0.24	0.42	0.34	0.7	4.61	4.3	4.86	9.73	18.95	4.19
에너지 화학	0.36	0.39	0.34	0.33	1.08	4.21	3.62	7.83	5.04	3.56	16.93
헬스	0.73	0.41	0.29	0.39	0.76	3.58	2.75	5.46	6.7	5.14	5.14
보험	0.72	0.58	0.24	0.38	1.23	4.47	9.95	4.65	5.21	4.57	3.82
IT	0.36	0.3	0.29	0.42	0.7	4.55	3.38	5.24	6.09	4.13	7.59
기계장비	0.3	0.46	0.28	0.27	0.99	4.1	3.94	9.43	5.15	3.79	9.55
엔터	0.57	0.24	0.38	0.3	0.65	3.9	3.47	4.27	7.33	5.61	5.06
통신	0.39	0.36	0.31	0.41	0.93	4.04	5.29	5.4	6.9	5.57	4.66
증권	0.25	0.28	0.25	0.3	0.74	4.8	7.6	6.13	6.65	4.71	5.34
반도체	0.34	0.34	0.28	0.45	0.64	3.91	2.99	4.78	6.31	4.22	5.61
철강	0.36	0.34	0.17	0.4	0.79	4.64	5.1	8.42	5.43	4.52	7.4
운송	0.33	0.31	0.35	0.32	0.86	4.84	4.8	6.24	7.04	5.38	5.21
유틸리티	0.51	0.38	0.21	0.5	1.02	3.2	4.38	6.36	5.1	4.38	4.76

Table 2. Second Connectedness result Table :

	헬스	보험	IT	기계장비	엔터	통신	증권	반도체	철강	운송	유틸리티
Coal	1.52	1.15	1	0.96	1.19	1.05	0.63	0.96	1.17	1	1.29
Dubai	0.62	0.93	0.89	1.14	0.64	1.16	0.85	0.59	0.95	1.03	0.99
JCC	0.75	0.9	0.76	1.08	0.83	0.81	1.3	0.75	0.74	0.82	0.77
JKM	1.3	1.15	1.38	1.11	1.13	1.49	1.38	1.11	1.47	1.16	1.36
WTI	1.07	1.13	1.06	1.45	0.8	1.29	1.09	0.86	1.21	0.96	1.56
자동차	3.32	3.76	5.94	4.86	3.88	4.19	6.17	4.36	5.41	5.2	2.64
은행	2.33	7.94	3.98	4.65	3.36	5.18	9.44	3.01	5.77	4.85	3.5
건설	3.82	3.21	5.1	8.88	3.21	4.39	6.05	4.01	7.65	5.21	4.2
경기 소비재	4.63	3.47	5.83	4.8	5.35	5.28	6.5	5.25	4.93	5.74	3.26
필수 소비재	4.41	3.53	4.87	4.21	5.08	5.31	5.67	4.29	4.94	5.51	3.35
에너지 화학	3.63	2.65	7.78	9.32	3.86	3.91	5.38	4.81	7.21	4.47	3.28
헬스	22.24	1.75	6.7	4.62	6.51	4.36	5.08	6.51	3.71	4.66	2.51
보험	1.71	26.21	2.66	3.58	1.65	4.91	7.81	1.94	5.37	4.01	4.33
IT	4.73	1.83	15.66	5.38	6.11	4.29	5.76	11.74	4.7	4.13	2.6
기계장비	3.36	2.64	5.7	17.14	3.27	3.87	5.59	3.88	7.6	4.81	3.86
엔터	6	1.54	8	4.13	21.02	5.08	5.81	6.5	3.48	4.31	2.36
통신	3.85	4.07	5.05	4.54	4.78	20.39	6.03	4.4	4.4	4.59	3.64
증권	3.6	4.98	5.64	5.48	4.45	4.83	15.43	4.62	5.74	5.01	3.16
반도체	5.43	1.6	13.96	4.31	5.92	4.31	5.44	18.61	3.91	4.09	2.56
철강	2.77	3.8	5.05	7.65	2.84	3.89	6.31	3.67	17.74	4.72	3.99
운송	3.98	3.25	4.93	5.47	3.88	4.39	6.01	4.22	5.2	19.11	3.88
유틸리티	2.74	4.2	3.83	5.53	2.84	4.52	4.85	3.1	5.5	5.04	27.03

## References

- Zhu, B., Deng, Y., Hu, X. (2023). Global energy security: Do internal and external risk spillovers matter? A multilayer network method, *Energy Economics*, 126, 106961.
- Lee, Y. J., Yoon, S. M. (2022). Impact of oil price shocks on stock prices by industry, *Environmental and Resource Economics Review*, 31(2), 233-260. (in Korean).
- Cui, J., Goh, M., Li, B., Zou, H. (2021). Dynamic dependence and risk connectedness among oil and stock markets: new evidence from time-frequency domain perspectives, *Energy*, 216, 119302.
- Bahmani-Oskooee, M., Ghodsi, S. H., Hadzic, M. (2019). Asymmetric causality between oil price and stock returns: A sectoral analysis. *Economic Analysis and Policy*, 63, 165-174.
- Wang, Z. X., Liu, B. Y., Fan, Y. (2023). Network connectedness between China's crude oil futures and sector stock indices, *Energy Economics*, 125, 106848.
- Pesaran, H. H., Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models, *Economics Letters*, 58(1), 17-29.
- Diebold, F. X., Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms, *Journal of Econometrics*, 182(1), 119-134.
- Antonakakis, N., Chatziantoniou, I., Gabauer, D. (2020). Refined measures of dynamic connectedness based on Time-Varying Parameter Vector Autoregressions, *Journal of Risk and Financial Management*, 13(4), 84.
- Koop, G., Korobilis, D. (2014). A new index of financial conditions, *European Economic Review*, 71, 101 - 116.

## Connectedness Analysis of Major Energy Futures and KRX Sector Indices

*Jihoon Lim<sup>1</sup>, Sung Y. Park<sup>2</sup>*

### Abstract

From the outbreak of Covid 19, the importance of the energy security is one of the emerging topics under condition of the increasing international uncertainty following the accidents like Russia-Ukraine war and Israel-Hamas war. Countries should monitor the energy channel to keep their industrial ecosystem work without malfunction under the extreme international uncertainty these days. But, most of the researches focus on the effect of energy price shock on the industry sectors, and the research containing the bi-directional analysis hasn't done enough. Since the price of energy and industry condition interact each other, the bi-directional analysis would be more appropriate to examine them. So, Adopting the Antonakakis, Chatziantoniou and Gabauer's methodology of connectedness network framework which is refined by using TVP-VAR, I conducted the connectedness analysis of the network of oil, coal and natural gas futures and KRX sector indices to reveal the bi-directional connectedness of spillover. The static network analysis result revealed that the energy futures are receiver of the spillover in the network, and there is an heterogeneity among the KRX sector indices.

*Keywords:* Network Connectedness, Spillover effect, TVP-VAR, Variance decomposition.

---

<sup>1</sup>Graduate student, School of Economics, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul, 06974, Republic of Korea. Email: jihoon700@cau.ac.kr

<sup>2</sup>(Corresponding Author) Professor, School of Economics, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul, 06974, Republic of Korea. Email: sungpark@cau.ac.kr

## 거시경제지표를 이용한 이산화탄소 배출량 예측

김민성<sup>1</sup>, 박성용<sup>2</sup>

### 요약

본 논문에서는 연간 이산화탄소 배출량을 예측하기 위해서 대한민국의 56개의 거시경제지표를 사용한다. 해당 56개의 거시경제지표는 FRED-MD와 유사한 대한민국의 1985년부터 2022년까지의 거시경제지표를 사용하였다. 본 논문에서는 모형을 구성하기 위해서 XGboost, CSR, 랜덤포레스트 등 다양한 머신러닝 기법의 앙상블 범주의 모형을 사용한다. 특히 다수의 변수 중 연간 이산화탄소 배출량 예측에 가장 큰 영향을 끼치는 변수를 찾기 위해서 Random Forest 모형을 기반으로 한 변수선택 기법인 보루타 알고리즘(Boruta algorithm)을 사용한다. 또한 각 예측모형별 다기간 예측을 통해서 단변량 다변량의 모형 중 연간 이산화탄소 배출 예측에 가장 우수한 성능을 보이는 모형을 판별한다. 이 결과 단기 예측에서 랜덤포레스트(Random Forest) 모형과 보루타 변수에 의한 랜덤포레스트(Boruta Selected Random Forest) 모형이 가장 우수한 예측 성능을 보이는 것을 확인했으며, 보루타 알고리즘(Boruta algorithm)을 통한 이산화탄소 배출량 예측을 위한 변수 선택에 있어서 전기,가스,증기 및 공기조절 공급업 생산지수 (PI\_E), 전기,가스,증기 및 공기조절 공급업 생산자제품 출하지수(SI\_E)가 중요한 변수로 선택되어 모형의 예측 성능을 높이는 것으로 나타났다.

주요 용어: 거시경제지표, 이산화탄소 배출량 예측, 머신러닝, 보루타 알고리즘(Boruta algorithm)

<sup>1</sup>06974 대한민국 서울시 동작구 흑석로 84, 중앙대학교 경제학부 석사과정. E-mail: ko99324@cau.ac.kr

<sup>2</sup>(교신저자) 06974 대한민국 서울시 동작구 흑석로 84, 중앙대학교 경제학부 교수. E-mail: sungpark@cau.ac.kr

## 실시간 해운 항만 데이터와 동적요인모형을 이용한 GDP Nowcasting

강경태<sup>1</sup>, 조용복<sup>2</sup>

### 요약

한 국가의 광범위한 부가가치를 다루는 GDP는 해당 경제 전체상황 파악하는데 중요한 지표로써 사용되지만, 필연적으로 방대한 자료를 다루기에 현시점의 정확한 GDP 값을 확인하기 위해서 많은 시간이 소요된다. 이 문제를 완화하고자 분기별로 발표되는 GDP보다 더 짧은 주기로 공표되는 자료를 이용해 실시간에 가깝게 추정하는 Nowcasting 방법론에 관한 연구가 이루어지고 있다. 그러나 기존에 사용되는 짧은 주기의 경제변수에도 공표 시점 불일치 문제가 존재하며, 최근 해당 문제의 완화를 위해 실시간 빅데이터를 활용한 다방면의 시도가 이루어지고 있다. 따라서 본 연구에서는 그 중 수출입 의존도가 높은 국내 경제 특성을 반영하여 실시간 항만 물동량 정보의 Nowcasting 성능에 대한 검증을 실시한다. 이를 근거로 Nowcasting 모형 중 세계 중앙은행 등에서 널리 사용하고 있는 동적요인모형(DFM :Dynamic Factor Model) 을 기반으로 기존 거시경제 지표와 항만 물동량 데이터를 추가한 모형의 예측성능을 확인하였다. 그 결과 항만 물동량 정보는 GDP지표에 선행하며 기존 DFM모형의 예측오차를 줄이는데 기여할 수 있음을 확인하였다.

주요 용어: GDP, Nowcasting, 실시간 항만 물동량, Dynamic Factor Model.

---

<sup>1</sup>49236 부산광역시 서구 구덕로 225, 동아대학교 경영정보학과 학석사연계과정. E-mail: kyung10110@gmail.com  
<sup>2</sup>(교신저자) 49236 부산광역시 서구 구덕로 225, 동아대학교 경영정보학과 조교수. E-mail: inofin.lab@gmail.com

## Aging study on the Triple-GEM GE2/1 M2 detector

정용호<sup>1</sup>, 김태정<sup>2</sup>, 박인규<sup>3</sup>, 이상훈<sup>4</sup>, 양운기<sup>5</sup>, 김동현<sup>6</sup>, 강다영<sup>7</sup>

### Abstract

We are conducting a study on the aging of Triple-GEM detectors. This experiment aims to evaluate the influence of the ionization ability of particles on the lifespan of gas detectors, and we have set up an experimental environment for aging phenomena in the laboratory. The experimental environment involves injecting gas at a ratio of Ar/CO<sub>2</sub> (70:30) and simultaneously projecting 10 laser devices with voltage and current of 40kV/100uA under operating conditions to assess the presence of aging phenomena in the detector. The detector used in the experiment is the GE2/1 M2 chamber of the GEM detector. We started the aging experiment in mid-February and plan to experiment for up to one year to determine whether aging phenomena occur and to assess the performance of the detector.

---

<sup>1</sup>한양대학교 물리학과, 박사후 연구원.

<sup>2</sup>한양대학교 물리학과, 교수.

<sup>3</sup>시립대학교 물리학과, 교수.

<sup>4</sup>시립대학교 물리학과, 교수.

<sup>5</sup>서울대학교 물리학과, 교수.

<sup>6</sup>한양대학교 물리학과, 박사후 연구원.

<sup>7</sup>시립대학교 물리학과, 대학원생 연구원.

## CMS iRPC 업그레이드 완료 및 빠른 타이밍 가스 검출기 위한 신규 연구개발\*

강민호<sup>1</sup>, 이경세<sup>2</sup>, 조영민<sup>3</sup>

### 요 약

개선된 저항성 판 챔버(iRPC) 검출기의 생산과 품질 보증을 2022년부터 신뢰성 있게 수행되었습니다. iRPC 가스 캡의 생산과 품질 관리는 한양대학교 연구원들의 주도 하에 2024년 6월에 성공적으로 마무리 단계에 있습니다. 또한, iRPC 가스 캡의 방사선 내성 및 새로운 친환경 가스를 위한 연구개발은 CERN과 고려대학교의 감마 조사 시설에서 강력한 Cs-137 감마선(661.7 keV) 소스를 사용하여 성공적으로 수행되었습니다. 현재, 우리는 CMS RPC 생산에 사용된 고려대학교의 검출기 생산 시설을 이용하여 대형 빠른 타이밍 가스 검출기의 새로운 연구개발을 진행하고 있습니다. 첫 번째 연구개발 항목은 약 200 피코초(ps)의 시간 해상도를 목표로 하는 마이크로 캡 유리 RPC입니다. 두 번째 연구개발 항목은 10 kHz cm<sup>2</sup> 이상의 높은 속도 처리 능력과 200 ps보다 나은 시간 해상도를 해결하는 저항성 판의 새로운 제조 공정을 탐색하는 것입니다. 새로운 연구개발은 국제 DRD1 협력체 내에서 수행될 것입니다. (MPGD, RPC, TPC, 대형 드리프트 챔버, 와이어 및 스트로우 튜브 챔버에 전념하는 가스 검출기 연구개발 협력체).

주요 용어: CMS, 뮤온 검출기, 타이밍 RPC.

---

<sup>1</sup>한양대학교, 연구원.

<sup>2</sup>고려대학교 극한 핵물질 연구센터.

<sup>3</sup>한양대학교.

## SRGM을 사용한 딥러닝 소프트웨어 신뢰성 모형에 관한 연구

공영선<sup>1</sup>, 장인홍<sup>2</sup>, 송광윤<sup>3</sup>, 김윤수<sup>4</sup>, 이다혜<sup>5</sup>

### 요약

소프트웨어는 현대 사회의 다양한 분야에서 사용되고 있다. 만약 소프트웨어에 오류가 생기면 일상적인 생활의 작은 불편함뿐만 아니라 다양한 피해가 생길 수 있다. 예를 들면 소프트웨어 오류로 인해 방사선 의료기기가 오작동하여 인명 피해가 발생한 사건이 있다. 또 다른 사건으로 테슬라는 소프트웨어 결함이 원인이 되어 리콜로 수많은 비용을 지불했다. 이런 사고를 예방하기 위해서 다양한 소프트웨어 신뢰성 연구가 진행되고 있다. 전통적인 소프트웨어 신뢰성 모형 연구로는 비동질 포아송 과정(Nonhomogeneous Poisson process; NHPP)을 가정하는 소프트웨어 신뢰성 성장 모형(Software Reliability Growth Model; SRGM)이 있다. 딥러닝과 머신러닝을 활용하는 비모수적 관점의 소프트웨어 신뢰성 모형 연구도 활발히 진행되고 있다. 한편 딥러닝을 활용한 소프트웨어 신뢰성 모형은 수학적 근거를 제공하지 않고 데이터에만 의존하는 경향이 있다. 이를 개선하기 위해 본 연구에서는 소프트웨어 오류가 S자 형태임을 이용하여 소프트웨어 신뢰성 모형을 활성화 함수로 사용하는 딥러닝 소프트웨어 신뢰성 모형을 제시한다. 기존의 전통적인 NHPP SRGM, 일반적인 신경망 모형, 본 연구에서 제안하는 모형에 실제 고장 데이터를 적합하여 10개의 Criteria를 추정한다. 그 결과, 다른 모형에 비하여 본 연구에서 제안하는 모형의 적합도가 가장 우수하였다.

주요 용어 : 소프트웨어 신뢰성, 소프트웨어 신뢰성 성장 모형, 딥러닝, 심층신경망.

\*이 논문은 2024학년도 조선대학교 학술연구비 및 한국연구재단의 지원을 받아 진행된 논문입니다 (NRF-2021R1F1A1048592, NRF-2021R1A6A3A01086716, RS-2023-00285353).

<sup>1</sup>61452 광주광역시 동구 조선대3길 30, 조선대학교 전산통계학과, 박사과정. E-mail: konyeongseon@gmail.com

<sup>2</sup>(교신저자) 61452 광주광역시 동구 조선대3길 30, 조선대학교 컴퓨터통계학과, 교수.

E-mail: ichang@chosun.ac.kr

<sup>3</sup>61452 광주광역시 동구 조선대3길 30, 조선대학교 컴퓨터통계학과, 조교수. E-mail: cssig@chosun.ac.kr

<sup>4</sup>61452 광주광역시 동구 조선대3길 30, 조선대학교 컴퓨터통계학과, 시간강사. E-mail: imk92315@naver.com

<sup>5</sup>61452 광주광역시 동구 조선대3길 30, 조선대학교 컴퓨터통계학과, 박사후연구원. E-mail: is\_hye@chosun.ac.kr

## A Case Study of Statistical Analysis on Autonomous Vehicles Accidents Data\*

*Ki-Yeong Sim<sup>1</sup>, In-Gyu Lee<sup>2</sup>, Jung-A Yang<sup>3</sup>, Kyupil Yeon<sup>4</sup>*

### Abstract

With the advent of the smart era, the technology of autonomous vehicles (AV) continues to evolve. As autonomous vehicles are commercialized, they are expected to grow as a core technology field of the 4th industrial revolution, leading to major changes that bring new experiences to users. However, as the technology advances, the risk of autonomous vehicle accidents is inevitable. Accidents of autonomous vehicles cannot be judged as the fault of the autonomous vehicle, and traffic accidents or fatalities may occur due to unexpected circumstances or inadequate response of the technology. Therefore, this study aims to investigate the effects of accident type and environment on the extent of vehicle damage in autonomous vehicles accidents through mixed effect logistic regression. The data was collected using the Autonomous Vehicles Collision Report provided by the California DMV (Department of Motor Vehicle). As a result of the analysis, we found that autonomous driving mode tends to mitigate the damage of AV in accidents while some factors related on the accident environments such as the darker lighting condition, vehicle-to-vehicle accident, whether there were any injuries, and whether the AV has caused the accident, tends to increase the damage of a AV. We expect this study can help to better understand the causes of AV accidents and provide a guide to the development of autonomous vehicle technology.

*Keywords:* autonomous vehicles, autonomous vehicle accident, autonomous vehicle damage, California DMV, mixed effects logistic regression model.

---

\*This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (MOE) (2021RIS-004). This research was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIT) (2022M3J6A1063595).

<sup>1</sup>Graduate student, Department of Data Science, Hoseo University. E-mail: kiyeong\_23@naver.com

<sup>2</sup>Graduate student, Department of Data Science, Hoseo University. E-mail: ingy\_u\_23@naver.com

<sup>3</sup>Graduate student, Department of Data Science, Hoseo University. E-mail: junga\_23@naver.com

<sup>4</sup>(Corresponding Author) Professor, Division of Big Data and AI, Hoseo University. 20, Hoseo-ro 79beon-gil, Baebang-eup, Asan-si, ChoongNam, 31449, Korea. E-mail: kpyeon1@hoseo.edu

## 자율주행차량 사고 원인 분석을 통한 안정성 제고 방안\*

김근아<sup>1</sup>, 이은학<sup>2</sup>, 연규필<sup>3</sup>

### 요 약

본격적인 자율주행 서비스 진행에 따라 운전자의 편리성이 극대화되고 있지만, 여전히 법적, 윤리적, 사회적 측면 등 다양한 부분에서 복합적인 문제점이 나타나고 있다. 이로 인해 자율주행 차량에 대한 사회적 인식과 수용성이 충분히 형성되지 않은 상태이며 현시점에서 안정성 제고를 위한 노력은 필수적이다. 따라서 본 연구에서는 California DMV에서 제공하는 자율주행차량의 사고 데이터 중 자율주행으로 운행하던 중 사고가 난 기록만을 선별하여 약 400여 건에 대한 사고 원인을 분석하였다. 분석 결과, 주차된 차량에 대한 낮은 인식률, 편의용품 부착으로 인한 차량의 크기 변동 파악 불가능, 방어 운전 시스템의 미흡, 도로의 특수 상황에서의 낮은 인지 속도 등 여러 사고 원인을 보였다. 이를 해결하기 위해 차량 인식률 상승의 필요성, 안전 영역 범위 재구축, 방어 운전을 위한 차량의 추가 기능 제안, 자율주행차량과 스마트도로 간 소통의 중요성 강조 등의 방안을 제시함으로써 앞서 발견한 문제점 보완과 동시에 안정성을 높일 수 있을 것으로 보인다. 결과적으로, 본 연구에서는 California DMV에서 제공하는 자율주행차량의 사고 데이터를 이용하여 자율주행차량의 안정성 제고 방안을 제안하였으며 이는 추후 자율주행차량의 안정성 향상을 통한 사회적 인식과 신뢰도 상승에 긍정적 영향을 미칠 것으로 기대한다.

주요 용어 : 자율주행, 자율주행 자동차, 안정성 제고, California DMV, 차량 사고.

\*본 과제(결과물)는 2024년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과임(2021RIS-004).

\*이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2022M3J6A1063595).

<sup>1</sup>31449 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 데이터사이언스학과 석사 석사과정.  
E-mail: kimguna\_23@naver.com

<sup>2</sup>31449 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 데이터사이언스학과 석사 석사과정.  
E-mail: eunhak\_23@nvaer.com

<sup>3</sup>(교신저자)31449 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 빅데이터AI학부 교수.  
E-mail: kpyeon1@hoseo.edu

## 주식투자 관련 유튜브 개인방송과 TV방송의 허위정보 판단을 위한 비교 및 분석: 텍스트, 음성, 이미지 데이터를 활용하여

조민지<sup>1</sup>, 양성병<sup>2</sup>

### 요 약

팬데믹 기간을 거치며 개인의 주식투자 열풍이 일어나고 있는 가운데, 소셜미디어를 통한 주식투자 관련 개인방송이 급증하고 있다. 특히, 유튜브와 같은 소셜미디어에서 비공식적으로 운영되는 주식투자 관련 개인방송은 법적 제재를 받지 않는다는 점에서 TV방송과 차별화된다. 이러한 비공식 주식투자 관련 유튜브 방송은 높은 조회수를 목표로 자극적인 표현과 확인되지 않은 정보를 유포하는 경향이 있으며, 이는 주가의 급등락과 자본시장의 왜곡 문제를 초래할 수 있다. 이에, 본 연구는 주식투자 관련 유튜브 개인방송과 TV방송의 텍스트, 음성, 이미지 데이터를 비교 분석하여 이들 방송의 특성과 차이점을 규명함으로써, 허위 정보 유포로 인한 자본시장의 혼란과 이해관계자의 피해를 최소화하는 전략 마련에 기여하고자 한다. 이를 위해, 주식투자 관련 두 유형의 유튜브 방송(개인방송 vs. TV방송) 데이터를 수집한 후, 음성 데이터를 STT(Speech-to-Text) 기술을 활용하여 텍스트 데이터로 변환한 후 감성분석, 소셜네트워크분석, 토픽모델링 등의 방법론을 통해 두 방송 간의 차이를 비교한다. 다음으로, 두 방송의 음성 데이터를 Waveform 데이터로 시각화함으로써 음성 간 톤(tone color)의 차이를 시각적으로 비교 분석한다. 마지막으로, 이미지 데이터를 분석하여 두 방송 간의 시각적 특성을 비교하고, 이를 통해 방송에서 사용되는 시각적 요소들의 차이를 도출한다. 본 연구의 결과가 주식투자 관련 개인방송과 TV방송 간의 차이를 규명하고, 자본시장의 안정성을 제고하는 데 중요한 자료를 제공할 수 있기를 기대한다.

주요 용어 : 주식투자, 개인방송, TV방송, 텍스트마이닝, 음성 데이터 분석, 이미지 데이터 분석, 유튜브.

<sup>1</sup>02447 서울특별시 동대문구 경희대로 26 경희대학교 일반대학원 빅데이터융합학과 석사과정.

E-mail: chominji@khu.ac.kr

<sup>2</sup>(교신저자) 02447 서울특별시 동대문구 경희대로 26 경희대학교 경영대학 경영학과/빅데이터융합학과 교수.

E-mail: sbyang@khu.ac.kr

## 초분광영상을 활용한 대면적 문화유산 내 아까시나무 탐지

권산<sup>1</sup>, 정재경<sup>2</sup>, 어수행<sup>3</sup>, 유세종<sup>4</sup>, 김충식<sup>5</sup>, 이재용<sup>6</sup>, 최호식<sup>7</sup>

### 요약

국가지정문화재는 「문화재보호법(제44조)」에 따라 정기조사를 통해 육안으로 보존과 관리 상태를 확인한다. 그러나 면적과 규모가 큰 대면적 문화유산의 경우, 조사 및 진단에 상당한 비용과 인력이 소요되며, 지형적 제한으로 인해 접근이 어려운 구간이 존재할 수 있다. 본 연구에서는 대면적 문화재에 분포하는 위해식물의 효율적 모니터링 기술로서 드론으로 촬영된 초분광 영상의 활용 가능성을 탐색한다. 일반적인 RGB 영상은 세 가지 파장 대역(빨강, 초록, 파랑)만을 사용하는 반면, 초분광 영상은 가시광선, 근적외선, 중적외선 등 넓은 범위의 파장 대역을 사용하여 세부적인 스펙트럼 정보를 측정한다. 본 연구에서는 초분광 영상 데이터(150개 파장 대역대)를 기반으로 공주 공산성 지역에 자생하는 위해식물인 아까시나무의 분포 영역을 탐지하기 위해 XGboost를 활용한다. 일반적으로 초분광 영상자료는 촬영환경 및 보정방법에 따라 큰 잡음이 내재하고 있으며, 이는 픽셀 단위별로의 탐지 성능에 영향을 미친다. 연구에서는 성능을 향상시키기 위한 방안으로, 복수의 대역대로부터 그룹화된 특징변수와 인접한 픽셀들의 스펙트럼들로부터 추출한 특징변수를 고려한다. 훈련 자료의 비율에 따라 예측한 실증 결과, 대면적 문화유산 내 아까시나무 탐지 문제에 대한 제안하는 대역대별, 공간상의 구조적 유사성을 반영한 XGboost 방법의 학습 가능성을 확인할 수 있었다.

주요 용어 : 대면적 문화유산, XGBoost, 초분광 영상, 머신러닝.

\*이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임  
(NO.2022M3J6A1084845)

<sup>1</sup>서울특별시 동대문구 서울시립대로 163 건설공학과 3-237, 서울시립대학교 도시빅데이터 융합학과 석사과정.

<sup>2</sup>서울특별시 동대문구 서울시립대로 163 건설공학과 3-237, 서울시립대학교 도시빅데이터 융합학과 박사과정.

<sup>3</sup>서울시 관악구 관악로 1 서울대학교, 27동 404호, (주)크립토랩 연구원.

<sup>4</sup>서울특별시 강서구 화곡로68길 82, 406-408호, (주)지오스토리 과장.

<sup>5</sup>충청남도 부여군 규암면 백제문로 367, 한국전통문화대학교 전통조경학과 교수.

<sup>6</sup>충청남도 부여군 규암면 백제문로 367, 한국전통문화대학교 전통조경학과 교수.

<sup>7</sup>(교신저자) 서울특별시 동대문구 서울시립대로 163 법학관 606호, 서울시립대학교 도시빅데이터 융합학과 교수. E-mail: choi.hosik@uos.ac.kr

## 언어모델을 활용한 미국 대선 영향 사건까지의 시차 예측

백우빈<sup>1</sup>, 최인수<sup>2</sup>, 손형진<sup>1</sup>, 김장호<sup>3</sup>, 김우창<sup>2</sup>

### 요약

미국의 대통령 선거가 미치는 영향은 단순 정계에서 그치지 않고 전세계의 경제계, 언론계 등 온갖 곳까지 끼친다. 이러한 대통령 선거의 경우 소셜 네트워크 서비스가 정치 관련 사건의 분위기와 추세를 결정하는 경우가 많기 때문에 미국 대선의 경우와 같이 중요한 행사를 분석할 때 이런 소셜 네트워크 서비스의 경향의 연구는 필수적이다. 이러한 점에 착안하여 본 연구에서는 대체 데이터이자 대표적인 소셜 네트워크 서비스인 트위터의 트윗 데이터를 분석하고 초거대 언어 모델에 파인튜닝(fine-tuning)시켜 대선에 영향을 끼치는 사건이 발생하기까지의 시차를 예측하고자 하였다.

주요 용어: 대통령 선거, 거대 언어 모델, 기계 학습, 소셜 네트워크 서비스(SNS).

---

<sup>1</sup>경희대학교 산업경영공학과

<sup>2</sup>KAIST 산업및시스템공학과

<sup>3</sup>고려대학교 기술경영전문대학원

## 프롬프트 엔지니어링을 활용한 LLM 속성 기반 감성 분석에 대한 연구

손민녕<sup>1</sup>, 임희석<sup>2</sup>

### 요약

속성 기반 감성 분석(Asspect-Based Sentiment Analysis, ABSA)은 전반적인 감성 분석보다 더욱 세분화된 속성별 평가를 통해 만족, 불만족 요인을 정확하게 파악하여 비즈니스 전략에 중요한 인사이트를 제공한다. 이전에는 BERT 모델을 활용하여 ABSA를 진행하는 경우가 많았으나, 산업별로 상이한 속성을 맞추기 위해 fine-tuning이 필수적이었고, 이 과정에서 많은 비용과 시간이 소요되었다. 본 연구는 누구나 간편하고 빠르게 적용할 수 있는 속성 기반 감성 분석의 프롬프트 엔지니어링 방법을 연구하였다. AI-Hub에 공개된 화장품 리뷰 및 평점 데이터를 활용하여, 화장품 구매 시 주요 고려 요인인 가격, 사용감, 사용 행태, 제품 구성, 효능, 향을 속성으로 정하고 GPT를 활용하여 속성 기반 감성 분석을 진행하였다. 연구 결과, 프롬프트 엔지니어링 시 고려할 사항을 도출할 수 있었다. 첫째, 속성 기반 감성 분석 task를 속성 추출(Asspect Term Extraction)과 속성 감성 분석(Asspect Term Polarity) 두 단계로 나누는 것보다 더 세분화하여 작업을 정의할 시 성능이 개선됨을 확인하였다. 둘째, few-shot learning 시 여러 속성 중 추출 정확도가 낮은 속성을 중심으로 예시를 제공할 경우 해당 속성의 성능이 크게 개선되었으며, 사례와 속성에 대한 정의를 함께 제공하는 것이 더욱 효과적임을 알 수 있었다.

주요 용어 : 속성 기반 감성 분석(Asspect-Based Sentiment Analysis, ABSA), 프롬프트 엔지니어링, large language models(LLMs), GPT.

---

<sup>1</sup>02841 서울특별시 성북구 안암로 145 고려대학교 SW·AI융합대학원 빅데이터융합학과 석사과정.

E-mail: showmetes@naver.com

<sup>2</sup>(교신저자) 02841 서울특별시 성북구 안암로 145 고려대학교 정보대학 컴퓨터학과 교수.

E-mail: limhseok@korea.ac.kr

## 경제뉴스 감성 분석과 딥러닝 모델을 활용한 판매량 예측 연구

박시은<sup>1</sup>

### 요약

판매량 예측 연구는 기업의 재고관리, 마케팅 전략 수립 및 시장 트렌드 파악에 중요한 역할을 한다. 예측 성능 향상을 위해 다양한 기법들이 활용되고 있으며, 본 연구에서는 경제뉴스 감성 분석과 딥러닝 모델을 활용하여 'A'사에서 수집한 자동차용품 판매량을 예측하는 모델과 지표의 조합을 제시하고자 한다. 차량관리 및 인테리어용품을 포함한 자동차용품은 기온, 강수량과 같은 환경적 변화와 소비 심리, 경제적 상황에 민감하게 반응하는 특징을 갖고 있다. 따라서 다양한 내외부 요인과 소비자 행동 지표를 찾아 판매량 예측 변수로 반영하는 것은 중요하다. 경제뉴스 감성 분석은 금융 데이터에 최적화된 FinBERT 언어모델을 활용하여 뉴스의 제목과 본문에 대해 각각 수행하여, 감성지수를 도출하였다. 예측 모델로는 시계열 데이터의 특징을 잘 포착하는 딥러닝 모델인 RNN(Recurrent Neural Network), LSTM(Long-Short-Term Memory), GRU(Gated Recurrent Unit)을 사용하였다. 그 결과, 경제뉴스 감성지수를 포함한 예측 모델들이 판매량 예측 성능을 유의미하게 향상시켰음을 확인할 수 있었다. 또한 모델별로 최적의 지표 조합을 도출한 결과, GRU 모델에서 경제뉴스 감성지수와 내외부 요인 및 소비자 행동 지표를 모두 예측 변수로 사용할 때 가장 높은 성능을 보이는 것으로 확인할 수 있었다. 이러한 연구 결과는 경제뉴스 감성 분석이 판매량 예측 성능을 향상시킨다는 것을 정량적으로 입증하여, 뉴스 및 소비자 행동 데이터를 적극적으로 활용할 가치가 있음을 시사한다.

주요 용어: 판매량 예측, FinBERT, 감성 분석, RNN, LSTM, GRU, 딥러닝.

## A Comprehensive Analysis of Multi-hop Question Answering Datasets: Insights and Implications

*Hyunseo Shin*<sup>1</sup>, *Wonseok Hwang*<sup>2</sup>

### Abstract

Multi-hop Question Answering (QA) is an important capability for a machine to understand natural language. It requires aggregation of information from multiple documents to answer a given question. Despite significant advances in many qa datasets, multi-hop QA datasets still remain an area of difficulty and challenge. This paper evaluates the structure, characteristics, and methodology of well-known datasets, and provides analysis and resulting insights on the multi-hop QA dataset. We conducted a specific and in-depth statistical analysis to identify the common characteristics and limitations present in this dataset. Additionally, we evaluate the performance of retrievers on these datasets to analyze the future direction for the development of multi-hop QA datasets. Our findings highlight the limitations and boundaries of currently existing multi-hop QA datasets and contribute to the development of more efficient natural language processing QA systems. Ultimately, our work analyzes the impact of the difference in performance of retrievers and the quality of the dataset in multi-hop QA. We hope that these analyses will lead to better future research in QA systems.

*Keywords:* Multi-hop Question Answering, Natural Language Processing, Dataset Analysis, QA Models.

---

<sup>1</sup>Graduate Student, Department of Artificial Intelligence, University of Seoul, 163 Seoulsiripdaero, Korea. E-mail: hseo98@uos.ac.kr

<sup>2</sup>(Corresponding Author) Professor, Department of Artificial Intelligence, University of Seoul, 163 Seoulsiripdaero, Korea. E-mail: wonseok.hwang@uos.ac.kr

## NASA 그랜트 데이터를 이용한 항공우주 연구토픽 분석\*

노경란<sup>1</sup>, 한혁<sup>2</sup>

### 요약

NASA는 우주 탐사 및 과학 연구의 발전을 촉진하기 위하여 혁신적인 아이디어와 기술개발에 편당을 지원함으로써 과학적 발견과 기술 발전에 앞장서고 있다. NASA R&D는 여러 미션본부와 연구센터들을 통해 추진/관리되고 있다. NASA는 연방정부의 정보공개 정책과 법률에 기반하여 R&D 편당 데이터를 공개하고 있다. 공개된 그랜트 데이터는 일반적으로 프로젝트 제목 및 설명, 편당금액, 수행기관, 연구기간, 프로그램 이름을 포함하고 있다. 본 연구에서는 NASA 그랜트 데이터에 수록된 프로젝트의 제목과 설명, 프로그램 이름을 이용하여 토픽모델을 적용하여 연구 토픽을 추출하였다. 분석을 통해 NASA가 중점을 두고 지원하는 우주항공 연구 분야와 우선순위를 파악하였다. 그랜트 데이터는 미래의 연구 방향을 예측하는 중요한 역할을 한다. 어떤 분야에 자금이 많이 할당되었는지, 어떤 기술이나 연구가 중점적으로 추진되고 있는지를 이해하면, 연구자와 기관이 향후 연구 방향을 설정하는 데 도움을 줄 수 있을 것이다.

주요 용어: 토픽모델링, 리서치 포트폴리오, 항공우주, 연구토픽, 편당 데이터, 그랜트 데이터.

---

\*이 논문은 2024년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다.(과제번호: K24L3M2C4)  
<sup>1</sup>02456 서울특별시 동대문구 회기로66 한국과학기술정보연구원 R&D투자전략연구소 투자분석팀 책임연구원.  
E-mail: infor@kisti.re.kr  
<sup>2</sup>02456 서울특별시 동대문구 회기로66 한국과학기술정보연구원 R&D투자전략연구소 투자분석팀 책임연구원.  
E-mail: hyhahn@kisti.re.kr

## 대전시 공공자전거의 효율적 운영을 위한 수요예측\*

박정연<sup>1</sup>, 이상인<sup>2</sup>

### 요약

이 논문은 대전시의 공공자전거 이용 현황과 특성을 파악하고 대여소별 수요 예측 모델을 제시하고 있다. 최근에는 공유 전동 킥보드, 공유 전기 자전거 등의 경쟁적인 공유 개인형 모빌리티 서비스들의 이용이 증가하면서, 대전시의 자전거 공유 시스템인 '타슈'의 운영 개선이 필요한 시점이다. 현재 모든 타슈의 정비는 교통공사 타슈센터에서 이루어지고 있어, 센터와 대여소 간 원거리로 인한 불편과 정비량 등의 문제가 발생하고 있다. 이에 대한 대책으로 타슈의 이용 특성 분석과 시공간적 패턴에 따른 군집분석 결과를 확인하여 군집별 특성을 파악하였다. 또한, 대여건수를 예측하는 모델을 생성하였는데, 회귀분석과 음이항 회귀분석을 활용하였으며, 대여소별과 시간대별 특성을 고려하기 위해 확률효과모형을 사용하였다. 이러한 분석을 통해 대전시의 공공자전거 서비스를 보다 효율적으로 운영하고 이용자들에게 더 나은 서비스를 제공할 수 있는 방안을 모색하고자 한다.

주요 용어 : 공유자전거, 군집분석, 선형회귀, 음이항회귀, 확률효과.

---

\*이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 연구되었음 (NRF-2022M3J6A1084843).

<sup>1</sup>34134 대전광역시 유성구 대학로 99, 충남대학교 통계데이터사이언스학과, 석사과정.

E-mail: wjddus3325@naver.com

<sup>2</sup>34134 대전광역시 유성구 대학로 99, 충남대학교 정보통계학과, 부교수. E-mail: sanginlee44@gmail.com

## 고등학교 학업성취에 영향을 미치는 요인 분석 및 예측 지표 개발: Propensity Score Matching과 인과추론 접근

이창현<sup>1</sup>

### 요 약

본 연구는 고등학교 학생들의 학업성취도에 영향을 미치는 요인들을 분석하고, 이를 기반으로 예측 지표를 개발하는 것을 목표로 한다. 연구는 포르투갈 고등학생 데이터를 사용하여 진행되었으며, 다양한 사회경제적 요인들을 고려하였다. PSM(=Propensity Score Matching) 기법을 통해 데이터를 실험군과 대조군으로 나누었으며, 카이제곱 테스트를 사용하여 각 요인의 유의미성을 평가하였다. 분석 결과, 학업성취도의 유의미한 영향을 미치는 주요 요인들은 다음과 같다: 부모의 교육 수준, 학생의 실패 경험, 추가적인 사교육 여부 등의 요인이 있다. 이러한 유의미한 요인들을 바탕으로 학업성취 예측 지표 EPI(= Educational Potential Index)를 개발하였다. 예측 지표는 각 유의미한 변수의 값과 해당 변수의 Lift를 곱한 값을 합산하여 산출한다. 이 지표를 통해 학생의 학업성취도를 평가하고, 낮은 점수를 보이는 학생들이 어떤 요인으로 인해 성취도가 낮은지를 진단할 수 있다. 이를 바탕으로 맞춤형 조치를 통해 학생들의 학업 성취를 개선할 수 있다. 비록 본 연구는 포르투갈 데이터를 기반으로 하였으나, 한국에서도 유사한 데이터셋이 존재한다면 동일한 방법론을 적용하여 학업성취 모델링을 수행할 수 있을 것으로 기대된다. 본 연구는 교육 정책 수립 및 학생 맞춤형 지원 방안 마련에 기여할 수 있는 중요한 자료를 제공할 것이다.

$$EPI = \frac{\sum_{k=1}(\alpha_k) * Var_k}{\sum_{k=1} \alpha_k}, \text{ where } EPI \text{ is between } 0 \text{ and } 1, \alpha_k \text{ is that weight for each } Var_k$$

그림 1. 학업성취 예측지표 EPI 산정방식

주요 용어: 고등학교 학업성취 능력, Propensity Score Matching, 인과추론, 지표 모델링.

## 2단계 집락 조건부 Quatember 확률화응답모형\*

이기성<sup>1</sup>, 홍기학<sup>2</sup>, 손창균<sup>3</sup>, 박근화<sup>4</sup>, 홍성준<sup>5</sup>

### 요약

본 논문에서는 모집단이 집락으로 구성되어 있고 얻고자 하는 속성이 민감할 때, 민감한 속성을 효율적으로 추정할 수 있는 조건부 Quatember(2009) 모형에 2단계 집락추출법을 적용한 2단계 집락 조건부 Quatember 모형을 제안하였다.

주요 용어: 덜 민감한 속성, 민감한 속성, Quatember 모형, 2단계 집락추출, 조건부 모형.

### 1. 서론

Warner(1965)는 민감한 질문 그리고 민감한 질문과 배반되는 질문으로 구성된 확률장치를 이용하여 민감한 속성을 추정하였다. 그 후 Quatember(2009)는 5개의 질문 문항 즉, 민감한 질문, 민감한 질문과 배반되는 질문, 무관한 질문, 예, 아니오로 구성된 새로운 형태의 확률화응답모형을 제안하였으며, Lee et al.(2023)은 Quatember 모형에 Carr, Marascuilo.(1982)의 조건부 모형을 적용한 조건부 Quatember 모형을 제안하여 총화추정이 가능하도록 총화 조건부 Quatember 모형으로 발전시켰다.

본 논문에서는 모집단이 집락으로 구성되어 있고 얻고자 하는 속성이 민감할 때, 민감한 속성을 효율적으로 추정할 수 있는 조건부 Quatember 모형에 2단계 집락추출법을 적용한 2단계 집락 조건부 Quatember 모형을 제안하고자 한다. 그리고 제안한 2단계 집락 조건부 Quatember 모형이 Quatember 모형에 2단계 집락추출법을 적용한 2단계 집락 Quatember 모형보다 효율적이 되는 조건을 제시하고자 한다.

### 2. 2단계 집락 조건부 Quatember 확률화응답모형

매우 민감한 조사에서 각 집락의 크기가  $M_i$  ( $i = 1, 2, \dots, N$ ) 인  $N$ 개의 집락으로 구성되어 있는 모집단으로부터  $n$  개의 집락을 단순임의비복원추출한 후, 추출된 각 집락에서 다시

\*이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2022R1F1A1063263).

<sup>1</sup>55338 전북 완주군 삼례읍 삼례로 443 우석대학교 아동사회복지학부 교수. E-mail: gisung@woosuk.ac.kr

<sup>2</sup>58245 전남 나주시 건재로 185 동신대학교 컴퓨터학과 교수. E-mail: khhong@dsu.ac.kr

<sup>3</sup>38066 경북 경주시 동대로 123 동국대학교 빅데이터·응용통계학과 교수. E-mail: ckson85@dongguk.ac.kr

<sup>4</sup>07511 서울특별시 강서구 금남화로 154 한국문화관광연구원 선임전문위원. E-mail: pkhkwen@kcti.re.kr

<sup>5</sup>13822 경기도 과천시 경마공원대로 107 마사회 말산업연구소 연구원. E-mail: hsj8129@naver.com

$m_i$  ( $i = 1, 2, \dots, n$ ) 개의 조사단위의 표본을 단순임의복원추출하는 2단계 집락추출법을 조건부 Quatember 모형에 적용해 보고자 한다.

$i$  번째 집락의 첫 번째 단계에서  $m_i$  명의 응답자에게 다음과 같은 2개의 질문으로 구성된 강요 질문모형의 확률장치를 통해 선택된 질문에 대하여 응답하도록 한다.

	질문내용	선택확률
질문 1	당신은 덜 민감한 속성 $B_i$ 를 가지고 있습니까?	$T_i$
질문 2	“예”라고 응답하십시오.	$1 - T_i$

이때, 응답자들은 확률장치에 의해서 질문 1이 선택되면 “예” 또는 “아니오”라고 응답하고, 질문 2가 선택되면 “예”라고 응답하게 된다.

$i$  번째 집락의 첫 번째 단계에서 응답자가 “예”라고 응답할 확률을  $\lambda_{i1}$  라 하면, 그 확률은 다음과 같다.

$$\lambda_{i1} = T_i \pi_{i1} + (1 - T_i).$$

여기서,  $\pi_{i1}$  은  $i$  번째 집락의 덜 민감한 속성  $B_i$  에 대한 모비율이다.

이때,  $i$  번째 집락의  $m_i$  명의 응답자들 중에서 “예”라고 응답한 사람의 수를  $m_{i1}$  명이라 하면,  $\hat{\lambda}_{i1} = \frac{m_{i1}}{m_i}$  이 되므로 덜 민감한 속성  $B_i$  에 대한 모비율  $\pi_{i1}$  의 추정량  $\hat{\pi}_{i1}$  는 다음과 같다.

$$\hat{\pi}_{i1} = \frac{1}{T_i} [\hat{\lambda}_{i1} - (1 - T_i)] = \frac{1}{m_i T_i} [m_{i1} - (1 - T_i) m_i].$$

두 번째 단계에서는  $i$  번째 집락의 첫 번째 단계에서 “예”라고 응답한  $m_{i1}$  명의 응답자에게만 다음과 같은 5개의 질문으로 구성된 Quatember 모형의 확률장치를 통해 선택된 질문에 대하여 “예” 또는 “아니오”라고 응답하도록 한다.

	질문내용	선택확률
질문 1	당신은 민감한 속성 $A$ 를 가지고 있습니까?	$p_{i1}$
질문 2	당신은 민감한 속성 $A$ 를 가지고 있지 않습니까?	$p_{i2}$
질문 3	당신은 무관한 속성 $Y$ 를 가지고 있습니까?	$p_{i3}$
질문 4	“예”라고 응답하세요.	$p_{i4}$
질문 5	“아니오”라고 응답하세요.	$p_{i5}$

따라서  $i$  번째 집락의 응답자가 첫 번째 단계에서 “예”라고 응답했다는 조건하에서 두 번째 단계에서 “예”라고 응답할 확률을  $\lambda_{i2}$  라 하면, 그 확률은 다음과 같다.

$$\lambda_{i2} = (p_{i1} - p_{i2}) \frac{\pi_{i2}}{\lambda_{i1}} + (p_{i2} + p_{i3} \pi_{iy} + p_{i4}).$$

여기서,  $\pi_{i2}$  는  $i$  번째 집락의 민감한 속성에 대한 모비율이며,  $\pi_{iy}$  는  $i$  번째 집락의 무관한 속성에 대한 모비율이며 알고 있다고 가정한다.

$i$  번째 집락의  $m_{i1}$  명의 응답자들 중에서 “예”라고 응답한 사람의 수를  $m_{i2}$  라 하면  $\hat{\lambda}_{i2} = \frac{m_{i2}}{m_{i1}}$  가 되므로  $i$  번째 집락의 민감한 속성에 대한 모비율  $\pi_{i2}$ 의 추정량  $\hat{\pi}_{i2}$ 는 다음과 같다.

$$\hat{\pi}_{i2} = \frac{1}{m_i(p_{i1} - p_{i2})} [m_{i2} - (p_{i2} + p_{i3}\pi_{iy} + p_{i4})m_{i1}], \quad (p_{i1} \neq p_{i2}).$$

이 때,  $\hat{\pi}_{i2}$ 의 기대값과 분산은 다음과 같다.

$$E(\hat{\pi}_{i2}) = \pi_{i2},$$

$$V(\hat{\pi}_{i2}) = \frac{1}{m_i} \left[ \pi_{i2}(1 - \pi_{i2}) + \frac{\pi_{i2}(1 - p_{i1} + p_{i2} - A_i)}{p_{i1} - p_{i2}} + \frac{\{T_i\pi_{i1} + (1 - T_i)\}A_i(1 - A_i)}{(p_{i1} - p_{i2})^2} \right].$$

여기서  $A_i = p_{i2} + p_{i3}\pi_{iy} + p_{i4}$  이다.

한편, 2단계 집락추출법에 의한 민감한 속성의 모비율  $\pi_2$ 는 다음과 같이 표현할 수 있다.

$$\pi_2 = \frac{1}{M_0} \sum_{i=1}^N M_i \pi_{i2}, \quad M_0 = \sum_{i=1}^N M_i.$$

응답자들이  $i$  ( $i = 1, 2, \dots, n$ ) 번째 집락으로부터 단순임의복원추출되었을 때, 이러한 절차에 의해 얻어진 민감한 속성에 대한 모비율  $\pi_2$ 의 추정량  $\hat{\pi}_2$ 는 다음과 같이 정의할 수 있다.

$$\hat{\pi}_2 = \frac{N}{nM_0} \sum_{i=1}^n M_i \hat{\pi}_{i2} = \frac{1}{nM} \sum_{i=1}^n M_i \hat{\pi}_{i2}, \quad \bar{M} = \frac{M_0}{N}$$

<정리 1> 추정량  $\hat{\pi}_2$ 는 모비율  $\pi_2$ 의 비편향추정량이다.

<정리 2> 추정량  $\hat{\pi}_2$ 의 분산은 다음과 같다.

$$V(\hat{\pi}_2) = \frac{N-n}{nN(N-1)} \sum_{i=1}^N \left( \frac{M_i \pi_{i2}}{M} - \pi_2 \right)^2 \quad (2.1)$$

$$+ \frac{1}{nNM^2} \sum_{i=1}^N M_i^2 \frac{1}{m_i} \left[ \pi_{i2}(1 - \pi_{i2}) + \frac{\pi_{i2}(1 - p_{i1} + p_{i2} - A_i)}{p_{i1} - p_{i2}} + \frac{\{T_i\pi_{i1} + (1 - T_i)\}A_i(1 - A_i)}{(p_{i1} - p_{i2})^2} \right].$$

### 3. 효율성 비교

이 절에서는 제안한 2단계 집락 조건부 Quatember 모형이 Quatember(2009) 모형에 2단계 집락추출법을 적용한 2단계 집락 Quatember 모형보다 효율적이 되는 조건을 제시하고자 한다. 2단계 집

락 Quatember 모형에서 민감한 속성에 대한 모비율  $\pi_2$ 의 추정량  $\hat{\pi}_q$ 의 분산은 다음과 같다.

$$V(\hat{\pi}_g) = \frac{N-n}{nN(N-1)} \sum_{i=1}^N \left( \frac{M_i \pi_{i2}}{M} - \pi_2 \right)^2 \quad (4.1)$$

$$+ \frac{1}{nNM^2} \sum_{i=1}^N \frac{M_i^2}{m_i} \left[ \pi_{i2}(1-\pi_{i2}) + \frac{\pi_{i2}(1-p_{i1}+p_{i2}-A_i)}{p_{i1}-p_{i2}} + \frac{A_i(1-A_i)}{(p_{i1}-p_{i2})^2} \right], \quad p_{i1} \neq p_{i2}.$$

제안한 2단계 집락 조건부 Quatember 모형의 분산 식(2.1)과 2단계 집락 Quatember 모형의 분산 식(4.1)을 이용하여  $V(\hat{\pi}_q) - V(\hat{\pi}_2)$ 를 구해보도록 하자.

$$V(\hat{\pi}_q) - V(\hat{\pi}_2) = \frac{1}{nNM^2} \sum_{i=1}^N M_i^2 \frac{1}{m_i} \frac{T_i(1-\pi_{i1})A_i(1-A_i)}{(p_{i1}-p_{i2})^2}. \quad (4.2)$$

식(4.2)는  $\pi_{i1} = 1$ 일 때  $V(\hat{\pi}_q) = V(\hat{\pi}_2)$ 이 성립되며,  $V(\hat{\pi}_2) < V(\hat{\pi}_q)$ 를 만족하는 조건을 구해보면 다음과 같다.

$$\pi_{i1} < 1. \quad (4.3)$$

이 조건은 일반적으로 성립되므로 제안한 2단계 집락 조건부 Quatember 모형은 2단계 집락 Quatember 모형에 비해 응답절차 면에서는 다소 복잡하지만, 효율성 면에서는 우월함을 알 수 있다.

#### 4. 결론

본 논문에서는 모집단이 집락으로 구성되어 있고 얻고자 하는 속성이 민감할 때, 민감한 속성을 효율적으로 추정할 수 있는 조건부 Quatember 모형에 2단계 집락추출법을 적용한 2단계 집락 조건부 Quatember 모형을 제안하였다. 그리고 제안한 2단계 집락 조건부 Quatember 모형이 Quatember 모형에 2단계 집락추출법을 적용한 2단계 집락 Quatember 모형보다 효율적이 되는 조건을 제시하였다.

#### Reference

- Carr, J. W., Marascuilo, L. A. (1982). Optimal randomized response models and methods for hypothesis testing, *Journal of Educational Statistics*, 7, 295-310.
- Lee, G. S., Hong, K. H., Son, C. K., Park, K. H., Hong, S. J. (2023). A stratified conditional Quatember's randomized response model, *Proceedings of the Korean Data Analysis Society*, January 26-27, 181-184. (in Korean).
- Quatember, A. (2009). A standardization of randomized response strategies, *Survey Methodology*, 35(2), 143-152.
- Warner, S. L. (1965). Randomized response ; A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, 60, 63-69.

## 주가 예측을 위한 LSTM, MAMBA 모델의 하이퍼파라미터 최적화 예비 연구

김희연<sup>1</sup>, 양하늘<sup>1</sup>, 정민서<sup>1</sup>

### 요약

세계 경제 불확실성이 확대되면서 주식에 대한 투자 심리로 최근 개인 투자자의 국내 주식시장 매수세와 해외 주식 투자가 빠르게 증가하고 있다. 주가 예측을 위하여 과거 주가 동향을 파악하는 것이 중요시됨으로 본 논문에서는 예측 모델의 향상을 위해 LSTM 모델과 MAMBA 모델의 하이퍼파라미터를 변경하여 모델들의 손실 값 변화를 통해 분석한다. 그 결과, 모델 성능의 향상과 최적의 예측 값을 도출하기 위해서는 하이퍼파라미터의 적절한 값을 찾는 과정의 필요성을 확인하였다. 후에 주식에 영향을 미치는 요인들을 파악하고 이를 적용하여 주가 예측 모델의 성능을 향상시키며, 최적의 하이퍼파라미터의 값을 도출할 것으로 기대된다.

주요용어 : 주식 예측, LSTM 모델, MAMBA 모델, 하이퍼파라미터, 최적화.

### 1. 서론

투자는 더 이상 전문가들만의 영역이 아니게 되었고, 일반인들도 자산관리를 위해 쉽게 접할 수 있는 활동이다. 주가를 분석하게 된 계기 중 하나는 바로 이러한 접근성의 확대와 개인 투자자들의 증가이다. 미래의 주가를 예측할 때 과거의 주가 동향을 분석하는 것은 투자 결정을 내릴 때 매우 중요하며, 예측의 정확도를 좌우한다. 과거 정보 기억을 바탕으로 데이터를 활용하여 장기 의존성 문제를 해결해주는 LSTM(Long Short-Term Memory) 모델과 최신 모델 Mamba를 사용하여 삼성전자와 애플의 주가 분석을 통한 주가 예측 연구를 진행했다. 2021년 5월 1일부터 2024년 5월 1일까지의 주가 데이터를 사용하여 각 주가의 최고가와 최저가를 예측할 수 있도록 모델을 구성했다. 과거의 주가 데이터를 입력으로 하여 모델을 학습시킬 시 모델의 성능에 영향을 미치는 하이퍼파라미터들을 변경해가며 학습시켰다. 두 모델 모두 에폭은 100으로 고정한 상태로 배치 사이즈를 32 64 128, 학습률을 0.1 0.01 0.001로 변경하였고, LSTM 모델에 대해서는 추가적으로 뉴런의 개수를 32 64 128로 변경을 진행했다. 최종적으로 테스트 손실 값이 최소로 나오는 최적의 하이퍼파라미터를 찾는 작업을 수행하였다.

### 2. 배경지식

LSTM(Long Short-Term Memory) 모델은 장/단기 기억을 가능하게 설계된 신경망으로, 시퀀스 데이터의 특성을 통하여 시간적 패턴과 의존성을 학습하고 데이터를 예측하는데 유용한 모델이다[1].

<sup>1</sup>성신여자대학교 AI융합학과

장기 의존성으로 시계열 데이터를 효과적으로 학습할 수 있으며 게이트 메커니즘을 활용하여 중요한 정보를 추출하여 정확한 패턴 학습이 가능하다. 이에 예측에 필요한 장기적인 패턴 학습이 가능하므로 시계열 데이터 문제 해결 시 유용하게 활용된다[2].

Mamba(Measurement Architecture for Model-Based analysis) 모델은 transformer 수준의 성능을 달성하는 최초의 선형 시간 시퀀스 모델이다. 이는 강력한 최신 transformer 훈련을 포함하여 광범위한 기준의 성능을 뛰어넘는다[3]. 또한, 소프트웨어 시스템의 정적 및 동적 모델 기반 분석을 위한 통합 측정 아키텍처로 임의의 통계 집계 함수 및 주기적 집계 함수와 같은 추가 모델링 기능을 활용한다[4].

### 3. 데이터 및 실험

#### 3.1. 데이터 세트

본 실험에서는 파이썬 금융 데이터 분석 라이브러리인 FinanceDataReader를 사용해 애플과 삼성전자 주가 데이터를 활용했다. 해당 라이브러리의 데이터는 날짜, 시가, 고가, 저가, 종가, 거래량과 전날 대비 종가 상승률로 구성되어 있다. 실험에는 2021년 5월 1일부터 2024년 5월 1일까지의 데이터를 사용하였고, 20일간의 주가 데이터로부터 그 다음날의 주가를 예측하도록 하였다. 그 중 2021년 5월 1일부터 2023년 12월 31일까지의 데이터는 훈련 데이터로, 2024년 1월 2일부터 2024년 5월 1일까지의 데이터를 테스트 데이터로 사용하였다.

#### 3.2. 하이퍼파라미터

하이퍼파라미터의 설정에 따라 기계학습에서 학습 성능이 달라지므로 좋은 성능을 위해서는 최적의 하이퍼파라미터 값을 결정해야 한다[5][6]. 본 연구에서는 최적의 값을 찾기 위해 LSTM의 뉴런 수, 학습률과 배치 크기를 조절해가며 학습시키며 MAMBA의 경우 학습률과 배치 크기를 조절하였다. 민감성이 높은 LSTM의 뉴런 수[7]는 32, 64, 128로 설정하였다. LSTM과 MAMBA 모델의 공통적으로 훈련 결과에 큰 영향을 주는 학습률[8]은 0.1, 0.01, 0.001, 배치 크기는 32, 64, 128로 조절하였고, 에폭은 100으로 고정하여 학습을 진행했다.

### 4. 실험 결과

삼성전자와 애플의 최고가와 최저가의 예측을 LSTM모델에 대해서는 표1과 같이, Mamba 모델에 대해서는 표2와 같이 각 하이퍼파라미터를 변경해가며 진행했다. 각 주가 데이터와 하이퍼파라미터에 대해 모델을 3번 돌린 후 출력된 손실 값들의 평균으로 평가하였다. LSTM 모델에 대해서는 어떠한 배치 크기에서도 학습률이 0.01일 때 최적의 결과를 도출했다. 반면에 배치 크기가 32, 그리고 뉴런의 개수가 128일 때 가장 좋지 않은 결과가 도출됐다. 그림1에 의하면 삼성전자 주가의 최고가와 최저가 예측 시 최적의 하이퍼파라미터는 배치 크기 128, 학습률 0.01, 그리고 뉴런의 개수 32일 때이다. 그림2에 의하면 애플 주가의 최고가와 최저가 예측 시 최적의 하이퍼파라미터는 배치 크기 64, 학습률 0.01, 그리고 뉴런의 개수 64일 때이다. Mamba 모델에 대해서는 삼성전자 주가 예측 시 최적의 하이퍼파라미터는 그림3에 의해 배치 크기 32와 학습률 0.001일 때였

고, 애플 주가 예측 시 최적의 하이퍼파라미터는 그림4에 의해 배치 사이즈 64일 때 0.01의 학습률로 학습할 때였다.

표 1. LSTM - 하이퍼파라미터에 따른 성능 비교 결과

항목	학습률	뉴런 수	배치 사이즈		
			32	64	128
삼성	0.1	32	0.1824459210	0.1571423487	0.1649079502
		64	0.1776711718	0.1549473156	0.1584286494
		128	0.2989856749	0.2180268462	0.1841595218
	0.01	32	0.1073294543	0.1030279945	0.0961694581
		64	0.1027349978	0.1033362113	0.0970431715
		128	0.1151046194	0.1090837514	0.0993253925
	0.001	32	0.1454169751	0.1233162843	0.1429741532
		64	0.1559162959	0.1107413903	0.1204885592
		128	0.1429121941	0.1066929015	0.1107351072
애플	0.1	32	0.1710671977	0.1390334297	0.2183267243
		64	0.1161121354	0.1211809193	0.1432767800
		128	0.1929929197	0.1391353561	0.1451053491
	0.01	32	0.1256902208	0.0981947786	0.1154577996
		64	0.0878827385	0.0832032959	0.0908257303
		128	0.0915439614	0.0874825878	0.0891636834
	0.001	32	0.1259855621	0.1208976529	0.1316848918
		64	0.1355016902	0.1016799062	0.1201246195
		128	0.1354429573	0.0947374988	0.1162650200

표 2. Mamba - 하이퍼파라미터에 따른 성능 비교 결과

항목	학습률	배치 사이즈		
		32	64	128
삼성	0.1	0.0676995	0.0076235	0.0057555
	0.01	0.0017226	0.0014767	0.0014627
	0.001	0.0013486	0.0017970	0.0019709
애플	0.1	0.0067737	0.0025629	0.0013644
	0.01	0.0011373	0.0010073	0.0013708
	0.001	0.0010412	0.0011315	0.0018570



그림 1. LSTM - 삼성전자 주가 예측

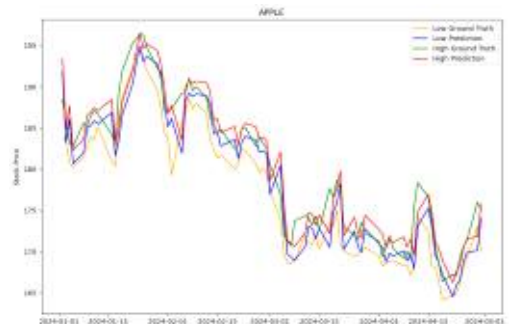


그림 2. LSTM - 애플 주가 예측



그림 3. MAMBA - 삼성전자 주가 예측

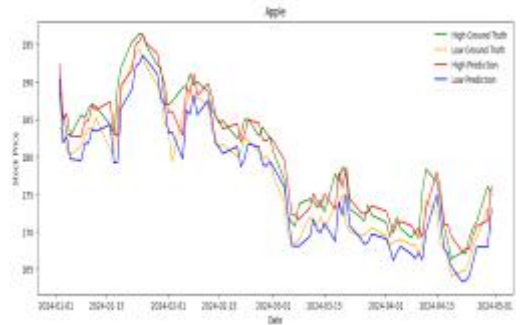


그림 4. MAMBA - 애플 주가 예측

## 5. 결론 및 향후 연구

본 연구는 주가 예측을 위한 LSTM과 MAMBA 모델에 직접적인 영향을 주는 파라미터들을 조절하여 최적의 값을 구하고자 예비 연구를 진행하였다. LSTM 모델과 MAMBA 모델의 하이퍼파라미터와 기본 매개변수의 튜닝을 통해 분석한 결과, 하이퍼파라미터에 의해 모델이 영향을 받으며 최적의 성능을 도출하기 위하여 적당한 조절의 필요성을 알 수 있었다.

향후 연구에서는 최적의 값을 도출하기 위해 다양한 알고리즘을 모델에 적용하여 하이퍼파라미터의 값을 도출할 계획이다. 또한 본 연구에서는 과거의 최저가와 최고가의 주가 데이터만을 활용하여 미래 주가를 예측하였지만, 향후에는 주가에 영향을 미치는 요인들을 고려하여 주가 예측 모델을 향상시킬 계획이다.

## References

- [1] 김지나, 고상근, 이수안. (2023). 통계적, 머신러닝, 딥러닝 모델을 이용한 주가 시계열 예측 및 비교. 한국정보과학회 학술발표논문집, 1809-1811.
- [2] 이정렬, 구명완. (2023). 주가 예측을 위한 Deep Learning: LSTM 및 Transformer 모델. 한국정보과학회 학술발표논문집, 1830-1832.
- [3] Gu, A., & Dao, T. (2023). Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752.
- [4] Frey, S., van Hoom, A., Jung, R., Hasselbring, W., & Kiel, B. (2011). MAMBA: A measurement architecture for model-based analysis.
- [5] 정종진, 김지연. (2020). LSTM을 이용한 주가예측 모델의 학습방법에 따른 성능분석. 디지털융복합연구, 18(11), 259-266.
- [6] 원종현, 신종민, 김재호, 이장원. (2023). 기계학습의 하이퍼파라미터 최적화 연구 동향. 한국통신학회논문지, 48(6), 733-747.
- [7] 이승훈, 김용수. (2021). 다구찌 실험계획법을 활용한 LSTM의 하이퍼파라미터 최적화에 관한 연구. 신뢰성 응용연구, 21(3), 238-245.
- [8] Yu, C., Qi, X., Ma, H., He, X., Wang, C., & Zhao, Y. (2020). LLR: Learning learning rates by LSTM for training neural networks. Neurocomputing, 394, 41-50.

## ARIMAX를 이용한 COVID-19 시기의 파생 변수와 국내 섹터 ETF의 가격 변동성 예측에 대한 연구

이태경<sup>1</sup>, 박성수<sup>1</sup>, 신경수<sup>1</sup>, 이수인<sup>1</sup>, 최인수<sup>2</sup>, 김장호<sup>3</sup>, 김우창<sup>4</sup>

### 요약

본 연구는 ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variables) 모형을 이용하여 COVID-19 시기 동안의 COVID-19 관련 변수와 국내 섹터 ETF 가격 변동 간의 관계를 분석하였다. 연구 목적은 COVID-19 시기 동안 COVID-19 관련 외생변수를 활용할 경우 섹터 ETF의 수익률이 유의미하게 증가하는지 확인하고자 하였다. COVID-19 시기 사망자수, 확진자수를 외생 변수로 사용한 ARIMAX 모형으로 섹터 ETF의 주가 변동을 예측한 결과, 일부 섹터의 주가 변동에 COVID-19 변수가 유의미한 영향을 미친 것으로 나타났다. 본 연구 결과는 COVID-19와 같은 외부 충격이 주식시장에 미치는 영향을 이해하고, 향후 유사한 팬데믹 상황에서 금융 시장의 변동성을 예측하고 대응 전략을 수립할 때 활용할 수 있음을 실증적으로 확인하였다.

주요 용어: ARIMAX, COVID-19, 수익률 예측, 섹터 분석.

---

<sup>1</sup>(공동 제1저자) 경희대학교 산업경영공학과

<sup>2</sup>KAIST 산업및시스템공학과

<sup>3</sup>고려대학교 기술경영전문대학원

<sup>4</sup>(교신저자) KAIST 산업및시스템공학과

## 다층 퍼셉트론을 활용한 화학 독성 데이터 다중 레이블 분류\*

오승현<sup>1</sup>, 노영민<sup>2</sup>

### 요약

Tox21 데이터는 화합물의 분자량 등 801개의 특성 정보와 12 종류의 독성 시험 결과 변수로 이루어져 있다. 잘 알려진 화합물 정보를 사용하여 독성을 예측하는 모델을 개발하는 것은 실험자의 안전과 다양한 실험을 진행하기 위해 중요하다. 그러나 일반적으로 독성이 존재하는 화합물의 샘플이 상대적으로 부족하기 때문에 분류 예측 모델을 개발하는 것은 클래스 불균형 문제로 인해 어려운 문제가 된다. 본 논문에서는 수치형 자료에서 일반적으로 사용되는 XGBoost와 같은 계산복잡도가 높은 모델을 사용하지 않고 단순히 다층 퍼셉트론(Multilayer Perceptron)을 활용하여 클래스 불균형 문제를 해결하고자 한다. 우리는 학습 데이터를 분석하여 주요한 입력 변수를 골라내고, 오버샘플링(Oversampling) 기법과 변수 정규화 기법을 사용한다. 더 나아가, 하나의 모델이 여러 종류의 독성 시험법 결과를 동시에 예측하도록 학습한다. 그 결과, 우리가 제안한 모델은 XGBoost를 사용할 때보다 더 짧은 시간에 여러 독성 시험법에 대해 독성 유무를 예측할 수 있다.

주요 용어: 화학 데이터셋, 화학 독성 분류, 다층 퍼셉트론, 딥러닝, 다중 레이블 분류.

\*본 연구는 환경부 주관 2024년도 화학물질 안전관리 특성화대학원 사업을 통한 성과물임을 밝힙니다.  
<sup>1</sup>02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 인공지능학과학과 석사과정.

E-mail: osh1795@naver.com

<sup>2</sup>(교신저자) 02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 인공지능학과학과 교수.

E-mail: youngmin.ro@uos.ac.kr

## Association between Oxidative Balance Score and Health-related Quality of Life Instrument with 8 Items (HINT-8) in Adults Based on Data from the 2021 Korea National Health and Nutrition Examination Survey (KNHNES)\*

*Subin Seong<sup>1</sup>, Jinheum Kim<sup>2</sup>*

### Abstract

The EQ-5D-3L is the most widely used health-related quality of life (HRQoL) instrument. However, there are only a total of 243 health conditions that can be expressed with the EQ-5D-3L, and in general population studies, the proportion of people responding to all questionnaires with ‘no problems’ is very high. Accordingly, in this presentation, instead of the EQ-5D-3L as an HRQoL instrument, we intend to employ the HINT-8 developed by Cho et al. (2014). In comparison to the EQ-5D-3L, the HINT-8 is reported to have a lower so-called ceiling effect, as it can express a total of 65,536 health conditions. This presentation aims to introduce the logistic quantile regression model, which was presented by Bottai et al. (2010), in order to investigate the impact of the oxidative balance score (OBS) on the HINT-8. A metric that takes the HINT-8 quality weight into account is the HINT-8 index, which was developed by Cho et al. (2017). As a representative value, the median is better than the mean because the index’s distribution skews to the left. A logistic transformation model was taken into consideration because it has a value in the interval [0.132, 1]. Among the 7,090 subjects collected in the 2021 KNHNES, those with missing analysis variables were excluded, the remaining 4,382 were defined as research subjects. OBS was defined using dietary and lifestyle variables; the dietary variables included 9 antioxidant and 3 prooxidant variables, and the lifestyle variables included 1 antioxidant and 2 prooxidant variables. A significant effect of OBS on HRQoL was found when OBS alone was taken into account (p-value = 0.0121). Furthermore, a statistically significant effect of OBS on HRQoL was observed even after controlling for socio-demographic factors (p-value = 0.0057), and this effect remained even after controlling for sleep and factors associated to disease. (p-value = 0.0063). It turned out that the p-value decreased with a decreasing quantile value. Finally, regardless of the three models, the odds for the median of the HINT-8 index increased by nearly 1.02 times when OBS increased by 1 point.

*Keywords* : Bounded outcome, HINT-8 index, Logistic transformation, Quantile.

---

\*This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00239958).

<sup>1</sup>Undergraduate student, Department of Data Science, College of Intelligent Software Convergence, 18323, Korea. E-mail 8800subin@naver.com

<sup>2</sup>Professor, Department of Data Science, College of Intelligent Software Convergence, 18323, Korea. E-mail jkimdt65@gmail.com

## Analysis of the Risk Factors towards Metabolic Syndrome for Korean Adults over 40 Years Old using Long-term Follow-up Cohort Data<sup>\*</sup>

*Myung Jun Kim<sup>1</sup>, Jinheum Kim<sup>2</sup>*

### Abstract

In this presentation, we aim to investigate factors affecting the prevalence of metabolic syndrome (MetS) in middle-aged and older adults using a community(Anseong and Ansan)-based cohort collected from the Korean Genome Epidemiology Survey (KoGES). Data were collected through follow-up surveys every two years from 2001 to 2018 and this study uses only the data up to the eighth follow-up survey. Kim & Yang (2023) proposed a Cox proportional hazards model with the time until MetS first occurrence as the outcome variable. However, because the presence or absence of MetS may change at each follow-up, this presentation conducted an analysis considering the entire follow-up period. One of them is a logistic regression model with the occurrence of MetS as the outcome variable, and the other is a Poisson regression model with the number of occurrences as the outcome variable. In addition, since some risk factors change over time, a logistic regression model was applied to analyze the prevalence of MetS from the baseline survey to the 8th follow-up survey. The odds of developing MetS were found to be higher in the Anseong cohort than in the Ansan cohort ( $e^{0.62}=1.86$ ), in women than in men ( $e^{0.43}=1.54$ ), in those without a spouse ( $e^{0.19}=1.21$ ), and in those without a family history of hypertension or diabetes ( $e^{0.48}=1.61$  or  $e^{0.31}=1.36$ ), respectively. The results regarding the rate of MetS occurrences showed similar characteristics to those on the prevalence of MetS occurrences, and the higher the education level ( $p$ -value $<0.001$ ), the longer the sleep time ( $p$ -value $<0.001$ ), and the lower the subjective health status ( $e^{0.09}=1.10$ ), the higher the rate of MetS occurrences. The analysis results for each follow-up survey and the integrated analysis results for the entire follow-up survey were analogous. However, the odds of developing MetS were higher in people with poor subjective health than in people with good subjective health, but the former was statistically significant ( $p$ -value $<0.001$ ) and the latter was not significant ( $p$ -value=0.202).

*Keywords* : Longitudinal outcome, Metabolic syndrome, Prevalence, Zero-inflated event.

---

<sup>\*</sup>This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2023-00239958).

<sup>1</sup>Undergraduate student, Department of Data Science, College of Intelligent Software Convergence, 18323, Korea. E-mail jinko70@naver.com.

<sup>2</sup>Professor, Department of Data Science, College of Intelligent Software Convergence, 18323, Korea. E-mail jkimdt65@gmail.com

## 결합된 제2종 점진적 복합 중도절단 상황에서 지수분포에 대한 정확한 가능성도 추론

황영은<sup>1</sup>, 김지현<sup>2</sup>, 이경준<sup>3</sup>

### 요 약

생존분석 및 신뢰성 연구에서 시간과 비용을 고려하기 위해 점진적 복합 중도절단의 적절한 추론이 중요하다. 본 연구에서는 결합된 제2종 점진적 복합 중도절단(type II progressive hybrid censoring scheme; ComT2PHCS)에서 지수분포의 모수를 추정하고, 이를 통해 평균 수명 시간을 도출한다. ComT2PHCS는 중도절단 상황에서 고장횟수와 실험 시간의 변동을 고려하는 방법론으로 일반화된 제1종 및 제2종 점진적 복합 중도절단을 결합한 형태이다. ComT2PHCS에서 지수분포 모수는 MLE(maximum likelihood estimator)와 베이지안 추론으로 유도된다. 또한, MLE의 conditional MGF(moment-generating function)을 계산하여 정확한 모수 추론을 수행한다. 마지막으로 예시 자료를 통해 지수분포의 모수 추정법별 성능을 비교한 결과, 사전분포의 정보가 주어졌을 때 계산된 모수가 가장 우수한 성능을 보였다.

주요 용어 : 결합된 제2종 점진적 복합 중도절단, 베이지안 추론, 지수분포, 정확한 가능성도 추론.

---

<sup>1</sup>39177 경상북도 구미시 대학로 61, 국립금오공과대학교 일반대학원 수학과, 박사과정.

E-mail: hye3775@naver.com

<sup>2</sup>39177 경상북도 구미시 대학로 61, 국립금오공과대학교 일반대학원 수학과, 석사과정.

<sup>3</sup>(교신저자) 39177 경상북도 구미시 대학로 61, 국립금오공과대학교 수리빅데이터학과, 조교수.

## 신장 및 심혈관질환 예측 및 위험평가 모델 개발\*

정상현<sup>1</sup>

### 요약

당뇨병은 뇌, 심혈관, 신장 등 여러 가지 합병증을 유발시키고 이 합병증들은 당뇨병으로 인한 사망의 주 원인이 된다. 따라서 본 연구에서는 한국인 유전체역학조사사업(KoGES)과 지역사회(안성·안산) 기본 코호트의 10년동안(2년 간격 추적검사)검진 및 설문한 10,030명의 자료를 활용하였다. 예측모델 개발을 위해 XGBoost, LightGBM, GBM 모델을 이용한 앙상블 모델을 만들었다. 이후 심혈관 질환과 신장질환의 발병을 예측하고, 이를 바탕으로 맞춤형 관리법 제공을 통한 의료시스템의 발전을 기대한다

주요 용어 : 당뇨병, 합병증, 머신러닝, 질병예측.

---

<sup>1</sup>국립금오공과대학교 수학과

## 점진적 중도절단 상황에서 적합도 통계량 개발

조영찬<sup>1</sup>, 박성희<sup>2</sup>, 이경준<sup>3</sup>

### 요약

많은 데이터 분석 방법이 데이터의 특정 분포를 가정하기 때문에 데이터 분석에 앞서 이를 확인해야 한다. 왜냐하면 데이터가 가정된 분포에 얼마나 잘 맞는지를 검증하는 것은 매우 중요하다. 확인하는 방법으로는 히스토그램이나 Q-Q플롯을 사용하는 방법이 있고, 경험적 분포 함수와 이론적 누적 분포 함수간의 거리 측정을 활용하여 데이터 분포를 평가하는 Goodness of fit Tests 방법이 있다. 하지만 제 2종 점진적 중도절단자료(progressive type II censored data)에서는 Goodness of fit Tests 방법을 사용해 데이터의 분포 가정이 맞는지 확인할 수 없다. 따라서 본 연구는 제 2종 점진적 중도절단자료(progressive type II censored data)에서 대칭 및 비대칭 위치 척도 분포의 Goodness of fit Tests를 제안했다. 제안된 검증 방법의 검증력은 몬테카를로(Monte Carlo) 시뮬레이션을 통해 추정했으며, 실제 대칭 및 비대칭 데이터인 파단강도 데이터, 로그 변환된 절연유 데이터에 적용하여 그 결과를 비교했다.

주요 용어 : 검정도, 로렌츠 곡선, 위치-척도 분포, 적합도 검증, 제 2종 점진적 중도절단 자료.

---

<sup>1</sup>39177 경상북도 구미시 대학로 61, 국립금오공과대학교 일반대학원 수학과, 석사과정.

E-mail: dudcks6436@naver.com

<sup>2</sup>39177 경상북도 구미시 대학로 61, 국립금오공과대학교 일반대학원 수학과, 석사과정.

<sup>3</sup>(교신저자) 39177 경상북도 구미시 대학로 61, 국립금오공과대학교 수리빅데이터학과, 조교수.

## 리그 오브 레전드(League of Legends) 게임 유저 데이터를 이용한 승패 예측

조혜령<sup>1</sup>, 이도현<sup>2</sup>, 손재우<sup>3</sup>, 송국영<sup>4</sup>, 박준범<sup>5</sup>

### 요약

본 연구는 리그 오브 레전드 게임에서 Riot API를 활용하여 데이터를 수집하고, 이를 전처리한 후, 데이터 모델링과 머신러닝 기법을 적용하여 승률을 예측하는 모델을 개발하였다. 리그 오브 레전드는 세계적으로 인기를 끌고 있는 e-스포츠 게임 중 하나로, 본 연구는 게임 데이터의 수집과 처리, 그리고 머신러닝 모델을 이용한 승률 예측을 통해 게임의 승패를 예측하는 기술적인 접근 방법을 탐구하여 승패 여부와 게임 데이터를 시계열로 정리하였다. 또한 로지스틱 회귀를 사용하여 승률 예측에 영향을 미치는 특징에 대해 파악하였다. 본 연구에서의 분석한 알고리즘 및 지표들인 연구결과는 향후 게임 분석 및 전략 수립에 도움을 주는 중요한 인사이트가 될 것으로 기대된다.

주요 용어: Riot API, 승률 예측, 로지스틱 회귀.

### 1. 서론

리그 오브 레전드(League of Legends)는 5명의 챔피언으로 구성된 양팀이 겨루는 PVP(Player Vs Player)게임으로, 상대방의 본진을 장악하는 게임이다. 플레이어들은 각자 챔피언 (Champion)이라 불리우는 고유의 능력과 플레이 스타일을 지닌 캐릭터를 선택하여 플레이한다.

경기가 시작되면 챔피언들은 경험치와 골드를 획득, 상대팀 공략을 위한 아이템을 구매하며 능력치를 높게 된다. 리그 오브 레전드의 기본 모드인 '소환사의 협곡'에서는 한 팀이 상대팀의 본진에 있는 넥서스를 파괴하면 승리하는 구조를 가진다.

본 논문은 리그 오브 레전드 게임에서의 승률 예측에 대한 연구를 다루고 있다. 우리는 머신러닝 기법을 활용하여 플레이어들의 게임 데이터를 분석하고, 승부 결과를 예측하는 모델을 개발하고자 하여 이를 위해 플레이어들의 게임전적 기록, 캐릭터 선택, 아이템 등 다양한 데이터를 수집하

<sup>1</sup>31962 충청남도 서산시 해미면 한서1로 46, 한서대학교 항공컴퓨터학과 학부연구생.

E-mail: johr0302@naver.com

<sup>2</sup>31962 충청남도 서산시 해미면 한서1로 46, 한서대학교 항공컴퓨터학과 학부연구생.

E-mail: dlehrms0124@gmail.com

<sup>3</sup>31962 충청남도 서산시 해미면 한서1로 46, 한서대학교 항공컴퓨터학과 학부연구생.

E-mail: sonjw0116@gmail.com

<sup>4</sup>31962 충청남도 서산시 해미면 한서1로 46, 한서대학교 항공컴퓨터학과 학부연구생.

E-mail: kysong0511@naver.com

<sup>5</sup>(교신저자) 31962 충청남도 서산시 해미면 한서1로 46, 한서대학교 항공컴퓨터학과 조교수.

E-mail: jbpark@hanseo.ac.kr

고 분석하였다. 그 후, 이러한 데이터를 기반으로 머신러닝 알고리즘을 학습시켜 각 플레이어나 팀의 승리 확률을 예측할 수 있는 모델을 연구하였다.

## 2. 본론

### 가. 데이터 수집

본 연구에서는 Riot API를 활용하여 리그 오브 레전드 게임의 데이터를 수집하였다. Riot API는 ‘라이엇 게임즈’에서 제공하는 공식적인 인터페이스로, 게임 내 정보에 대한 접근을 제공하여 다양한 게임 데이터를 손쉽게 수집할 수 있는 도구이다. 매치 데이터를 중심으로 수집하여 승률에 영향을 미치는 상세한 정보를 수집할 수 있다.

Figure 1은 API를 사용하여 데이터를 수집한 과정이다. 리그 오브 레전드(League of Legends)의 랭크 게임을 기준으로 브론즈, 실버, 골드, 플래티넘, 다이아에 속한 플레이어들의 데이터를 수집하였으며, 게임 데이터를 얻기 위해 필요한 정보를 단계별로 호출하여 사용하는 방식으로 진행하였다. 4개의 API를 사용하여 최종적으로 JSON 형태의 다양한 매치 데이터를 얻을 수 있었고, Pandas(판다스) 라이브러리를 사용해 데이터를 CSV 파일 형태로 저장 및 관리하였다.

### 나. 데이터 전처리

CSV 파일을 Pandas로 불러와 데이터를 필터링하고, 승패에 따라 데이터를 분리하는 데이터 전처리 과정을 진행하였다. ‘win’을 포함하는 헤더와 ‘lose’를 포함하는 헤더를 필터링하여 데이터를 추출하였다. 필터링된 열을 새로운 DataFrame으로 생성 후, 각 열의 이름을 정리하고 재설정한다. 이후 경기 시간을 나타내는 ‘duration’ 열을 데이터에 추가하고, 이 데이터의 승패 여부를 나타내는 ‘win’ 열을 1로 ‘lose’ 열은 0으로 표현하였다.



Figure 1. Data Collection Process Using API

승/패	데이터 수	
0(Lose)	7459rows X 16columns	X
1(Win)	7459rows X 16columns	
Total	14918rows X 16columns	
		duration

Figure 2. Win/Lose Data Classification

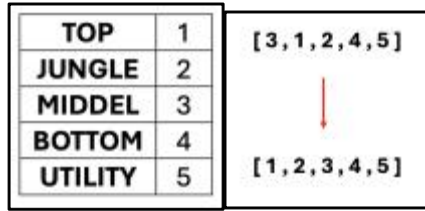


Figure 3. Position Mapping and Sorting

Figure 2 은 0(Lose)와 1(Win)의 데이터에서 머신러닝 모델에 입력하기 용이하게 만들기 위한 세 부적인 분류를 보여준다. Figure 3 와 같이 각 플레이어의 포지션을 리스트에 1부터 5까지의 숫자 로 매핑하는 과정을 진행하였고, 포지션 리스트를 기준으로 다른 데이터들의 정렬을 진행 하였다.

#### 다. 모델링 과정과 성능 결과

본 연구는 주요 목표인 게임의 승패(이진 결과)를 예측하는 것으로 이러한 이진 분류 문제에 특 화된 알고리즘인 로지스틱 회귀를 사용하였다. 로지스틱 회귀 모델을 학습하기 위해 ‘train\_test\_split’ 함수를 사용하여 데이터를 80%는 훈련 세트, 20%는 테스트 세트로 분할하였다. 훈련 세트를 사 용하여 모델을 학습시키고, 테스트 세트를 사용하여 모델의 성능 평가를 진행하였다.

로지스틱 회귀 모델의 계수는 각 특성 변수(‘X’, feature)가 타겟 변수(‘y’, target)에 미치는 영향을 나타낸다. Figure 5 는 계수를 시각화한 막대 그래프로 게임의 승패를 예측하는 데 중요한 특성들을 보여준다. 이 과정을 통해 어시스트(assists)와 킬(kills)의 양수 계수가 높으므로 이 특성들이 게임에 서 승리에 큰 영향을 미치고, 데스(deaths)의 음수 계수가 높으므로 데스가 많을수록 패배에 큰 영향 을 미친다는 확인할 수 있었다.

모델의 성능을 평가하기 위해, 학습된 모델을 사용하여 테스트 세트에 대한 예측을 수행하였다. 예측 확률을 계산하고, 다양한 성능 지표를 통해 모델의 성능을 평가하였다. Figure 6 은 ROC 곡선 을 시각화한 그래프이다. AUC가 1에 가까울수록 모델의 성능이 우수하고, 0.5에 가까울수록 무작위 로 예측하는 것과 비슷한 성능을 나타낸다. 결과적으로 AUC 값이 0.84 로 나왔고, 높은 예측 성능 을 보였다.

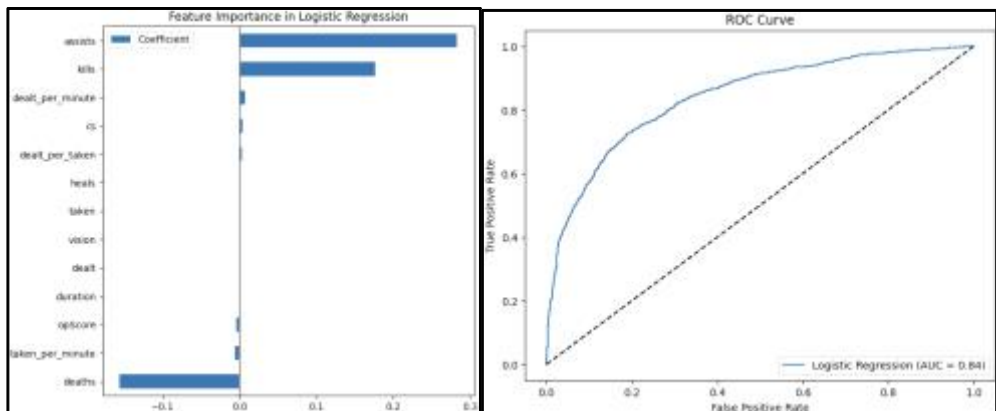


Figure 5. Feature Importance in Logistic Regression      Figure 6. ROC Curve Visualization

### 3. 결론

본 연구는 리그 오브 레전드 게임 데이터를 기반으로 승패 예측 모델을 구축하고, 로지스틱 회귀와 결정 트리 모델을 통해 성능을 평가하였다. Riot API를 활용하여 다양한 랭크의 데이터를 수집하고, 전치리를 통해 분석에 적합한 형태로 변환하여 데이터의 신뢰성과 일관성을 확보하였다. 승패 예측을 위해 로지스틱 회귀 모델을 사용하였으며, 모델의 성능을 평가한 결과 AUC 값이 0.84로 높은 예측 성능을 보였다. 본 연구를 이어서 향후 연구로 승패 요인에 간접적인 영향을 주는 다양한 요소에 대해 연구할 예정이다.

### References

- [1] Wikipedia. (n.d.). 리그 오브 레전드. In Wikipedia, The Free Encyclopedia. Retrieved May 23, 2024, from [https://ko.wikipedia.org/wiki/리그\\_오브\\_레전드](https://ko.wikipedia.org/wiki/리그_오브_레전드)
- [2] Ku, J.-M., & Yu, K. (2017). Design and application of a winning forecast model of the AOS genre game. *KIISE Transactions on Computing Practices*, 23(1), 37-44.
- [3] Heo, M.-G., & Park, C.-H. (2023). Production of skill classification AI model using process data and setting data in League of Legends. *Journal of Korea Game Society*, 23(5), 101-110. <https://doi.org/10.7583/JKGS.2023.23.5.101>
- [4] Kim, Y.-J., & Min, J.-H. (2023). Win-loss prediction using AOS game user data. *Journal of the Korea Society of Computer and Information*, 28(12), 23-32.
- [5] Kim, J. (2023). Study on predictive factors for League of Legends game churn using match record data. *Journal of Korea Game Society*, 23(4), 67-79. <https://doi.org/10.7583/JKGS.2023.23.4.67>
- [6] Oh, S. S., & Kim, D. H. (2012). Analysis of the academic research trend of e-sports. *Journal of Korean Society for Wellness*, 7(2), 113-121.
- [7] Riot Games. (n.d.). League of Legends official web site. Retrieved May 23, 2024, from <http://www.leagueoflegends.co.kr/>
- [8] Riot Games. (n.d.). League of Legends developers web site. Retrieved May 23, 2024, from <https://developer.riotgames.com/>
- [9] Kim, J. H. (2015). *R multivariate statistical analysis*. Seoul: KyoWooSa.
- [10] Lee, J.-W., Lee, J.-W., & Kim, H.-N. (2020, January 21). A study on the revitalization of e-sports by analyzing e-sports events. In *Proceedings of the Korean Society of Computer Information Conference* (pp. 1-8). Busan.

## 음원데이터를 활용한 주간 차트 특징 분석 및 예측\*

이민주<sup>1</sup>, 이하은<sup>2</sup>, 박준범<sup>3</sup>

### 요약

음원 콘텐츠는 현대 음악 시장에서 매우 중요한 역할을 하며, 다양한 부가가치를 창출한다. 특히, 음원의 순위는 그 흥행 여부를 판단하는 주요 지표로 작용한다. 본 연구에서는 음원의 흥행 여부를 결정짓는 여러 요소를 체계적으로 분석하였다. 연구 방법으로는 온라인 음원 플랫폼에서 데이터를 추출한 후, 이러한 데이터가 주간 차트 순위에 미치는 영향을 분석하였다. 이를 위해 데이터는 시계열로 정리되었으며, 주간 차트 100위 이내에 진입한 곡들의 특징을 종합적으로 분석하였다. 분석 결과 아티스트의 인지도, 발매시점, 장르 등이 음원의 흥행에 결정적인 영향을 미치는 것으로 나타났다. 특히, 이 전주 순위가 주간 차트 순위에 가장 큰 영향을 미치는 요인으로 도출되었다. 이러한 결과는 음원 마케팅 및 프로모션 전략 수립에 있어 중요한 시사점을 제공하며, 향후 음원 산업의 발전 방향에 있어서도 유의미한 통찰을 제시한다.

### 1. 서론

한국콘텐츠진흥원의 ‘2023 음악 이용자 실태조사’에 따르면 전국 만 10~69세 음악 콘텐츠 이용자들의 음악 이용 수단 또는 서비스에서 음원 스트리밍이 69.0%로 2위를 차지했으며 특히, 음원 스트리밍 및 다운로드 주 이용서비스에서 멜론 차트가 32.8%로 1위를 차지했다. 이러한 이유로 음악 기획사나 팬덤 등에서도 음원의 흥행 여부를 판단할 때 그 기준을 멜론의 음원차트 결과로 삼는 경우가 많기 때문에 데이터를 사용화하기 좋은 사이트로 선정했다.(Ga Yeon Kim·Myoung Jun Kim, 2018) 따라서 본 논문에서는 멜론 차트 데이터를 사용해 모의실험 설계에서 데이터 수집과 전처리, 모델링 과정을 거친 후 필요 데이터들을 모아 분석해 시각화하였다. 이후 실험결과에서 주간차트 순위에 미치는 영향과 순위 예측을 진행하였다.

### 2. 모의실험 설계

#### A. 데이터 수집 및 전처리

데이터 수집은 ‘멜론 주간 차트’ 1위 ~ 100위를 기준으로 하며, 수집 기간은 2023.01.01 ~ 2024.04.30로 한다. rank, song, artist, like, genre, release\_date, week 정보를 수집한다. 해당 정보를 Table 1에 넣은 후 ‘week’를 기준으로 지난 주의 순위를 ‘previous\_rank’ 열에 저장한다. 하지만 순

<sup>1</sup>31962 충청남도 서산시 해미면 한서1로 46, 한서대학교 공학행정학과 학부연구생. E-mail: lmj010110@gmail.com

<sup>2</sup>31962 충청남도 서산시 해미면 한서1로 46, 한서대학교 항공컴퓨터학과 학부연구생. E-mail: lhe93ee@gmail.com

<sup>3</sup>(교신저자) 31962 충청남도 서산시 해미면 한서1로 46, 한서대학교 항공컴퓨터학과 조교수. E-mail: jbpark@hanseo.ac.kr

위가 100위 밖이었던 경우 ‘previous\_rank’에 결측치가 생기게 된다. 이는 음원이 신곡일 경우에도 마찬가지이다. 고로 다음과 같은 기준으로 결측치를 대체한다.

(1) 2023. 01. 01 이후 발매곡은 신곡으로 분류해 해당 주차와 같은 순위의 값으로,

(2) 2023. 01. 01 이전 발매곡은 순위 하락으로 간주해 101 ~ 150 중 임의의 값으로 대체한다.

‘week’를 기준으로 (1) 3, 4, 5월일 경우 ‘봄’ (2) 6, 7, 8월일 경우 ‘여름’ (3) 9, 10, 11월일 경우 ‘가을’ (4) 12, 1, 2월일 경우 ‘겨울’로 분류하여 표 1의 ‘season’ 열로 저장한다.

Table 1. DataFrame shape

rank	song	artist	like	release_date	genre	week	previous_rank	season
1	Ditto	NewJeans	284484	2022-12-19	Dance	2023년1월1주차	1	Winter

## B. 데이터 모델링

Table 2와 같이 원본 데이터 프레임에서 필요한 부분만 추출하기 위해, 곡명을 파라미터로 받는 함수로 노래에 대한 정보만 반환 받아 인덱스를 키로 사용하는 딕셔너리로 저장한다. 그리고 순위 변동의 기울기를 구하기 위해 ‘week’ 열에 포함된 문자를 삭제하고 int 타입으로 변환해 0부터 시작하는 연속된 숫자로 재설정한다. 다음으로 ‘week’와 ‘rank’열에 diff() 메소드로 각각의 차이를 계산하고, 이를 ‘week\_diff’와 ‘rank\_diff’ 열로 저장한다. 이후 ‘rank\_diff’를 ‘week\_diff’로 나누어 기울기를 구해 ‘slope’열로 저장한다.

‘week’ 열을 여러 구간으로 나누어 각 구간에 대해 ‘slope’의 평균을 계산해 ‘week\_bin’ 열로 저장한다. 이를 기준으로 그룹화하고, 각 그룹의 ‘slope’ 평균을 계산한다.

‘slope’ 열에 -1을 곱하여 음수로 변환한다. 이는 순위가 높아질수록 값이 작아지는 특성을 반영함과 동시에 ‘Graph 0’과 같이 순위가 높아지는 경우 기울기 또한 양의 방향으로 올라가는 것을 보여주기 위한 것이다. 이 데이터를 토대로 ‘slope’ 값에 따른 ‘target’ 변수를 예측하도록 모델 학습을 시킨다.

Table 2. DataFrame of the song ‘Attention’

	rank	song	artist	week	week_diff	rank_diff	slope	week_bin
0	7	Attention	NewJeans	1	1.0	-2.0	-2.0	(0.939, 1.984]
:	:	:	:	:	:	:	:	:
61	92	Attention	NewJeans	62	1.0	-7.0	-7.0	(61.016, 62.0]

Table 3. df\_slope\_avg shape

	week_bin	slope
0		
1	0	2.0
:	:	:
61	61	7.0

## C. 데이터 시각화

### (1) 계절별 top10에 들어간 장르의 빈도

주간 차트 순위에 들어간 곡들의 특징을 분석하기 위해 먼저 모델링한 데이터를 이용해 ‘genre\_by\_season’이라는 새로운 데이터프레임을 생성한다. 이후 Top10에 많이 들어간 장르를 groupby하여 Figure 2.로 시각화한다.

Figure 2의 결과를 보면 사계절 모두 공통적으로 Dance와 Ballad의 빈도가 높으며 각 계절별로 가을에는 Folk/Blues와 Rap/Hiphop, 봄에는 Pop과 R&B/Soul, 여름에는 Pop/Ani/Webtoon/Kids/Comic, 겨울에는 Ballad/Indi와 R&B/Soul이라는 각기 다른 장르가 많은 비율을 차지하는 것을 볼 수 있다. 반면 계절별 Top50에 들어간 아티스트 10팀의 비율을 계산한 결과 장르의 분포와는 달리 차이가 거의 발생하지 않았다. 즉, 차트 순위에 들어가는 아티스트들은 계절의 영향을 많이 받지 않으며 아티스트의 인지도나 소속사의 크기 등 해당 음원 데이터에서는 분석할 수 없는 외부적인 영향을 받는 것으로 추측이 가능하다.

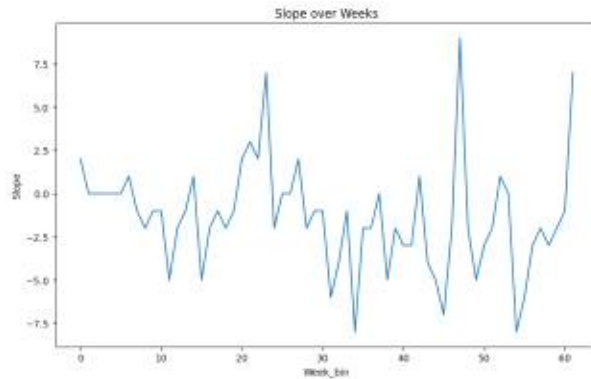


Figure 1. ‘Attention’ slope over Weeks

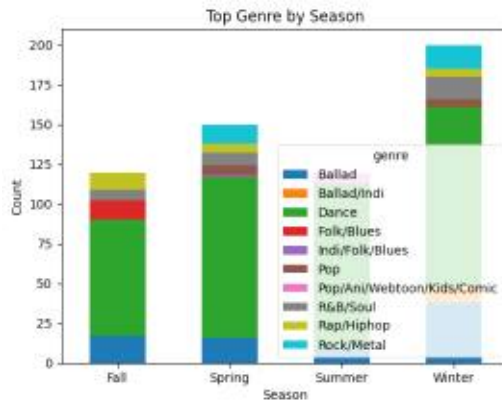


Figure 2. Top Genre By Season

rankUp_2					
song	artist	like	release_date	genre	count
All I Want for Christmas Is You	Mariah Carey	236402	2019-11-01	R&B/Soul	2
Candy	NCT DREAM	178150	2022-12-16	Dance	2
Dangerously	Charlie Puth	211068	2016-11-11	Pop	2
Get A Gutter	RIZE	93409	2023-09-04	Dance	2
Last Christmas	Ariana Grande	70394	2013-01-01	Pop	2
NIGHT DANCER	imase	124866	2022-08-19	J-POP	2
Snowman	Sia	265457	2022-12-30	Pop	2
Still With You	정국	126588	2020-06-05	R&B/Soul	2
Summer (Feat. BE'O (비오))	Paul Blanco	135431	2022-06-15	Rap/Hiphop	2
간절 사이에 피어난 장미 (Rose Blossom)	H1-KEY (하이키)	111185	2023-01-05	Dance	2
미리 미리 크리스마스 (Feat. 천도 OF MBLAQ)	아이유	180138	2010-12-09	Dance	2
별빛을	DK(다블배)	18794	2023-03-30	Ballad	2
알 자레자, 우리 (여름날 우리 X 로이킴)	로이킴	81654	2023-07-06	Ballad	2
잠만 시간 빌까	이무진	70261	2023-05-09	Rock/Metal	2
빛 눈	EXO	287051	2013-12-09	Ballad	2
크리스마스나타	성시경	171242	2012-12-06	Ballad	2

Figure 3. List of Selected Songs

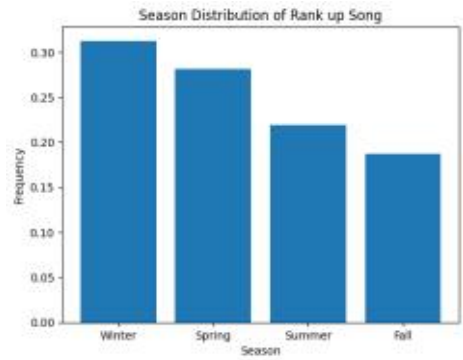


Figure 4. Season Distribution of Rank up Song

### (2) 역주행 곡의 계절 분포

역주행 곡을 추출하는 ‘rankUp\_2’ 데이터프레임을 만든다. 이때 ‘역주행 곡’이란 ‘rank’와 그 전주 차 ‘rank’의 차이가 20위 이상인 곡들을 선정해 ‘rankUp\_20’ 데이터프레임에 담고, ‘rankUp\_20’에 2번 이상 나온 곡들로 한번 더 작업해 나온 결과들을 말한다.(Figure 3.) ‘count’열은 ‘rankUp\_20’에 올라간 횟수를 뜻한다.

이렇게 ‘역주행 곡’이라고 판단된 곡들의 계절 정보를 추출해 역주행 했을 때의 계절을 비율로 나타낸 Figure 4를 생성한다.

### (3) 역주행곡 텍스트마이닝

Figure 3을 통해 역주행 곡의 계절 분포를 살펴보면 ‘겨울’에 역주행 곡이 가장 많이 발생한 것을 알 수 있다. 이를 토대로 위 곡들의 특징을 분석하기 위해 Table 4에서 season이 겨울인 노래 3곡을 뽑아 텍스트 마이닝을 진행한다. 이후 단어의 빈도를 쉽게 식별할 수 있도록 워드클라우드를 분석한 결과 노래 제목과 가사에 대표적으로 ‘겨울’을 나타내는 ‘눈’이나 ‘크리스마스’, ‘하얀’ 등의 단어가 사용된 것을 알 수 있었다.

## 3. 실험 결과

주간 차트 순위에 미치는 영향이 무엇인지 분석하기 위해 차트 순위 ‘rank’와 그 전주 차트 순위인 ‘previous\_rank’의 상관계수와 p-값을 계산한다. 상관계수는 -1에서 1까지의 값을 가지며, 1과 -1에 가까울수록 두 변수 사이의 상관관계가 강함을 나타낸다. Table 5.와 같이 ‘rank’와 ‘previous\_rank’의 상관계수는 0.95로 매우 강한 양의 상관관계를 뜻하며 ‘previous\_rank’가 ‘rank’에 미치는 영향이 크다고 볼 수 있다.

다음으로 p-값은 두 변수의 관계가 우연에 의한 것인지 아닌지를 나타낸다. p-값이 0.05보다 작으면 두 변수 사이의 관계가 통계적으로 유의하다고 볼 수 있다. ‘rank’와 ‘previous\_rank’의 p-값을 계산한 결과 소수점 6자리까지 표현했을 때 0이라는 결과를 도출한다.(Table 4) p-값이 0이라는 것은 귀무가설(두 변수 사이의 상관관계가 없다는 가설)이 기각되어 두 변수 사이가 통계적으로 유의미한 상관관계를 가짐을 뜻한다.

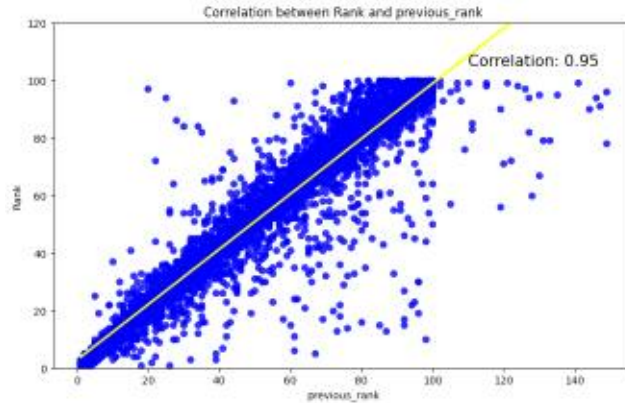


Figure 5. Correlation Between Rank and Previous\_rank

Table 4. Correlation and p-value

Correlation	p-value
0.95	0.000000

#### 4. 결론

이와 같이 분석한 결과 주간 차트 순위에는 발매시점, 이전 차트 순위와 계절에 따른 장르, 특정 계절에서의 곡 제목에 들어가는 단어 등이 영향을 끼치는 것을 알 수 있었다. 특히, 전주 순위가 주간 차트 순위에 가장 큰 영향을 미치는 요인으로 분석되었다. 이외에도 아티스트의 인지도나 소속사의 규모 등 본 논문에서 수집하지 못한 정보들도 영향을 끼칠 수 있다는 결론이 도출되었다. 본 연구에서는 음원차트 순위에 미치는 사회적인 요인까지는 분석하지 못했지만 음원 마케팅 및 프로모션 전략 수립에 있어 중요한 시사점을 제공하며, 향후 음원 산업의 발전 방향에 있어서 유의미한 통찰을 제시하였다. 본 연구를 이어서 향후 연구로 아티스트의 팬덤이나 사회적 요인들의 정보를 추가해 예측 정확도를 높이는 방향으로 연구할 예정이다.

#### References

- Chang Bin Lim. (2017). A Study on Musical Characteristics of Popular Music by Season : Focused on the lists on the Gaon Digital Chart after 2012
- Na Hee Kim. (2017). A study on the characteristics intentional chart reversal (Focused on the characteristics intentional music chart reversal)
- Min Young Lee. (2023). Diversity and performance in the K-Pop industry : policy implications from a decade of music chart analysis
- Seung Youp Cho. (2016). Dynamics of Chart Success in Korean Digital Music Industry : Is It Different from Traditional Physical Music Records Industry?. Korea Business Review
- A Yeong Gwak. (2017). An empirical study on the effecting factors of online music chart abusing
- Ga Yeon Kim·Myoung Jun Kim. (2018). A Study on the Prediction Index for Chart Success of Digital Music Contents based on Analysis of Social Data
- Korea Creative Content Agency. (2023). 2023 Survey of Music Users. (KOCCA 23-08). URL

## 라만 분광법을 이용한 10 $\mu$ m 이하 미세플라스틱의 신속한 검출 및 분류를 위한 단일 선형 네트워크 분석 및 제안\*

이동현<sup>1</sup>, 노영민<sup>2</sup>

### 요 약

미세 플라스틱의 무분별한 확산은 심각한 생태적 영향을 초래할 수 있는 중요한 환경 문제로 대두되고 있다. 이에 따라 미세 합성 폴리머 입자를 분석하여 환경적 영향을 평가하는 방법론 개발이 필수적이다. 라만 분광법은 미세 플라스틱 분석에 효과적이지만 높은 레이저 출력이나 긴 분석 시간이 요구되고 적절한 신호 대 잡음 비율을 얻는 데 어려움이 있으며, 이는 전체 분석 시간을 지연시키는 요인이 된다. 이 문제를 해결하기 위해, 이전 연구에서는 잡음을 추가한 분석 결과 및 미분된 입력을 활용하여 CNN 기반의 미세 플라스틱 탐지 및 분류 프레임워크를 제안하였다. 그러나 CNN은 특정 스펙트럼에서 발생하는 값을 인지하기 어려우며, 이를 인지하기 위해 많은 계산 자원이 소요되는 단점이 있다. 본 연구에서는, 단순한 Linear layer를 사용해 시계열 데이터를 처리하는 이전 연구에 영감을 받아, 단일 Linear layer만으로도 기존 CNN 기반 연구보다 약 1% 이상 정확도가 향상되는 것을 확인한다. 더 나아가 미분된 입력을 활용하는 것 대신 학습 가능한 Convolution 필터를 거친 입력을 활용하여 추가적인 성능 향상을 얻는다. 이러한 접근 방식은 미세 플라스틱 분석의 효율성을 크게 향상시킬 수 있으며, 신속하고 정확한 환경적 영향 평가를 가능하게 할 것으로 기대된다.

주요 용어: 미세 플라스틱 검출, 딥러닝.

\*본 연구는 환경부 주관 2024년도 화학물질 안전관리 특성화대학원 사업을 통한 성과물임을 밝힙니다.

<sup>1</sup>02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 인공지능학과 석사과정.

E-mail: dslisleadh@gmail.com

<sup>2</sup>(교신저자) 02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 인공지능학과 교수.

E-mail: yongmin.ro@uos.ac.kr

## A New Large-Scale Knowledge Graph for Link Prediction in Chemicals inducing Diseases\*

*Gyeongdong Woo<sup>1</sup>, Soyoung Cho<sup>1</sup>, Wonhyung Shiri<sup>2</sup>, Kione Kim<sup>2</sup>,  
Donghyeon Kim<sup>3</sup>, Jinhee Choi<sup>4</sup>, Jong-June Jeon<sup>5</sup>*

### Abstract

As societal concerns about environmental and health issues increase, studying and understanding the impact of chemicals on the human body is becoming increasingly important. With continuous research on the interactions between chemical compounds, genes, and diseases and the accumulation of these research results in publicly accessible databases, we are presented with new opportunities to understand and utilize these findings integratively. A representative success case is the exploration of drug development and therapeutic strategies using biological knowledge graphs and artificial intelligence. To extend this to the field of chemical toxicity prediction, we aim to construct a large-scale knowledge graph processed from the existing Comparative Toxicogenomics Database (CTD), termed the Comparative Toxicogenomics Database Knowledge Graph (CTDKG). Using this graph, we present baseline experimental results for link prediction tasks and provide negative samples for validation/test data to ensure reproducibility of experiments. Through this data, we anticipate promoting the development of artificial intelligence that can more accurately infer the potential hazards of chemicals according to the Adverse Outcome Pathway.

*keywords* : Knowledge graph, Link prediction, Comparative Toxicogenomics Database.

---

\*This work is supported by a grant from the Korean Ministry of Environment through ‘Environmental Health R&D Program’ (2021003310005) and ‘the National Research Foundation of Korea grant’ (2022R1A4A3033874, 2022R1F1A1074758).

<sup>1</sup>(Equal contribution)Department of Statistical Data Science, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504, Korea.

<sup>2</sup>Department of Statistical Data Science, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504, Korea.

<sup>3</sup>School of Environmental Engineering, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504, Korea.

<sup>4</sup>(Co-Corresponding Author)School of Environmental Engineering, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504, Korea.

<sup>5</sup>(Co-Corresponding Author)Department of Statistics, University of Seoul, 163 Seoulsiripdae-ro, Dongdaemun-gu, Seoul 02504, Korea.

## Direct Optimization Approach for Query Performance Prediction

*Jae-Hwan Jung*<sup>1</sup>, *Jong-June Jeon*<sup>2</sup>

### Abstract

Query Performance Prediction (QPP) predicts the quality of queries in Information Retrieval (IR) systems without relying on human-judged relevance. The advent of large language models (LLMs) like BERT has significantly impacted QPP, resulting in the development of supervised QPP models that have achieved high performance. QPP models are typically evaluated using rank correlation measures, such as Kendall's tau or Spearman's rho, to compare actual and predicted query performance. These rank correlation measures are non-differentiable, posing challenges for direct optimization. Consequently, existing supervised QPP approaches often resort to using mean square error or cross-entropy as objective functions. In this study, we propose a novel direct optimization method that approximates the rank correlation evaluation measure with a differentiable function, allowing it to be utilized as an objective function. This approach aims to enhance the alignment between the QPP model's training objective and its evaluation measure, thereby improving the prediction accuracy of queries in the context of IR systems.

*Keywords* : Query Performance Prediction, Direct Optimization, Rank Correlation.

---

<sup>1</sup>02504 Graduate Student, Department of Statistics, University of Seoul, Seoulsiripdaero 163, Korea.  
E-mail: y18630@uos.ac.kr

<sup>2</sup>(Corresponding Author) 02504 Professor, Department of Statistics, University of Seoul, Seoulsiripdaero 163, Korea. E-mail: jj.jeon@uos.ac.kr

## Clustering sleep patterns and exploring their correlation with hs-CRP Levels\*

*Hyeonjeong Lim<sup>1</sup>, Heejin Kim<sup>2</sup>, Eunjee Lee<sup>3</sup>, Boram Yang<sup>4</sup>*

### Abstract

As maintaining a healthy life necessitates sufficient sleep, as numerous studies have illuminated the connection between sleep patterns and inflammation levels. High-sensitivity C-reactive Protein (hs-CRP) serves as a prominent inflammatory marker. Individuals with higher hs-CRP levels are likely to belong to a high-risk group for cardiovascular. In this study, we investigate the association between sleep patterns and high risk group by using the Korea National Health and Nutrition Examination Survey data. We utilize latent class analysis (LCA) to detect sleep patterns, offering clinicians easily interpretable and applicable clustering results. Three sleep patterns were identified : Balanced Sleeper(little difference in sleep time and bedtime between weekends and weekdays while keeping the recommended sleep time), Poor Sleeper(Sleep less than the recommended and big difference between weekend and weekday sleep time and bedtime), Moderate Sleeper(sleep as much as the recommended sleep time and the difference between bedtime on weekends and weekdays is relatively large). We conducted logistic regression analysis to explored the effects of the three sleep patterns on being in a high-risk group for cardiovascular. Results indicate that the increase in sleep time and bedtime of weekday compared to weekend is significantly related to higher odds.

*Keywords:* Latent Class Analysis, logistic regression.

---

\*This material was based on work partially supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant (No. 2020-0-01441, Artificial Intelligence Convergence Research Center(Chungnam National University)). This work was also partially supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2022M3J6A1084843, No. NRF-2021R1C1C1013936, No. 2022R1A5A7085156).

<sup>1</sup>Daehak-ro, Yuseong-gu, Daejeon (34134), Department of statistics and data science, Chungnam National University, dlaguswjd123@o.cnu.ac.kr

<sup>2</sup>Daehak-ro, Yuseong-gu, Daejeon (34134), Department of statistics and data science, Chungnam National University, black7cat@naver.com

<sup>3</sup>Daehak-ro, Yuseong-gu, Daejeon (34134), Department of statistics and data science, Chungnam National University, eunjee.cnu@gmail.com

<sup>4</sup>Daehak-ro, Yuseong-gu, Daejeon (34134), Department of Pharmacy, Chungnam National University, br.yang@cnu.ac.kr

## 전기 노인과 후기 노인의 신체활동 제한 사유와 악력 정도의 비교

노효련<sup>1</sup>

### 요약

본 연구에서는 전기 노인과 후기 노인의 일상생활과 사회활동을 제한하는 건강문제를 알아보고 악력 정도 비교하였다. 본 연구는 한국질병관리청에서 실시하는 제 7기 국민 건강영양조사(2016~2018)의 원시자료를 사용하였다. 한국질병관리본부에서 제시한 절차대로 복합표본화하여 통계를 실시하였다. 본 연구 대상자는 65세 이상의 노인인구 중 활동제한을 가지고 있는 8,374명(남성: 3,620명, 여성 :4,754명). 활동 제한을 받고 있는 전기 노인은 15.4%, 후기 노인은 23.1%로 나타났다. 노인들의 활동제한 사유 중 가장 높은 빈도를 나타내는 3가지 주요 질환은 등과 목의 문제, 관절염 및 류마티즘, 무릎 및 다리통증 순이었다. 한국 노인들의 88% 이상에서 주 사용손으로 오른손을 사용하고 있었다. 일상생활 및 사회활동에 제한이 있는 경우 주로 사용하는 손의 악력을 기준으로 전기 노인의 활동제한군의 악력은 활동제한이 없는 집단의 -2.462배였으며 악력은 후기 노인의 근력은 활동 제한이 없는 그룹의 -1.562배였다. 후기 노인군에서는 관절염으로 인해 활동제한이 있었던 경우에 활동제한의 원인이 3배 더 높았다. 전기 노인과 후기 노인의 악력은 활동제한과 높은 상관관계를 보였지만, 개인의 활동제한 원인에 따라 악력과의 상관관계가 달랐다. 노인에서는 무릎 및 다리 통증으로 인해 활동 제한과 악력 간의 상관 관계가 발견되었습니다. 그러나 활동제한 이유에 따른 악력의 차이는 일관되지 않았다. 활동 제한이 모든 노년층에서 악력 감소로 이어진다고 하기는 어렵다.

주요 용어 : 건강문제, 악력, 전기노인, 후기노인, 활동제한.

## 우리나라 여성의 골다공증 영향요인-50세 이상을 대상으로\*

홍희정<sup>1</sup>

### 요 약

골다공증은 골의 강도가 약화되어 구조적 노화로 쉽게 골절이 일어나는 전신 질환으로 (Scholz-Ahrens, Schrezenmeir, 2004) 노년층과 폐경기 이후 여성에게 많이 발생하며, 특히 50세 이상의 골절과 밀접한 관련이 있는 것으로 나타났다(Abrahamsen et al, 2009). 또한 국민건강영양조사에서도 만 50세 이상 골다공증 유병률 결과 여자 34.9%, 남자 7.8%로 나타나 남성에 비해 여성의 골다공증 유병률이 더 높은 것으로 확인되었다(질병관리청, 2012). 따라서 본 연구의 목적은 질병관리본부의 국민건강영양조사 제8기(2019~2021년) 원시자료를 이용하여 50세 이상 한국 여성의 골다공증 영향 요인을 확인하고자 한다. 이를 위하여 국민건강영양조사 원시자료를 이용하여 제8기(2019~2021년)에 실시한 조사대상자 전체 22,559명 중 50세 이상 여성 대상자 1,553명의 자료를 분석하였다. 자료분석에서 연구대상자의 일반적 특성과 생활습관 특성은 복합표본분석을 시행하여 빈도, 가중된 백분율, 평균 및 표준편차로 검증하였다. 정상군, 골다공증군 간의 일반적 특성과 생활습관 특성 차이는  $\chi^2$  test, ANOVA 및 Scheffé test로 분석하였다. 골다공증에 영향을 미치는 영향 요인은 로지스틱 회귀분석으로 시행하였다. 그 결과, 골다공증 영향요인으로 연령, 가구소득, 교육수준, BMI가 나타났다. 골다공증에서는 연령이 70세 이상에 비해 50~59세 0.23배(95% CI: 0.16~0.34,  $p < .001$ ), 60~69세 0.65배(95% CI: 0.50~0.84,  $p = .001$ )로 나타났고, 가구소득(4분위)은 '상'에 비해 '하'는 1.60배 높게(95% CI: 1.05~2.42,  $p = .027$ ), '중'은 1.60배 높게(95% CI: 1.14~2.25,  $p = .007$ ) 보였다. 교육수준에서는 '대졸 이상'에 비해 '초졸 이하'가 1.75배 높게(95% CI: 1.14~2.70,  $p = .011$ ), '중졸'이 1.64배 높게(95% CI: 1.04~2.58,  $p = .034$ ), '고졸'이 0.90배(95% CI: 0.57~1.42,  $p = .644$ )이다. BMI가 한 단위 증가할수록 골다공증 위험이 0.94배였다(95% CI: 0.91~0.97,  $p = .001$ ). 따라서 본 연구는 골다공증 영향 요인을 파악하여 추후 골다공증 예방 및 중재 프로그램 개발에 도움이 될 것으로 사료된다.

주요 용어: 여성, 골다공증, 영향요인.

### 1. 연구의 필요성과 목적

골다공증은 골의 강도가 약화되어 구조적 노화로 쉽게 골절이 일어나는 전신 질환으로 (Scholz-Ahrens, Schrezenmeir, 2004) 노년층과 폐경기 이후 여성에게 많이 발생하며, 특히 50세 이상의 골절과 밀접한 관련이 있는 것으로 나타났다(Abrahamsen et al, 2009). 또한 국민건강영양조사에서도 만 50세 이상 골다공증 유병률 결과 여자 34.9%, 남자 7.8%로 나타나 남성에 비해 여성의 골다공증 유병률이 더 높은 것으로 확인되었다(질병관리청, 2012). 따라서 본 연구의 목적은 질병관리본부의 국민건강영양조사 제8기(2019~2021년) 원시자료를 이용하여 50세 이상 한국 여성의 골다공증 영향 요인을 확인하고자 한다.

<sup>1</sup>50578 경상남도 양산시 명곡로 321, 동원과학기술대학교 간호학과 겸임교수. E-mail: hjhong1117@hanmail.net

- 1) 정상군과 골다공증군 간 일반적 특성의 차이를 확인한다.
- 2) 정상군과 골다공증군 간 생활습관 특성의 차이를 확인한다.
- 3) 골다공증에 영향을 미치는 영향 요인을 확인한다.

## 2. 연구방법

### 2.1. 연구 설계

본 연구는 50세 이상 한국 여성의 골다공증 영향 요인을 확인하기 위한 서술적 조사 연구이다.

### 2.2. 자료수집 및 연구절차

국민건강영양조사는 매년 192개 표본조사구를 추출해 우리나라 만1세이상 국민 1만 명에 대한 건강수준과 건강행태, 식품 및 영양섭취 실태에 대해 국가 및 시도 단위로 통계 산출을 위한 조사를 실시한다. 표본추출방법으로 조사구와 가구를 1, 2차 추출단위로 하는 2단계 층화집락표본추출 방법을 사용하였고, 제8기(2019~2021년)의 경우에는 시도, 동·읍면, 주택유형(일반주택, 아파트)을 기준으로 하여 추출틀을 층화하고, 주거면적 비율과 가구주 나이와 1인 가구 비율 등을 내재적 층화 기준으로 사용하였다. 조사구는 연간 192개로 제8기(2019~2021년) 3년간 576개를 추출하였다(질병관리청, 2023).

본 연구에서 연구자료는 국민건강영양조사 홈페이지에서 소정의 절차를 거친 후 원시자료를 제공받아 분석하였다. 본 연구에서는 제8기(2019~2021년)에 실시한 조사대상자 전체 22,559명 중 50세 이상 여성 대상자 1,553명을 분석대상자로 하였다.

### 2.3. 윤리적 고려

본 연구는 연구 참여자의 윤리적 측면의 보호를 위하여 자료수집에 앞서 D대학교 생명윤리위원회(DIRB-202405-HR-W-21)의 승인을 받았으며, 질병관리본부의 국민건강영양조사 제8기(2019~2021년) 원시자료를 이용하여 50세 이상 한국 여성의 골다공증 영향 요인을 확인하고자 한다.

### 2.4. 자료분석방법

본 연구의 자료는 IBM SPSS WIN 26.0 program을 이용하여 분석하였다.

## 3. 연구결과

### 3.1. 대상자의 일반적 특성

본 연구에서 정상군, 골다공증 군 간 일반적 특성 차이의 분석 결과, 혼인 상태는 ‘유배우자\_동거’가 골다공증군이 56.7%로 가장 많았고, 정상군은 72.7%로 두 그룹 간 차이가 나타났다( $\chi^2=97.35$ ,  $p<.001$ ). 연령은 70세 이상이 골다공증군 50.7%로 가장 높게 보였고, 정상군 26.2%이며, 두 그룹 간 차이가 있다( $\chi^2=314.41$ ,  $p<.001$ ). 가구소득(4분위)에서는 ‘중’이 골다공증군 48.2%로 나타나

가장 많았고, 정상군 48.6%로 두 그룹간 차이를 나타냈다( $\chi^2=129.04$ ,  $p<.001$ ). 또한 교육수준은 ‘초졸 이하’가 골다공증군이 55.3%로 가장 높았고, 정상군 28.4%로 두 그룹간 차이를 보였다( $\chi^2=217.16$ ,  $p<.001$ )(Table 1).

Table 1. Comparison of general characteristics between groups with normal, osteoporosis

Variables	Categories	Normal	Osteoporosis	$\chi^2$	p
		n(%)	n(%)		
Marital status	With a spouse_ Cohabitation	2745(72.7)	511(56.7)	97.35	< .001
	With a spouse_ separation	52(1.1)	14(1.9)		
	bereavement	917(19.3)	348(34.4)		
	divorce	298(6.8)	62(7.0)		
Age(year)	50~59	1488(44.5)	96(13.2)	314.41	< .001
	60~69	1304(29.3)	327(36.0)		
	≥70	1286(26.2)	519(50.7)		
Household income (4th quartile)	Low	1169(24.0)	430(39.9)	129.04	< .001
	Medium	1922(48.6)	415(48.2)		
	Top	954(27.4)	95(11.9)		
Education level	Elementary graduates and younger	1257(28.4)	562(55.3)	271.16	< .001
	Middle school graduates	554(14.1)	157(17.9)		
	High School Graduates	1157(35.8)	146(17.8)		
	College graduate or above	687(21.6)	67(9.0)		

\*Rao-scott 수정된 카이제곱 통계량

### 3.2. 대상자의 생활습관 특성 차이

정상군, 골다공증 간 생활습관 특성 차이로 음주 여부( $\chi^2=15.25$ ,  $p<.001$ ), 유산소 신체활동 실천율( $\chi^2=6.51$ ,  $p=.011$ ), 폐경 여부( $\chi^2=44.46$ ,  $p<.001$ )가 유의미하게 나타났다.

### 3.3. 골다공증에 미치는 영향

골다공증 영향요인을 알아보기 위해 일반적 특성 및 생활습관 특성을 독립변수로, 골다공증 여부(정상군 및 골다공증군)를 종속변수로 로지스틱 회귀분석을 시행하였다. 그 결과, 골다공증 영향요인으로 연령, 가구소득, 교육수준, BMI가 나타났다. 골다공증에서는 연령이 70세 이상에 비해 50~59세 0.23배(95% CI: 0.16~0.34,  $p<.001$ ), 60~69세 0.65배(95% CI: 0.50~0.84,  $p=.001$ )로 나타났고, 가구소득(4분위)은 ‘상’에 비해 ‘하’는 1.60배 높게(95% CI: 1.05~2.42,  $p=.027$ ), ‘중’은 1.60배 높게(95% CI: 1.14~2.25,  $p=.007$ ) 보였다. 교육수준에서는 ‘대졸 이상’에 비해 ‘초졸 이하’가 1.75배 높게(95% CI: 1.14~2.70,  $p=.011$ ), ‘중졸’이 1.64배 높게(95% CI: 1.04~2.58,  $p=.034$ ), ‘고졸’이 0.90배(95% CI: 0.57~1.42,  $p=.644$ )이다. BMI가 한 단위 증가할수록 골다공증 위험이 0.94배였다(95% CI: 0.91~0.97,  $p=.001$ )(Table 2).

Table 2. Influencing factors osteoporosis in women in Korea

Variables	Categories	OR	95% CI	<i>p</i>
Alcohol consumption	Don't drink alcohol	1.11	0.88~1.41	0.379
	drinking alcohol	1		
Smoking	No smoking	1.43	0.90~2.30	0.134
	Smoking	1		
Aerobic physical activity practice rate	Don't do it	0.88	0.66~1.17	0.374
	Doing	1		
Menopause	Premenopausal	0.78	0.53~1.15	0.205
	Menopause	1		
Age(year)	50~59	.23	0.16~0.34	< .001
	60~69	.65	0.50~0.84	.001
	≥70	1		
Household income (4th quartile)	Low	1.60	1.05~2.42	.027
	Medium	1.60	1.14~2.25	.007
	Top	1		
Education level	Elementary graduates and younger	1.75	1.14~2.70	.011
	Middle school graduates	1.64	1.04~2.58	.034
	High School Graduates	0.90	0.57~1.42	.644
	College graduate or above	1		
BMI(Weight/Height)	-	.94	0.91~0.97	.001
Walking days	-	.96	0.92~1.01	.111

\*OR=odds ratio ; CI=confidence interval.

## References

질병관리청 (2012). 건강행태 및 만성질환 통계. 국민건강영양조사 제5기 3차년도.

질병관리청 (2023). 국민건강영양조사 원시자료 이용지침서 제8기(2019~2021). 3-5.

Abrahamsen B., van Staa T., Ariely R., Olson M., Cooper C. (2009). Excess mortality following up hip fracture: a systematic epidemiological review. *Osteoporos Int.* 20(10):1633-1650. Doi: 10.1007/s00198-009-0920-3

## 코로나19 예방접종 영향 요인 분석

김은경<sup>1</sup>

### 요 약

본 연구는 2021년 전국적으로 수행된 지역사회건강조사(Community Health Survey, CHS) 자료를 활용하여 코로나19 예방접종에 영향을 미치는 요인을 파악하기 위한 이차자료 분석연구이다. 19세 이상 231,784명을 대상으로 자료 분석은 IBM SPSS statistics SPSS 21.0 프로그램을 이용하여 분석하였다. 연구대상자의 코로나19 예방접종에 영향을 미치는 요인은 로지스틱 회귀분석을 이용하여 분석하였다. 코로나19 예방접종에 대한 영향요인은 64이하, 남성, 직업이 없는 경우, 배우자가 없는 경우, 미충족의료 경험, 국민기초생활보장 수급자, 주관적 건강이 낮은 경우, 우울증, 현재 흡연자, 아침식사를 하지 않는 경우 백신 미접종 위험이 높았고 고등학교 미만 학력, 코로나19 감염에 대한 심리적 우려, 만성질환자, 현재 음주자에서 코로나19 백신 미접종 위험이 낮았다. 코로나19 예방을 위한 가장 효과적인 방법은 백신접종이므로 본 연구결과를 바탕으로 취약계층의 코로나19 백신 접종을 촉진하기 위한 개입 전략과 실질적인 지원체계가 마련되어야 할 것이다.

주요 용어 : 코로나19, 예방접종, 지역사회건강조사.

## 발목 관절의 통증과 가동범위 제한에 대한 봉대형 파라핀 적용 단일사례연구

노효련<sup>1</sup>, 이경봉<sup>2</sup>, 정우경<sup>3</sup>, 정은지<sup>4</sup>, 최희원<sup>5</sup>, 김혜미<sup>6</sup>, 남채림<sup>7</sup>,  
이나린<sup>8</sup>, 이승현<sup>9</sup>, 이유진<sup>10</sup>, 길우진<sup>11</sup>, 이수빈<sup>12</sup>

### 요약

본 연구는 발목관절의 통증과 관절가동범위의 제한을 받고 있는 발목에 봉대형 파라핀을 적용하고 그 효과를 알아본 단일사례 연구이다. 요추 디스크 및 협착증 후 종아리 외측부와 발목에 통증이 있으며 관절가동범위의 제한이 있었다. 관절가동범위를 측정하는 앱을 이용하여 발목관절 가동범위를 측정하였고 시각 상사척도를 이용하여 통증 정도를 측정하였다. 파라핀이 침부된 신장력이 있는 봉대를 발목관절 중심으로 6겹 정도 감싼 후 15~20분 가량 유지 후 파라핀 봉대를 제거한 후 발목관절 운동을 실시하였다. 주 2회 총 7회기를 실시하였다. 발목관절의 수동적 관절가동범위는 35° 증가하였으며 통증은 4점 감소하였다. 봉대형 파라핀치료는 발목의 통증과 가동범위를 증가시킬 수 있을 것으로 보인다.

주요용어 : 관절가동범위, 발목관절, 봉대형 파라핀, 통증.

### 1. 서론

파라핀 치료는 물리치료 영역 중 열치료에 해당하는 것으로 골관절염(Osteo-arthritis), 류마티스관절염, 수근관 증후군(capal tunnel syndrem), 수족 냉증 등의 손의 질환, 전신 경화증과 같은 혈관 질환, 습진과 같은 피부 질환과 같은 신경 질환, 골절이나 뺨 후의 관절 경직, 급성기를 지난 윤활낭염, 섬유염, 힘줄염, 과도긴장 등으로 광범위하다(Lee et al, 2003). 치료를 원하는 특정 부위를 파라핀 왁스가 녹여진 통(bath)에 담그는 침습법과 통에 담글 수 없는 무릎 등의 신체 부위는 붓으로 칠을 하는 방식으로 적용한다(Lee et al, 2003). 그러나 의료 현장에서는 침습법으로 손과 손목 질환에만 주로 적용되고 있다(Kim et al, 2023). 이와 같이 파라핀 치료의 적용이 가능한 질환이 다양함

<sup>1</sup>25940 대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 교수. E-mail: withtry@kangwon.ac.kr

<sup>2</sup>25940(교신저자)대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 부교수. E-mail: kblee@kangwon.ac.kr

<sup>3</sup>25940 대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 학부재학생. E-mail: wookjz97@gmail.com

<sup>4</sup>25940 대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 학부재학생. E-mail: check1215@naver.com

<sup>5</sup>25940 대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 학부재학생. E-mail: gmlndnjs02140@naver.com

<sup>6</sup>25940 대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 학부재학생. E-mail: senbekim@naver.com

<sup>7</sup>25940 대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 학부재학생. E-mail: jjalomy104@naver.com

<sup>8</sup>25940 대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 학부재학생. E-mail: nmr0815@naver.com

<sup>9</sup>25940 대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 학부재학생. E-mail: 040909lsh@naver.com

<sup>10</sup>25940대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 학부재학생. E-mail: qazwsx467979@naver.com

<sup>11</sup>25940 대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 학부재학생. E-mail: jj2094@naver.com

<sup>12</sup>25940 대한민국 삼척시 도계읍 황조리 강원대학교 물리치료학과 학부재학생. E-mail: pchyp16@gmail.com

에도 불구하고 치료 기구와 적용 방법의 한계로 인하여 다양한 질환에 적용이 되지 않고 있다. 이에 본 연구는 허리 디스크 및 협착증 수술 후 발목 관절의 통증과 가동범위의 제한을 받고 있는 대상자에게 붕대형 파라핀을 적용을 한 후 그 효과를 알아보고자 하였다.

## 2. 연구 방법

### 2.1 연구 참여자 및 측정

64세 여성으로 2022년 요추 2~3번의 디스크로 수술을 한 후 같은 해 10월 척추 협착증으로 다시 수술하였다. 수술 후 걸을 수가 없어서 2023년 10월 3번째 허리수술 후 다시 걸을 수 있게 되었고 현재는 무릎 아래로 종아리 외측부와 발목에 통증이 있으며 발목의 등쪽 굽힘이 일어나지 않고 있다. 관절가동범위를 측정하는 앱을 이용하여 발목관절 가동범위(range of motion: ROM)를 측정하였고 시각 유사척도(visual analogue scale)를 이용하여 통증 정도를 측정하였다.

### 2.2 중재 방법

본 연구자가 개발한 파라핀이 첨부된 신장력이 있는 붕대(가출원 일자, 2024. 6. 15)를 적용한 것으로 발목관절을 중심으로 6겹 정도 감싼 후 15~20분 가량 유지하였다. 파라핀 붕대를 제거한 후 수동적인 발목관절 운동을 실시하였다. 주 2회 총 7회기를 실시하였다.

Table 1. Intervention method

Session	Intervention method	Intervention method
1 ~ 2	Bandage type paraffin,	15 ~ 20 minutes
3 ~ 4	Bandage type paraffin, Passive ROM exercise	15 ~ 20 minutes 5 minutes
5 ~ 7	Bandage type paraffin, Passive ROM exercise, Active assisted ROM exercise	15 ~ 20 minutes 7 minutes

## 3. 연구 결과

### 3.1 연구 결과

중재 전후의 발목관절의 가동범위와 통증의 차이는 Table 2와 같다.

7회기 중재 후에 중재전보다 발목관절의 수동적 관절가동범위는 35° 증가하였으며 통증은 4점 감소하였다.

Table 2. Intervention result

변수		Before	After
Passive ankle ROM	Dorsi flexion	70°	90°
	Plantar flexion	131°	145°
	Degree	7	3
Pain	Type	Aching pain sensation, Numb pain sensation, A tingling pain sensation	Aching pain sensation, Numb pain sensation,

## References

- Lee, Suk-Min, & Kwon, Ki-Ja (2003). The Effect of Paraffin Therapy on Release of pain and Range of motion with Joint Disease Patients. *Journal of Korean Physical Therapy Science*, Vol. 10, No. 1, pp. 130-139.
- Kim SG et al. (2023). Effectiveness of paraffin bath therapy for the symptoms and function of hand diseases: A systematic review and meta-analysis of randomized controlled trials, *Journal of Hand Therapy*, *Journal of Hand Therapy*, Vol. 36, No. 3, pp. 706-712.

## 성별에 따른 직무유연성의 차이가 비자발적퇴직 후 고용에 미치는 영향

정은은<sup>1</sup>, 홍지훈<sup>2</sup>

### 요약

본 연구는 한국패널조사(2018-2022) 자료를 사용하여 성별에 따른 직무유연성의 차이가 고용에 미치는 영향을 분석하였다. 노동이동을 외생변수로 추정하기 위하여 여성과 남성이 예기치 못한 직업상실을 경험한 경우 즉, 비자발적퇴직을 경험한 근로자를 처리군으로 두고 실직하지 않은 근로자를 대조군으로 두었으며 CEM(Coarsened Exact Matching) 방법을 사용하여 실직근로자와 비실직근로자의 특성을 동질적으로 구성하였다. 분석방법으로 이중차분법(Difference in Difference)을 활용하였으며 근로자가 비자발적 퇴직 후 일정기간이 경과한 경우에 고용에 미치는 영향을 성별에 따라 분석하였다. 분석결과, 외생적 충격으로 실직한 근로자의 고용상태는 실직 후 12개월 동안 비실직 근로자에 비해 3.03% 낮았으며, 실직 후 12개월 동안 여성 근로자의 고용 상태는 남성 근로자에 비해 약 0.85% 낮다는 결과를 도출하였다. 이는 외생적 충격으로 인한 비자발적 실직이 고용에서 성별 격차를 확대 할 수 있다는 사실을 시사한다.

주요 용어 : 직무유연성, 성별고용유연성, 이중차분법, 여성고용.

<sup>1</sup>부산대학교 경제통상대학 경제학과 박사수료. hih9529@naver.com

<sup>2</sup>(교신저자) 부산대학교 경제통상대학 경제학과 교수. gh9x@pusan.ac.kr

# Option-Ownership Alignment and Downside Risks of Multinationality

*Sangcheol Song*<sup>1</sup>

## Abstract

We extend our understanding of the relationship of multinationality to downside risk reduction by tackling two problematic assumptions that are commonly made under the multinational operational flexibility literature. We posit that since multinationality is a portfolio of growth, switching and dual options, downside risk reduction effects of multinationality tend to vary by actual portfolio composition of the three options ascribed to foreign subsidiaries. In addition, we postulate that the effect of option types may also vary in terms of their alignment with minority, balanced, and majority ownership. From the Tobit two-stage model on a large sample of Korean multinational corporations, we observe that multinationality predominantly comprised of dual options reduce downside risks more than multinationality predominantly comprising of either growth or switching options. We also find that balanced ownership stakes are best aligned with dual options for downside risk reduction. These findings imply strategic alignments between options and ownership types in order to maximize the value of multinational flexibility.

## 1. Hypotheses

Hypothesis 1: Downside risk reduction effects of multinationality will be stronger when an MNC portfolio is predominantly composed of dual options than when an MNC portfolio is predominantly composed of either growth or switching options.

Hypothesis 2: The downside risk reduction effect of dual options predominantly comprising a MNC portfolio will be stronger when the MNC portfolio has balanced ownership stakes mainly than when it has either minority or majority ownership stakes mainly.

## 2. Research Methodology

For empirical testing, we utilized a panel dataset of 345 Korean MNCs that were publicly listed and active for at least five years on the Korea Stock Exchange during 1985 to 2014 and had presence in 71 foreign countries. We followed the methods employed in Tong and Reuer's (2007) study for downside risks as the dependent variable of this study. These studies commonly measured downside risks of multinationality based on the difference in annual return

---

<sup>1</sup>Associate professor of Management, Saint Joseph's University, Philadelphia, USA. E-mail: [ssong@sju.edu](mailto:ssong@sju.edu)

on assets (ROA) between an MNC and its reference group (i.e., other MNCs in the same two-digit industry). The squared difference term between industry average ROA (IROA) and a firm's ROA was summed in the form of a second-order root lower partial moment over past five years (see formula below). By the definition of downside risks as the indicator of real option performance, only below-target performance is counted in a Tobit model.

$$\begin{aligned} & \text{Downside Risk}_{t=0} \\ &= \sqrt{\frac{1}{5} \sum_{t=0}^{t=-4} (\text{IROA}_{t-1} - \text{ROA}_t)^2 | \text{IROA}_{t-1} > \text{ROA}_t,} \end{aligned}$$

**Multinationality:** We calculated the logarithm of 1 plus the number of countries in which an MNC had its foreign subsidiaries in order to remedy the significant positive skew that was evident for the pre-transformed count measure.

**Option types:** We referred to Chung et al.'s (2015) definitions and measurements of three option types and before creating three option groups in two steps. First, subsidiaries are categorized in any one of the three options. A growth option is typically ascribed to a foreign subsidiary whose sales were predominantly focused on its host country by more than 66.6% of its total sales. On the other hand, a switching option is ascribed to a foreign subsidiary whose sales were predominantly directed at sister subsidiaries in other host countries by more than 66.6% of its total sales. Finally, the dual option is ascribed to a foreign subsidiary whose sales to either other companies in the same host country or sister affiliates in other host countries are between 33.4 and 66.6% of its overall sales. Second, we classified all MNCs into three groups, adopting a 50% cut-off. To illustrate, an MNC portfolio comprises of dual options by more than 50%, the portfolio is classified as high dual option group, while being as low dual option group otherwise. Similarly, the creation of high vs. low growth and switching options groups takes place. Thus, three different regressions run on high vs. low group per each of three options. We referred to audit reports from the database of the TS2000 Warehouse and the Data Analysis, Retrieval, and Transfer (DART) System for each subsidiary's internal and external sales information.

**Ownership types:** We first checked three ownership levels that foreign subsidiaries possess: minority ownership (less than 50%), balanced ownership (50%), and majority ownership (more than 50%). Then, we recognized three ownership groups in the same manner to three option groups using 50% cut-off before creating three dummies associated with those three ownership groups; 1 is assigned to the portfolio comprising of one of three ownership types (majority, minority, or equal ownership by more than 50%), and 0 otherwise.

## 3. RESULTS

Table 1. Estimates for the Tobit models with selectivity

	Full sample		Regression 1 with focus on dual option	
	Model 1	Model 2	High dual option portfolio	Low dual option portfolio
			Model 3	Model 3
Intercept	-0.037(0.094)	-0.050(0.040)	-0.017(0.143)	
Firm size	-0.178(0.030)	-0.249(0.028)	-0.117(0.110)	
Firm performance	-0.427(0.000)	-0.516(0.000)	0.302(0.008)	
R&D intensity	-0.128(0.074)	-0.145(0.066)	-0.098(0.106)	
Organizational slack	-0.070(0.082)	0.099(0.072)	0.114(0.040)	
International experience	-0.070(0.032)	-0.116(0.020)	-0.084(0.060)	
Expatriates	-0.117(0.090)	-0.113(0.060)	-0.520(0.102)	
GDP/capita range	0.070(0.108)	0.086(0.110)	0.030(0.118)	
Low labor cost correlation	-0.133(0.038)	-0.150(0.032)	-0.082(0.125)	
Firm dummies	Included	Included	Included	
Year dummies	Included	Included	Included	
Industry dummies	Included	Included	Included	
Multinationality		-2743(0.005)	-0.670(0.150)	
Correction for self-selection	-0.178(0.042)	-0.140(0.034)	-0.152(0.041)	
# of MNCs	345	115	230	
Wald c2	126.23(0.000)	150.44(0.000)	138.456(0.000)	
	Regression 2 with focus on growth option		Regression 3 with focus On switching option	
	High growth option portfolio	Low growth option portfolio	High switching option portfolio	Low switching option portfolio
	Model 4	Model 5	Model 6	Model 7
Intercept	-0.043(0.054)	-0.012(0.130)	-0.040(0.046)	-0.015(0.128)
Firm size	-0.220(0.044)	-0.109(0.092)	-0.203(0.038)	-0.094(0.102)
Firm performance	-0.404(0.000)	-0.156(0.006)	-0.501(0.000)	-0.190(0.000)
R&D intensity	-0.172(0.038)	-0.098(0.128)	-0.156(0.050)	-0.099(0.123)
Organizational slack	0.112(0.042)	0.117(0.090)	0.120(0.050)	0.122(0.030)
International experience	-0.102(0.060)	-0.114(0.038)	-0.098(0.044)	-0.117(0.055)
Expatriates	-0.119(0.103)	-0.128(0.050)	-0.1016(0.078)	-0.088(0.116)
GDP/capita range	0.56(0.120)	0.074(0.114)	0.074(0.112)	0.110(0.106)
Low labor cost correlation	-0.144(0.094)	-0.168(0.044)	-0.170(0.028)	-0.114(0.114)
Firm dummies	Included	Included	Included	Included
Year dummies	Included	Included	Included	Included
Industry dummies	Included	Included	Included	Included
Multinationality	-1.090(0.026)	-0.404(0.144)	-2110(0.009)	-0.890(0.134)
Correction for self-selection	-0.151(0.028)	-0.124(0.038)	-0.124(0.020)	-0.115(0.042)
# of MNCs	157	188	192	143
Wald c2	144.69(0.000)	136.26(0.000)	154.28(0.000)	142.37(0.000)

p-values in parenthesis

#### 4. FINDINGS & DISCUSSIONS

In this study, we tackle two problematic assumptions made in the extant literature on the linkage of multinationality with downside risks. The literature assumes that multinationality is a portfolio of switching options ascribed to foreign subsidiaries, which is why MNCs are able to shift operations among those subsidiaries easily and flexibly for the sake of their own benefits. However, we posit that multinationality can also comprise of growth, switching, and dual options; thus, the downside risk reduction effects of multinationality may vary based on the actual portfolio compositions of these three options. The extant literature also assumes that majority portfolio ownership allows MNCs to coordinate scattered subsidiaries and shift operations easily and quickly. In this study, we offer a novel approach in that we argue for the potential strategic alignment between portfolio option and ownership compositions from the standpoint of downside risk reduction. Our empirical results support our hypothesized relationships in the data of Korean MNCs for the period between 1985 and 2014. Our first finding is that when an MNC portfolio is predominantly composed of dual options, it reduces downside risks more than a portfolio predominantly that comprises of switching or growth options. This examination creates a gap between the growth and switching option literature, which has been developed in separate streams. Our findings regarding the positive and significant impacts of dual options in reducing downside risks of multinationality underpin the importance of fine-grained examinations of complementary and conflictual interactions among different option types in a portfolio such as multinationality.

#### REFERENCES

- Chung, C. H., Lee S-H., Beamish, P., & Isobe, T. 2010. Subsidiary expansion/contraction during times of economic crisis. *Journal of International Business Studies*, 41, 500 - 525.
- Tong, T., & Reuer, J. J. 2007. Switching options and coordination costs in multinational firms. *Journal of International Business Studies*, 38, 215 - 230.

## NASA SBIR 지원과제의 Topic Code에 따른 메타정보를 활용한 토픽모델링 연구

한혁<sup>1</sup>, 노경란<sup>2</sup>

### 81요 약

본 연구에서는 NASA의 중소기업 혁신연구 프로그램(Small Business Innovation Research, SBIR)에서 제공하는 오픈 데이터의 토픽 코드와 메타정보를 기반으로 NASA SBIR의 세부적인 지원 경향을 분석하고자 한다. NASA SBIR 프로그램은 우주 및 항공 분야의 혁신적 기술 개발을 촉진하기 위해 중소기업에게 연구개발 및 상용화 자금을 지원하는 중요한 정책 도구이다. 본 연구는 NASA SBIR 지원과제 데이터를 체계적으로 분류하고 이해하기 위해 토픽모델링 기법을 적용하여 Topic Code별 지원 경향을 제시하는 것을 목적으로 한다. 연구 방법으로는 NASA SBIR 오픈 데이터에서 제공하는 지원과제의 제목, 초록 등을 포함한 메타정보와 토픽 코드를 수집한 후, 이를 전처리하여 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 모델을 적용하여 토픽모델링을 수행하였다. LDA 모델은 문서 집합에서 주제를 자동으로 추출하는 기법으로, 각 지원과제의 텍스트 데이터를 기반으로 잠재적인 주제를 식별하였다. 연구 결과, 토픽모델링을 통해 도출된 주제들은 NASA SBIR 지원과제를 분류하는 역할을 수행하는 Topic Code별로 다양한 연구 분야와 기술적 특성을 반영하고 있음을 확인하였다. Topic Code별로 주요 주제들을 도출함으로써 각 Topic Code에 대한 세부적인 분석을 통해 지원과제의 경향성과 주제 분포를 분석하였다. 이러한 분석 결과는 항공우주분야의 정책 입안자들에게 항공우주 분야 기술혁신 프로그램의 방향성을 제시하고, 중소기업의 혁신 활동을 지원하는 데 유용한 정보를 제공할 수 있으리라 기대된다.

주요 용어: NASA SBIR, 토픽모델링, 잠재 디리클레 할당(LDA), 메타정보, 우주 및 항공 기술, 연구 트렌드, 중소기업 혁신.

---

\*이 논문은 2024년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다.(과제번호: K24L3M2C4)  
<sup>1</sup>02456 서울특별시 동대문구 회기로66 한국과학기술정보연구원 R&D투자전략연구센터 투자분석팀 책임연구원.  
E-mail: hyhahn@kisti.re.kr  
<sup>2</sup>02456 서울특별시 동대문구 회기로66 한국과학기술정보연구원 R&D투자전략연구센터 투자분석팀 책임연구원.  
E-mail: infor@kisti.re.kr

## Visual Analysis of Arctic Sea Ice Imagery

*Yunseo Cho<sup>1</sup>, Jaesung Park<sup>2</sup>, Sungchul Hong<sup>3</sup>, Jong-June Jeon<sup>4</sup>*

### Abstract

Arctic sea ice is closely linked to the global climate and serves as a critical indicator of global climate change. With the advancement of technology, numerous deep learning models have been employed in predicting the overall sea ice concentration and boundaries in the Arctic. However, challenges persist in capturing both the temporal and spatial characteristics inherent in sea ice imagery. Additionally, despite ongoing research on overall sea ice concentration and boundaries of the Arctic, studies focusing on specific regional analysis remain relatively less explored. Therefore, in this study, we provide a visual regional analysis as a subsequent research using forecasted sea ice imagery from our deep learning model Unicorn (U-Net for sea Ice forecasting using Convolutional Operation Node) developed by our research team. Our model Unicorn modifies the bottleneck of the U-Net model with convolutional neural ordinary differential equations to effectively capture the characteristics of the images. The dataset we utilized includes weekly sea ice concentration provided by the Korea Polar Research Institute (KOPRI) and regional Arctic masks from the National Snow and Ice Data Center (NSIDC). This study aims to provide insights into the regional variability of sea ice change through detailed visual analysis.

*Keywords* : sea ice, spatiotemporal forecasting, U-Net, deep learning.

---

<sup>1</sup>Graduate student, Department of Statistical Data Science, University of Seoul, 163, Seoulsiripdae-ro, Dongdaemun-gu, 02504, Seoul, Korea. Email: friend0429@uos.ac.kr

<sup>2</sup>Graduate student, Department of Statistical Data Science, University of Seoul, 163, Seoulsiripdae-ro, Dongdaemun-gu, 02504, Seoul, Korea. Email: wkds0804@uos.ac.kr

<sup>3</sup>Postdoctoral researcher, Department of Statistics, University of Seoul, 163, Seoulsiripdae-ro, Dongdaemun-gu, 02504, Seoul, Korea. Email: shong@uos.ac.kr

<sup>4</sup>(Corresponding Author) Professor, Department of Statistics, University of Seoul, 163, Seoulsiripdae-ro, Dongdaemun-gu, 02504, Seoul, Korea. Email: jj.jeon@uos.ac.kr

## 강수량에 의한 미세먼지의 변화

김도현<sup>1</sup>, 문우형<sup>2</sup>, 김준현<sup>3</sup>, 박준범<sup>4</sup>

### 요약

미세먼지의 노출이 많은 현대에 인체에 유해하며, 미관에 영향을 끼치게 된다. 미세먼지에는 입자크기가  $10\mu\text{m}$ 이하인 PM10과  $2.5\mu\text{m}$ 이하인 PM2.5(초미세먼지), 이산화질소, 일산화탄소, 아황산가스, 오존이 포함되어 있다. 통상적으로 비가 내리면 대기중의 미세먼지를 씻어준다는 이야기가 있다. 본 고에서는 이를 토대로 서울시 강남구를 대상으로 하여 2023년도의 강수량과 미세먼지 수치의 데이터를 활용하여 이를 분석하였다. 그리고 분석한 내용을 그래프로 시각화하였으며 강수량으로 인한 미세먼지 변화에 대해서 분석하였다. 분석 결과, 계절별, 히트맵으로 나타내어 강수량이 많으면 미세먼지 농도의 수치가 낮아지는 상관관계를 확인할 수 있었다. 또한 본 연구에서 변수로 사용한 강수량을 적용했을 때 3-4%의 정확도(Accuracy)가 향상되었다. 본 연구에서 진행한 미세먼지와 강수량에 대한 분석이 향후 미세먼지 예측에 도움이 될 것으로 기대된다.

### 1. 서론

미세먼지의 노출은 인체에 유해하며, 미관에 영향을 끼치는 등 현대 사회의 심각한 문제 중 하나로, 대기 오염의 주요 원인으로 작용하고 있다. 미세먼지에는 입자의 크기가  $10\mu\text{m}$ 이하의 입자인 미세먼지 PM10과  $2.5\mu\text{m}$ 이하인 초미세먼지 PM2.5이 있으며, 그 안에는 이산화질소, 일산화탄소, 아황산가스, 오존이 포함되어 있다. 이들은 다양한 부정적인 영향을 끼치는 요인이 되어 호흡기 및 심혈관계에 심각한 질환을 유발할 수도 있다. 때문에 현재 많은 국가와 단체에서 미세먼지에 대한 원인과 해결 방안에 대해 연구를 하고 있다. 통상적으로 많이 들려오는 것 중 비가 내리면 미세먼지를 씻어준다는 이야기가 있다. 본 고에서는 이를 토대로 미세먼지와 강수량에 따른 미세먼지 농도 변화에 대한 상관관계를 명확히 알아내기 위해 서울시 강남구를 대상으로 2022년 12월부터 2023년 11월까지의 강수량과 미세먼지 수치의 데이터를 가지고 이를 분석하였다. 그리고 분석한 내용을 그래프로 시각화하여 강수량에 의한 미세먼지의 변화와 계절별 변화에 대해서 상관관계를 알아보며, 미세먼지 변화 패턴에 대해 예측 강수량에 따른 예측의 accuracy 변화를 분석하였다.

<sup>1</sup>31962 충청남도 서산시 해미면 한서1로 46 한서대학교 항공컴퓨터전공 학부연구생.  
E-mail: kdh1102@naver.com

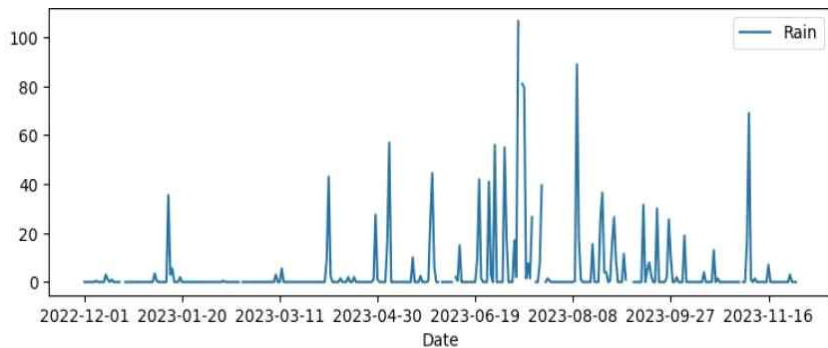
<sup>2</sup>31962 충청남도 서산시 해미면 한서1로 46 한서대학교 문화재보존학과 학부연구생.  
E-mail: dngud1013@naver.com

<sup>3</sup>31962 충청남도 서산시 해미면 한서1로 46 한서대학교 항공컴퓨터전공 학부연구생.  
E-mail: kjh03031000@naver.com

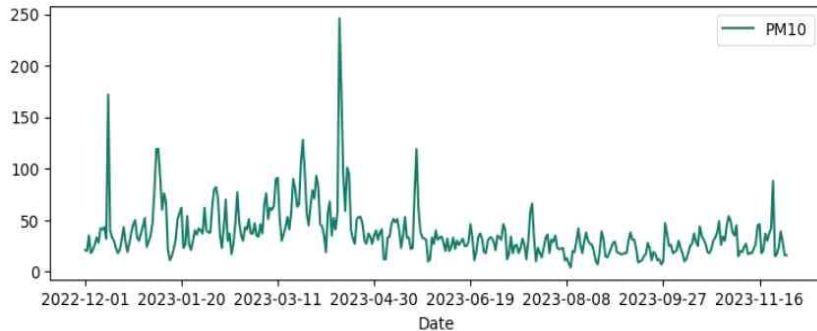
<sup>4</sup>(교신저자) 31962 충청남도 서산시 해미면 한서1로 46 한서대학교 항공컴퓨터학과 조교수.  
E-mail: jbpark@hanseo.ac.kr

## 2. 미세먼지와 강수량 데이터 분석 및 시각화

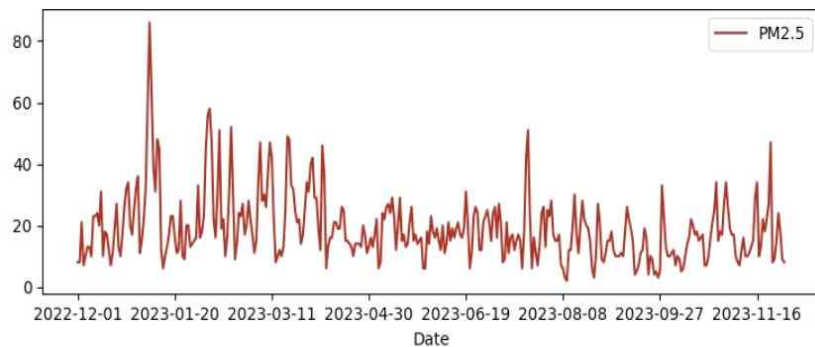
본 고에서 서울시 강남구의 미세먼지 데이터와 강수량 데이터를 수집하여 분석하고 데이터를 그래프로 시각화하였다. 미세먼지 데이터는 에어코리아에서, 강수량 데이터는 서울 열린데이터광장에서 얻어 사용하였다. 데이터 수집 기간은 2022년 12월부터 2023년 11월까지 설정하여 일별로 정리하였다. 정리한 데이터는 각각의 수치를 백분율로 환산하고 그래프로 시각화 하였으며, 미세먼지 데이터의 경우 확실한 입자가 존재하고 가장 대표적으로 언급되는 미세먼지 PM10과 초미세먼지인 PM2.5만을 다루었다.



Graph1. Graph of Rainfall



Graph2. Graph of PM10



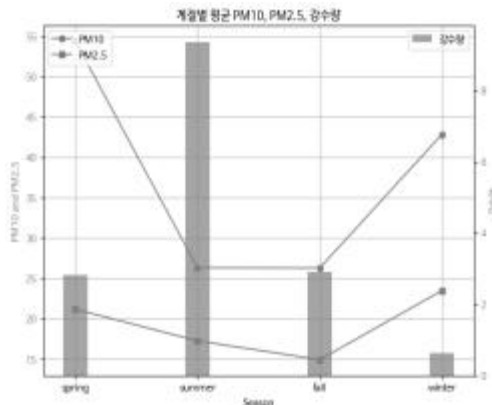
Graph3. Graph of PM2.5

그래프에서 확인할 수 있듯, 서울시 강남구의 미세먼지 농도는 강수량에 따라 확연한 변화를 보인다. 강수량이 적은 날의 다른 미세먼지 그래프를 확인해보면 PM10과 PM2.5의 수치가 높게 나오는 것을 확인할 수 있으며, 강수량이 많았던 날과 동일한 날짜의 PM10과 PM2.5의 그래프를 보면 앞의 경우에 비해 미세먼지 농도가 낮게 나오는 것을 확인할 수 있다. 이러한 시각화와 분석의 결과를 통해 미세먼지 농도가 강수량에 의한 변동이 있음을 확인할 수 있다.

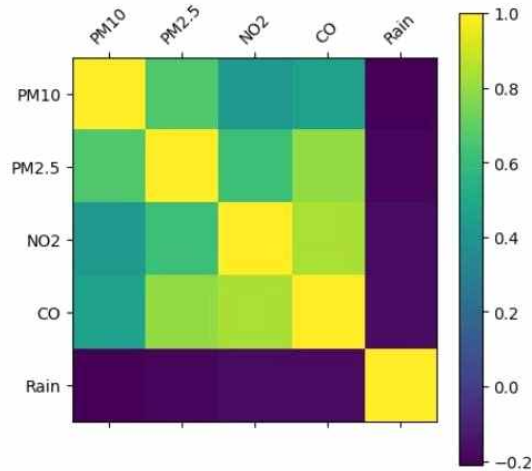
### 3. 미세먼지와 강수량의 상관관계

앞서 미세먼지 PM10과 초미세먼지 PM2.5, 강수량 데이터를 각각의 그래프로 시각화한 뒤 분석했으며, 이를 토대로 계절별 미세먼지 수치와 강수량의 상관관계에 대해서 분석하기 쉽도록 시각화해보았다. 미세먼지와 강수량의 데이터의 상관관계를 직관적으로 시각화하여 변수들의 상관성을 파악할 수 있는 히트맵을 만들어 강우가 미세먼지 농도에 미치는 영향을 분석하였다.

그래프4는 미세먼지 농도와 강수량의 계절별 상관관계를 다뤘으며 이는 계절에 따라 다르게 나타난다. 구체적인 분석을 위해 본 그래프를 참고하여 봄, 여름, 가을, 겨울에 대한 미세먼지 농도와 강수량 사이의 연관성을 분석하였다. 봄은 기온이 상승하며 대기의 순환으로 미세먼지의 확산이 큰 계절이며, 황사와 같은 외부 요인으로 미세먼지의 농도가 높은 계절이다. 사계절 중 강수량이 비교적 적은 계절에 속해있어, 그래프에서 볼 수 있듯이 비교적 낮은 강수량과 높은 미세먼지 수치를 확인할 수 있다. 여름은 강수량이 가장 많으며 장마가 발생하는 계절이다. 위와 같이 사계절 중 강수량이 제일 높은 수치를 기록하며 앞선 시각화에서 본 것처럼 확연히 낮아지는 미세먼지 농도를 확인할 수 있다. 가을의 강수량은 봄과 비슷한 수준이지만 가을의 경우 황사와 같은 외부 요인이 없어 봄과는 다르게 비슷한 강수량에 비해 낮은 수치가 기록됨을 알 수 있다. 겨울은 대기 정체가 발생하는 계절로 가장 낮은 강수량과 비교적 높은 수준의 미세먼지 농도를 확인할 수 있다.



Graph4. Graph of Average Rainfall, PM10, PM2.5 by season



Graph5. Heatmap of correlation between dust and precipitation

그래프5에서 PM10, PM2.5, 이산화질소, 이산화탄소, 비를 다루어서 미세먼지와 비와의 상관관계를 분석하였다.

PM10은 PM2.5와 높은 수치의 상관관계를 나타내며 서로 밀접한 연관이 있음을 확인할 수 있으며, NO2와 CO는 보통의 수치의 상관관계를 나타내며 서로 일부의 연관이 있음을 확인할 수 있으며, Rain과는 낮은 수치의 상관관계를 나타낸다. PM2.5는 PM10과 높은 수치의 상관관계를 나타내며 서로 밀접한 연관이 있음을 확인할 수 있으며, NO2와 CO는 PM10과 마찬가지로 보통의 수치의 상관관계를 나타내며 Rain과는 낮은 수치의 상관관계를 확인할 수 있다. NO2와 CO는 PM10, PM2.5, CO와 보통의 수치의 상관관계를 나타내며 일부 연관이 있음을 확인할 수 있으며 Rain과는 낮은 수치의 상관관계를 확인할 수 있다. Rain은 모든 미세먼지 변수들과 무의미할 정도의 낮은 수치의 상관관계를 확인할 수 있다.

또한, Rain을 변수 유무에 따라 SVM Classification 모델 성능 테스트시 Rain을 변수로 넣었을 때 3-4%의 향상된 Accuracy를 갖는 것으로 나타났다. 위와 같이 시각화한 상관관계를 확인함으로써, 미세먼지는 강수량에 의해 그 농도의 수치에 영향을 받고 변화한다는 결과를 보였다.

#### 4. 결론

우리가 직접 강수량과 미세먼지 농도의 관계를 측정된 결과, 강수량이 많을수록 미세먼지 농도가 낮아지는 경향이 있음을 확인했다. 이러한 관찰을 바탕으로 강수량이 높을수록 미세먼지 수치가 낮아진다는 예측을 세울 수 있다. 이 예측은 여러 과학적 원리에 의해 뒷받침될 수 있다. 우리의 측정 결과는 이러한 이론적 배경과 일치한다. 강수량이 많은 날에는 미세먼지 농도가 낮아지는 패턴이 반복적으로 관찰되었으며, 이는 비가 대기 중의 오염 물질을 효과적으로 제거한다는 것을 시사한다. 따라서 강수량이 많을수록 미세먼지 수치가 낮아진다는 예측은 우리 측정 결과에 의해 뒷받침될 수 있다.

그러나 이 예측이 항상 모든 상황에서 절대적으로 적용되는 것은 아니다. 비가 내린 후에도 새로운 미세먼지가 발생할 수 있고, 강수량의 크기와 지속 시간, 지역의 지리적 특성 등에 따라 미세먼지 제거 효과는 다를 수 있다. 예를 들어, 짧은 시간 동안 많은 비가 내린 경우와 오랜 시간 동안 적은 비가 내린 경우, 미세먼지 제거 효과는 달라질 수 있다.

종합적으로, 우리가 수행한 측정 결과는 강수량과 미세먼지 농도 사이의 명확한 상관관계를 보여준다. 강수량이 많을 때 미세먼지 농도가 낮아지는 경향을 확인한 바, 이러한 관찰을 토대로 강수량이 높을수록 미세먼지 수치가 낮아진다는 예측을 세우는 것은 과학적으로 타당하다고 할 수 있다. 이 예측은 대기 오염을 줄이고, 공기 질을 개선하기 위한 기상 조건의 중요성을 강조하는 데 기여할 수 있는 것으로 보여진다.

본 연구에서는 SVM 분류 모델에서 강수량 변수를 포함했을 때, 예측 정확도가 약 3-4% 향상되는 결과를 보였다. 이러한 상관관계를 통해 미세먼지 농도가 강수량에 의해 영향을 받는다는 것을 확인할 수 있다. 이는 강수량이 PM10 농도 예측에 중요한 변수임을 시사한다.

본 연구를 이어서 향후 연구로 장기간의 강수량 및 미세먼지 데이터를 수집·분석하여 기후 변화와 계절적 변동성이 미세먼지 농도에 미치는 영향을 명확히 파악하고, 다양한 기후 조건에서의 비교 분석을 통해 지역별 맞춤형 대기 질 관리 방안에 대해 연구할 예정이다.

## References

Fine dust data - Air Korea

Precipitation data - Meteorological Data Opening Portal

Park.H.J(2022) "Analysis and prediction of fine dust based on weather data and atmospheric environment data" 328-337 (in Korea)

Explanation of Climate Change Monitoring Terminology

Shin.J.H, Kim.H.W, Park.S.Y(2018) "Correlation between Fine Dust and Precipitation" 194-195 (in Korea)

Han.S.J, Eom.J.S, Lee.S.S, Jeong.C.H(2022) "Long-term Spacetime Distribution of Precipitation and Fine Dust in Seoul" 453-453 (in Korea)

Lee.J.H(2020) "A Study on the Influencing Factors related to Fine Dust Concentration in Seoul" (in Korea)

## 머신러닝과 딥러닝 모형을 이용한 '신고' 배나무의 수액 흐름 예측에 관한 연구\*

최원기<sup>1</sup>, 김동연<sup>2</sup>, 김민수<sup>3</sup>

### 요 약

나무의 생장을 이해하기 위해서는 수액 흐름을 측정하는 것이 필요하다. 그러나 모든 농가에서 수액 흐름을 측정하는 것은 현실적으로 어렵다. 본 연구에서는 이러한 농가들을 위하여 기상 변수와 DAFB(Days After Full Bloom)만을 이용하여 '신고' 배나무의 수액 흐름을 예측하는 모델을 개발하였다. 기상 변수와 수액흐름 간의 시간 지연 효과를 파악하기 위해 스피어만 상관 분석을 실시하였고 일사량의 경우 5시차 전의 계수가 최대 계수임이 나타났다. 비선형 관계를 고려하여 예측 모델로는 머신러닝 모델인 Random Forest, XGBoost를 사용하였고 딥러닝 모델인 LSTM(Long Short-Term Memory), GRU(Gated Recurrent Unit), BiLSTM(Bidirectional LSTM), CNN-GRU-BiLSTM 모델을 사용하여 수액 흐름을 예측하였다. 일반적으로 나무의 수액 흐름은 시민박명시간을 기준으로 가우시안 분포를 따르는 경향을 보인다. 이러한 패턴을 효과적으로 학습하기 위해, 시민박명시간을 기준으로 한 새로운 시간 변수를 도입함으로써 모델의 예측 성능을 개선하였다. 예측 결과 CNN-GRU-BiLSTM 모델이 가장 좋은 성능을 보였고 기상 변수와 DAFB만을 이용하여 수액 흐름을 예측하는 것이 가능하다는 것을 보여주었다. 이러한 연구결과는 해외 사례를 적용하기 어려웠던 우리나라 농가에 도움을 줄 수 있을 것으로 생각되며 향후 이러한 모형을 적용한다면 데이터 수집이 어려운 증산에 대한 연구의 기초 자료로 사용될 수 있을 것으로 기대된다.

주요 용어 : 신고, 수액 흐름, DAFB, CNN-GRU-BiLSTM.

\*본 연구는 농림축산식품부의 재원으로 농림식품기술기획평가원의 노지분야 스마트농업기술 단기고도화사업의 지원을 받아 수행되었음 (과제번호: 322034-3).

<sup>1</sup>61186 광주광역시 북구 용봉로 77, 전남대학교 통계학과 학부재학생. E-mail: wongi8994@gmail.com

<sup>2</sup>61186 광주광역시 북구 용봉로 77, 전남대학교 수학/통계학과 석사. E-mail: ehddjs0717@gmail.com

<sup>3</sup>(교신저자) 61186 광주광역시 북구 용봉로 77, 전남대학교 통계학과 교수. E-mail: kimms@chonnam.ac.kr

## 생육기간 기상 및 생산량, 수출량을 활용한 배추 가격 예측

김혜진<sup>1</sup>, 김민수<sup>2</sup>

### 요약

농산물의 가격은 매년 소비자와 재배 농가에 파동을 일으킨다. 배추 가격 안정을 위한 방안으로, 본 연구에서는 배추 가격을 형성하는 데에 영향을 미치는 여러 요인들의 상관성과 인과관계를 살펴본 후 변수들을 선별해 가격을 예측하고자 한다. 배추 가격에 직접적인 영향을 주는 공급 요인인 생산량과 생산량에 영향을 줄 수 있는 주산지 기상데이터, 대표적인 수요 요인인 수출량 데이터를 비롯한 기타 요인들이 가격과 어떤 상관성을 띄고 있는가에 대해 분석했다. 각 변수들과 가격의 상관성을 파악하고자 피어슨 상관계수(Pearson correlation coefficient)를 확인했으며, 변수들이 시간에 따른 인과성을 띄고 있는지 확인하기 위해 Granger 인과성 검사를 시행했다. 가격과 작형별 주산지 기상데이터 상관계수는 파종기, 정식기, 수확기, 전체기간 중 수확기와 전체기간의 데이터가 0.50 이상의 값을 보여 주요 변수로 선별되었다. 인과성 검사를 통해 생산량, 수출량 데이터와 가격에 대한 상호간 인과성을 확인했다. 위 결과를 토대로 변수를 선별해 LSTM 모형으로 예측을 시행했으며 가격과 생산량, 수출량 데이터를 활용한 경우가 가장 성능이 좋았다. 본 연구를 통해 배추의 가격 예측에 효과적으로 활용되는 변수를 선별할 수 있으며 그로 인해 예측 모형의 성능을 향상시킬 수 있을 것으로 기대된다.

주요 용어: 배추 가격, 피어슨 상관계수, Granger 인과성 검사, LSTM.

## 기계학습기반 모형의 토양수분 예측성능 비교평가\*

임호진<sup>1</sup>, 김민수<sup>2</sup>

### 요약

토양 수분은 농업 생산성과 생태계 건강을 유지하는 데 중요한 역할을 하며, 가뭄 및 홍수 예측, 농작물 관리, 수자원 관리 등에 필수적인 요소이다. 따라서 정확한 토양 수분 예측은 농업 효율성 증대와 환경 보호에 중요한 역할을 한다. 본 연구에서는 기상청 기상자료 개방포털의 방재기상관측(AWS) 및 농업기상관측(AAOS)에서 제공하는 '화순시 능주면'의 기상 및 토양 데이터를 활용하여 시간 단위 토양 수분 예측을 위한 여러 기계학습 및 딥러닝 모형의 성능을 비교 분석하였다. 연구에 사용된 모형은 기존의 토양 수분 예측 연구에 주로 활용된 장단기 메모리(Long Short-Term Memory: LSTM), 서포트 벡터 회귀분석(Support Vector Regression: SVR), 번역 모형으로 처음 등장했으나 시퀀스 데이터의 예측에 효과적인 것으로 알려진 Transformer, 그리고 시간적 및 정적 특성을 통합하여 예측 정확도를 향상시킨 Temporal Fusion Transformer (TFT)이다. 2021년 1월 1일부터 2023년 7월 12일까지의 데이터를 사용하여 2023년 8월 4일부터 2023년 12월 31일까지의 토양 수분을 예측하였다. 예측 성능 평가는 평균 제곱근 오차(Root Mean Square Error: RMSE)로 비교한 결과, Temporal Fusion Transformer가 가장 높은 예측 성능을 보였다. 또한 최소 무강우 시간을 고려하여 시간당 강수량이 아닌 강수 이벤트별 누적 강수량을 이용했을 때 예측 성능이 증가하였으며, 특히 강우 기간의 예측 성능이 크게 향상되었다.

주요 용어 : LSTM, SVR, Transformer, Temporal Fusion Transformer, 최소무강우시간.

\*이 논문은 교육부 및 한국연구재단의 4단계 두뇌한국21 사업 (4단계 BK21 사업, NO.5120200913674)으로 지원된 연구임.

<sup>1</sup>61186 광주광역시 북구 용봉로 77 전남대학교, 수학/통계학과 석사과정. E-mail: ghwls7001@naver.com

<sup>2</sup>(교신저자) 61186 광주광역시 북구 용봉로 77 전남대학교, 교수. E-mail: kimms@chonnam.ac.kr

## Large Scale Road-Traffic Noise Mapping with AI Technology

*Sungsu Choi<sup>1</sup>, Taeho Park<sup>2</sup>, Jong-June Jeon<sup>3</sup>*

### Abstract

Our research begins with whether artificial intelligence can understand the mechanism of traffic noise through data generated by a physical model and predict the results. The typical noise prediction model consists of a noise generation function that takes inputs such as traffic volume, speed, and vehicle type, and an energy attenuation function due to structural surfaces, transmission medium like air, etc. We construct a Convolutional Neural Network (CNN) regression model that learns noise predictions from multi-channel images representing the information used as inputs in traditional noise prediction models. Applying this CNN regression model to predict the average noise levels of peak-over time in Gwangju, South Korea, we observed improvement in generalized predictive performance compared to existing machine learning models that summarize input values on a grid basis. Through the artificial neural networks, we successfully reproduced the physical model of noise prediction and efficiently propagated noise prediction values. Additionally, we investigate model structures and database construction methods using CNN to enhance noise prediction performance.

*Keywords* : Artificial Intelligence, Convolutional Neural Network, Domain Adaptation, Noise Map, Regression model.

---

<sup>1</sup>Graduate student, Department of Statistical Data Science, University of Seoul, 163, Seoulsiripdae-ro, Dongdaemun-gu, 02504, Seoul, Korea. Email: ttjh1234@uos.ac.kr

<sup>2</sup>Researcher, Korean Environment Institute, 370, Sicheong-daero, 30147, Sejong-si, Korea.  
Email: thpark@kei.re.kr

<sup>3</sup>(Corresponding Author) Professor, Department of Statistics, University of Seoul, 163, Seoulsiripdae-ro, Dongdaemun-gu, 02504, Seoul, Korea. Email: jj.jeon@uos.ac.kr

## Bibliometric Analysis of Depression Research Among Young Adults in Korea: A Preliminary Analysis

*Areum Kim<sup>1</sup>, Haim Lee<sup>2</sup>, You Kyung Kim<sup>3</sup>, Han Wool Jung<sup>4</sup>, Byunghoon “Tony” Ahn<sup>5</sup>*

### Abstract

This study examined trends and key themes in depression research among young Korean adults using topic modeling to understand core themes, semantic structures, and changes in research topics over time. We analyzed 602 PubMed publications from 2007 to 2024 using VOSviewer, and 14,973 Naver news articles from 2021 to 2024 via TEXTOM. By combining academic research and news media, we aimed to understand individual and societal factors influencing young adults' depression in Korea. From the PubMed publications, identified topics included 'socioeconomic factors', 'COVID-19', 'sex factors', 'age', and 'smartphone use'. From the news articles, topics included 'crimes', 'suicide', 'public health', 'social support', and 'the economic aspects of depression treatments'. Academic research highlighted links between young adults' depression, eating habits, and body image, affecting quality of life. News articles emphasized connections to social and economic aspects, such as sexual crimes and digital mental health services. The study underscores the multifaceted nature of depression, influenced by societal factors such as gender roles and economic pressures, as well as individual characteristics like age and eating habits. These findings highlight the need for comprehensive understanding and interventions for young adults' depression.

*Keywords* : Depression, Young adults, Bibliometric analysis, South Korea.

---

<sup>1</sup>용인세브란스병원 정신건강의학과, 연구원

<sup>2</sup>이화여자대학교 심리학과, 상담사

<sup>3</sup>용인세브란스병원 정신건강의학과, 연구원

<sup>4</sup>용인세브란스병원 정신건강의학과, 연구원

<sup>5</sup>백길대학교 의학 및 건강 과학과, 연구원

## 일반고 학생의 학교참여 유형 탐색

조은수<sup>2</sup>

### 요 약

본 연구는 인문계 일반고 고등학생의 학교참여 유형을 탐색하여, 사회복지 개입을 위한 기초 자료를 마련하고자 수행되었다. 연구대상은 2005한국교육종단연구(KELS2005) 4차(2008년)자료이며, 직업계 교육을 받지 않는 일반고 학생 중에서 주요 변수에 응답한 고등학교 1학년 학생 4065명을 대상으로, SPSS요인분석, AMOS측정모형 적합도를 이용하여 학교참여에 대한 요인을 추출하였다. 분석 결과, 첫째, 인지적 참여 구성 변수는 “인지조절전략”의 하위 4개의 범주 중에서, 3개(조직화 전략, 초인지 전략, 정교화 전략)가 추출되었다. 둘째, 정서적 참여 구성 변수는 학교에 대한 정서적 반응, 교사에 대한 정서적 반응, 학우에 대한 정서적 반응의 3가지를 투입하여 교사와 학우에 대한 애착만 추출되었다. 셋째, 행동적 참여 구성 변수는 학습에 대한 참여적 행동을 의미하는 “토론 참여”, “학원/과외 시간”, “학교 방과후 프로그램 참여”, “학생회 활동” 등은 요인분석과 모델 적합도를 토대로 탈락하고, “숙제공부시간”, “수업 전 내용 환기”, “교과서 외 자료활용”, “수업집중도”, “독서시간”, “문화활동”, “동아리활동”으로 구성되었다. 본 연구에서 발견된 변수들을 토대로 학생들의 참여 유형을 식별한다면, 사회복지적 개입을 위한 낮은 참여 집단을 찾아낼 수 있을 것이며, 고참여집단과 저참여집단에 영향을 미치는 요인들도 찾아낼 수 있을 것이다.

## 문화 이벤트 관람과 만족도 및 향후 참여 의향 분석: 인천광역시를 중심으로

김하니<sup>1</sup>

### 요 약

본 연구는 문화빅데이터플랫폼이 제공하는 문화향수실태조사 통계 데이터를 기반으로 문화 이벤트 직접 관람횟수에 대한 분석을 하고, 상관분석을 통하여 인천광역시의 문화 이벤트 관람 수, 만족수 및 향후 참여 의향수간의 관계를 알아보고자 한다. 또한 어떤 그룹들이 있는지를 알기 위해 계층적 군집분석을 수행한다. 데이터 안에는 문학 행사, 미술 전시, 해외 음악, 전통 예술, 연극, 뮤지컬, 무용, 영화, 대중음악 엔터테인먼트로 9가지로 구성 되어있었으며 엔터테인먼트 각각의 관람수, 만족수 및 참여 의향수 값이 레코드로 있었다. 원활한 분석을 위해 IQR이상치 제거를 한 뒤 막대그래프를 그려 문화 이벤트 직접 관람 횟수 추세를 알아보았으며, 상관분석을 진행하고 히트맵으로 시각화하였다. 막대그래프 분석 결과 인천광역시 거주 국민이 가장 많이 즐기는 문화 이벤트는 영화인 것으로 나타났으며, 반대로 전통예술 관람수는 저조한 것을 알 수 있었다. 또한 해가 거듭될수록 특정 문화 이벤트 관람수(무용, 영화, 대중음악)가 줄어들고 있는 것을 확인하였다. 상관분석 결과 여러 개의 유의미한 상관이 도출되었다. 그 중에서도 성격이 매우 다른 예술 장르인데도 상관이 높다는 점이 흥미로웠다. 이러한 분석을 통해 인천광역시의 지역 맞춤형 정책 수립을 세우는 데 도움이 될 것으로 기대된다.

주요 용어 : 인천광역시, 문화 이벤트, 막대그래프, 상관분석, 군집분석.

### 1. 서론

현대 사회에서 문화 이벤트의 역할은 단순히 여가 시간의 채움을 넘어서, 개인의 삶의 질을 향상시키는 중요한 요소로 자리 잡고 있다. 문화 이벤트는 스트레스 해소, 창의성 증진, 사회적 교류 강화 등의 다양한 긍정적 효과를 제공하여 개인의 삶의 질을 전반적으로 향상시킨다. 이러한 문화 이벤트는 음악, 미술, 연극, 영화 등 다양한 분야에서 개최되며, 이는 지역사회 내에서 문화적 다양성을 증진시키는 동시에, 경제적 발전을 촉진하는 중요한 매개체 역할을 한다. 특히, 이런 문화 이벤트들은 지역사회에 활력을 불어넣고, 시민들의 삶의 만족도를 높이는 데에 크게 기여한다.

본 연구는 인천광역시를 중심으로 문화 이벤트 관람과 만족도 및 향후 참여의향을 분석하여 문화 정책 수립에 기여하고자 한다. 이를 위해 문화빅데이터플랫폼에서 제공하는 2019~2022년 문화향수실태조사(문화체육관광부)데이터를 활용한다. 인천광역시 거주 시민의 문화 이벤트 참여수 변화를 확인하고, 다양한 문화 이벤트 참여(관람)수와 만족도 및 향후 참여 의향에 대한 상관분석을 통해 어떠한 상관을 보이고 있는지 연구한다. 또한 어떠한 그룹이 있는지 확인을 위해 계층적 군집분석을 수행한다. 이러한 분석 결과를 바탕으로 인천광역시가 적합한 문화정책을 수립하는 데

<sup>1</sup>12807 경기도 광주시 초월읍 독고개길 14-24 강남대학교 인공지능융합공학부 학부재학생.

참고할 수 있도록 기여하고자 한다. 또한, 본 연구는 인천광역시뿐만 아니라 다른 지역사회에서도 문화 이벤트의 가치를 재평가하고, 이를 통해 지역사회 발전에 기여할 수 있는 방안을 모색하는데에 참고 자료가 될 것이다.

## 2. 연구대상 및 분석방법

### 2.1. 연구대상

본 연구에서는 문화빅데이터플랫폼에서 배포하는 문화향수실태조사 통계 데이터를 분석에 활용하였다. 해당 데이터는 일반 사용이 허용된 데이터이다. (문화체육관광부 2019~2022 국민문화예술 활동조사(보고서))조사대상은 만 15세 이상 남녀 10,000명이다. 관람수: 지역별 직접 관람횟수(평균), 만족도: 지역별 관람 만족도(평균), 의향: 지역별, 행사별 직접관람 의향 이렇게 크게 세 파트로 구성이 되어있다. 조사 방법은 조사원에 의한 1:1 가구방문 면접조사이다. 문화 이벤트는 순서대로 문학 행사, 미술 전시, 해외 음악, 전통 예술, 연극, 뮤지컬, 무용, 영화, 대중음악 엔터테인먼트로 9 가지 종류로 구성되어있다.

### 2.2. 분석방법

보다 정확한 분석을 위하여 데이터 전처리를 수행한다. 해당 연구에서는 인천광역시를 중심으로 하기에 인천광역시를 제외한 나머지 지역은 삭제한다. 결측이 있는 경우 해당 데이터는 평균값으로 대체한다. 이후 이상치 처리를 위하여 먼저 박스플롯을 그려보고 이상치의 유무를 파악한다. 이후 IQR 범위를 벗어나는 값을 이상치로 간주하고 제거한다. 아래 수식과 같이 정리된다.

$$IQR = Q_3 - Q_1$$

이상치 제거 후 직접 관람횟수(평균) 데이터를 활용해 인천광역시, 연도별(2019년-2022년)로 그룹화하여 막대그래프로 시각화 한다.

이후 각 문화 이벤트의 관람수와 만족도 및 향후 참여 의향 컬럼을 그룹화하여 상관 분석을 진행할 것이며, 상관분석을 통하여 인천광역시의 문화 이벤트 관람수, 만족수 및 향후 참여 의향수간의 관계를 알아볼 것이다. 또한 계층적 군집분석을 수행하여 어떤 그룹이 있는지 알아보려고 한다.

## 3. 분석결과

### 3.1. 문화 이벤트 관람수에 대한 분석

인천광역시의 문화 이벤트 관람수에 대한 분석을 막대그래프로 시각화하였다. X축은 2019~2022 년도이다. 시각화 결과 전반적으로 인천광역시 거주 국민이 가장 많이 즐기는 문화 이벤트는 영화인 것으로 나타났으며, 반대로 가장 즐기지 않는 문화 이벤트는 전통예술로 나타났다. 또한 해가 거듭될수록 특정 문화 이벤트 관람수(무용, 영화, 대중음악)가 줄어들고 있는 것을 확인하였다. 이는 Figure 1과 같이 나타났다. 범례 순서대로 문학 행사, 미술 전시, 해외 음악, 전통 예술, 연극, 뮤지컬, 무용, 영화, 대중음악의 관람수이다.

### 3.2. 문화 이벤트 참여와 만족도 및 향후 참여 의향에 대한 상관분석

인천광역시의 문화 이벤트에 대한 상관분석을 진행한 결과 흥미로운 상관관계가 관찰되었다. 첫 번째, 성격이 매우 다른 예술 장르인 문학행사관람수와 해외음악관람수의 상관계수가 0.91로 도출된 점이다. 두 번째, 미술전시관람수와 영화관람수의 상관계수가 0.83으로 도출된 점이다. 이 둘은 시각적 예술이라는 공통점이 있으나 성격이 매우 다르다. 세 번째, 무용 관람수와 연극만족수의 상관계수가 0.92로 도출된 점이다. 네 번째, 전통예술관람수와 뮤지컬관람수의 상관계수가 0.72로 도출된 점이다. 다섯 번째, 영화관람수와 전통예술의향수의 상관계수가 0.99로 매우 높게 도출된 점이다.

### 3.3 문화 이벤트 참여와 만족도 및 향후 참여 의향에 대한 군집분석

분류에는 계층적 군집분석(위드법, 대상 간의 비유사도는 제곱 유클리드 거리의 1/2배)을 이용하였다. 덴드로그램을 관찰하고 엘보우 기법을 수행한 결과 3개의 군집일 때 해석가능성이 높음을 나타내었다. 각 군집의 군집평균으로 고찰하면 클러스터 1은 예술 전시, 극장, 댄스, 영화, 인기 음악 공연 참여와 만족도, 그리고 다양한 문화 이벤트에 대한 의도가 높은 그룹이다. 클러스터 2는 문학 이벤트, 해외 음악, 전통 예술, 극장, 뮤지컬, 댄스, 영화, 인기 음악 공연 참여와 만족도, 그리고 특정 문화 활동에 대한 의도가 높은 그룹이다. 클러스터 3은 문학 이벤트, 예술 전시, 해외 음악, 문화 이벤트 만족도, 뮤지컬과 영화 만족도, 그리고 특정 문화 활동에 대한 의도가 높은 그룹이다. 클러스터 4는 전통 예술, 극장, 뮤지컬 참여와 만족도, 해외 음악 만족도, 그리고 특정 문화 활동에 대한 의도가 높은 그룹이다.

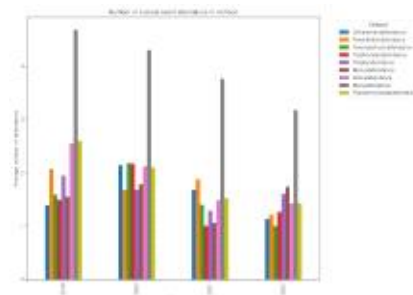


Figure 1. 인천광역시의 문화 이벤트 관람수 막대그래프 분석

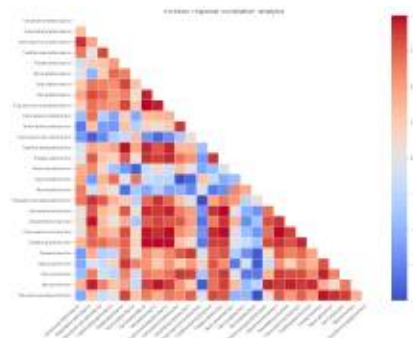


Figure 2. 인천광역시의 문화 이벤트 관람수, 만족도 및 향후 참여 의향 상관분석

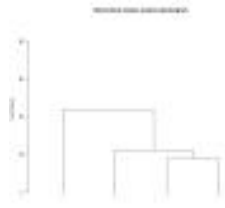


Figure 3. 인천광역시의 문화 이벤트 관람수, 만족도 및 향후 참여 의향 군집분석

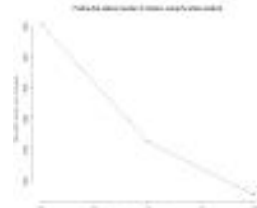


Figure 4. 엘보우 기법 결과

#### 4. 결론 및 제언

본 연구에서는 2019년부터 2022년까지의 인천광역시의 문화 이벤트 관람수의 변화를 분석하였다. 특히 영화 관람이 가장 인기 있는 문화이벤트임에도 불구하고 전반적인 관람수가 감소하는 추세를 확인하였다. 이러한 추세는 디지털 콘텐츠의 소비 증가, 경제적 요인, 코로나19 팬데믹 등 여러 외부적 요인들이 문화 이벤트에 대한 참여에 영향을 미쳤을 가능성을 시사한다. 그리고 상관분석을 통해 각 파트에 관한 흥미로운 상관 관계를 확인하였다. 또한 계층적 군집분석을 수행하고 덴드로그램을 그려 어떤 그룹들이 있는지 알아보았다.

이러한 분석을 통해 인천광역시의 지역 맞춤형 정책 수립을 세우는 데 도움이 될 것으로 기대된다. 제언으로는, 인천광역시는 문화 이벤트의 다양성을 증가시키고 시민들의 참여를 유도하기 위해 지역 맞춤형 정책을 수립할 필요가 있다. 특히 Figure 1에서 인천광역시의 전통예술관람수가 저조하였는데, Figure 2의 상관분석 결과 영화관람수와 전통예술의향수 사이에 매우 높은 상관계수가 나타났다. 영화 상영 전 광고 시간에 인천 지역의 전통예술 공연 홍보 영상을 상영하거나, 영화와 관련된 전통예술 이벤트 정보를 영화 티켓 뒷면에 홍보, 영화와 전통예술을 연결하는 다양한 콘텐츠를 제작, 영화 티켓 지참시 전통 예술 티켓값 할인 등의 전략을 세운다면 인천광역시 주민들이 전통예술에 더 많은 관심을 가지고 참여할 수 있도록 유도하는 데 도움이 될 수 있다.

마지막으로, 본 연구는 인천광역시에 한정되어 수행되었으므로, 타 지역에 대한 분석을 하기 위해서는 추가적인 연구가 필요하다. 더불어 다른 지역사회에서의 문화 이벤트 참여와 만족도 및 향후 참여 의향에 관한 연구를 통해 비교 분석을 실시한다면 더욱 포괄적인 문화 정책 수립에 기여할 수 있을 것이다. 이를 통해 문화 이벤트가 지역사회의 문화적 다양성과 경제적 발전에 더욱 기여할 수 있도록 하는 방안을 모색해야 할 것이다.

#### References



본 저작물은 '문화체육관광부에서 '2018~2022'작성하여 공공누리 제4유형으로 개방한 '문화향수실태조사(작성자:이나영)'을 이용하였으며, 해당 저작물은 '문화체육관광부, <https://mcst.go.kr/kor/main.jsp>'에서 무료로 다운받으실 수 있습니다.

## 중국유학생의 한국유학과 한국대학 선택에 대한 영향요인분석

도몽이<sup>1</sup>, 양정호<sup>2</sup>

### 요 약

대학생 국제 이동성은 지난 20년 동안 증가하는 추세 있고 세계적인 현상이 되었으며 아시아 국가들은 외국유학생의 중요한 수용 국가가 되고 있다. 본 연구에서는 한국에서 계속해서 증가하는 재학 중인 중국인 유학생이 유학국가로서 한국과 한국대학을 선택한 요인을 분석하는 데 목적이 있다. 본 연구는 push-pull 요인 프레임워크와 설문분석을 사용하여 유학지 선택에 관한 연구모형을 만들었다. 이를 바탕으로 국내에 재학 중인 중국인 유학생(대학생, 대학원생)을 대상으로 설문조사를 실시하였으며, 총 300부를 본 연구에 활용하였다. 자료 분석을 위해서 SPSS를 이용하여 신뢰도 분석, 빈도분석, 요인분석, ANOVA분석을 하였다. 연구 결과는 첫째, 학년에 따라 학부생과 대학원생이 한국 유학 선택 시 개인요인, 수업요인, 추천요인, 취업요인, 사회 환경 요인에 유의미한 차이가 있었으며, 학부생은 일반적으로 대학원생보다 개인, 수업, 추천, 취업 및 사회 환경 요인의 영향을 더 많이 받았다. 둘째, 학년에 따라 학부생과 대학원생이 한국 대학을 선택 시 추천요인, 대학 명성요인 및 대학조건 요인에 유의미한 차이가 있었으며, 학부생은 대학원생보다 추천과 대학 여건상 더 많은 영향을 받지만 대학원생은 학부생보다 대학의 명성을 더 중시한다는 것을 나타냈다. 외국인 유학생 유치 정책이 지속적인 성공을 거두기 위해서, 특히 매년 증가하는 중국인 유학생을 계속해서 유치하기 위해서는 한국과 한국대학에 대한 관심을 갖게 된 유학생들의 유학지 선택이유를 바탕으로 한국과 국내 대학이 국제화 경쟁력을 높이기 위한 유학생 유치 정책 및 전략이 될 필요가 있다.

주요 용어 : 한국유학, 한국 대학 선택 이유, 중국유학생, push-pull이론, ANOVA분석.

<sup>1</sup>03063 서울특별시 종로구 성균관로 25-2, 성균관대학교 교육학과 석사과정. E-mail: taomengyi27@gmail.com

<sup>2</sup>(교신저자) 03063 서울특별시 종로구 성균관로 25-2, 성균관대학교 교육학과 교수.

E-mail: jyang@skku.edu

## 중국 공립 중학교 교사승진제도분석 -중국 산둥성 연해 지역의 청도, 동영, 빈주 중심으로

조개열<sup>1</sup>, 양정호<sup>2</sup>

### 요 약

중국의 교사승진제도 세부 평가기준이 명확하지 않아 교사의 질적격차가 나타나고 승진유인에도 영향을 미치고 있다. 본 논문은 중국 산둥성 연해 지역에서 근무하는 공립 중등교사를 인터뷰를 통해 분석해 교육공무원 승진정책이 적합성, 적정성, 효과성, 효율성, 대응성 등을 기준으로 어느 정도 제대로 운영되고 있는지 분석하려고 한다. 산둥성 연해 3개 도시인 청도, 동영, 빈주에서 근무하는 중학교 교사 총 8명의 인터뷰를 통해 승진제도를 어떻게 생각하고 있는지 분석하였다. 인터뷰를 내용 분석을 보면, 첫째, 신입교사, 경력이 많은 교사, 정년을 앞둔 교사간의 차이가 뚜렷했다. 모든 지역에서 나이 많은 교사들은 경력중심의 승진제도에서 가장 많은 비중을 차지하기 때문에 승진 욕구가 더 강하게 나타났다. 신입 교사는 저 연령에 따라 빠른 승진이 어렵고, 정년을 앞둔 교사들은 정년을 앞두고 있어 상대적으로 승진에 둔감했다. 둘째, 도시교사와 농촌교사간의 승진 환경이 다르기 때문에 승진제도에 대한 태도에 차이가 나타났다. 농촌 교사의 승진 기회는 대부분 보직교사에 집중되어 있다. 승진이 불공평한 배경에도 불구하고 오히려 농촌교사들은 승진에 더 관심이 많았다. 도시 교사들은 보직교사에 우선 승진 기회를 주기보다 교사에게 공정한 경쟁 기회를 부여하고 있어, 오히려 승진 욕구가 크지 않았다. 셋째, 명예증서와 공개수업은 교사의 승진에 영향을 미치는 핵심요소 중 하나이며 농촌교사에 있는 일반교사는 보직교사보다 적시에 정보를 습득한 쉽지 않았다. 시 단위 공개수업을 하는 교사는 주로 교장이며 일반 교사는 참여할 기회가 적었다. 넷째, 도시지역에서는 석사나 박사학위 받은 우수교사는 승진에서도 유리한 부분이 있지만, 대부분 복리후생 급여가 높더라도 농촌 지역에는 가려고 하지 않는 경향을 보인다. 교사승진제도는 지역, 연령, 보직여부에 따라 차이가 크게 나타나고 있어, 앞으로 이에 대한 개선이 필요해 보인다.

주요 용어 : 교사승진제도, 중국 중학교 교사, 도시 교사, 농촌 교사, 질적연구.

## 중국의 온라인 교육 플랫폼 비교분석: 신동방온라인, 후지앙왕샤오, 요우다오 프리미엄 코스중심으로\*

손정욱<sup>1</sup>, 양정호<sup>2</sup>

### 요 약

2020년 초에 발생한 코로나19는 전 세계 대부분의 국가에 큰 영향을 주었으며, 중국의 교육업 계도 상당한 영향을 받았다. 코로나19의 방역 및 통제 동안 각급(級) 학교는 “수업이 멈춰도 가르침은 멈추지 않고 수업이 멈춰도 배움이 멈추지 않는다”(停課不停教, 停課不停學)를 모토로 삼고, 온라인 교육으로 학습 방향을 돌리게 되었다. 이에 그 기간 동안 중국의 온라인 교육 플랫폼은 중요한 역할을 담당하게 되었다. 중국 온라인 교육 시장 규모는 점차 커지고 있으며 교육 대상에 따라 취학 전 교육, K12 교육, 고등 교육, 직업 훈련 등 범주로 나눌 수 있다. 본 연구에서는 중국 주요 온라인 교육 플랫폼의 현황과 특성을 파악하고 이들의 차이점과 강점을 비교하였다. 현재 중국의 가장 대표적인 온라인 언어 교육 플랫폼인 신동방온라인, 후지앙왕샤오, 요우다오 프리미엄 코스를 중심으로, 온라인 교육 플랫폼 구축 현황, 커리큘럼 내용, 기술 지원 및 사용자 경험, 플랫폼의 강점 차원에서 분석하였다. 우선, 신동방온라인은 광범위한 커리큘럼과 강력한 교사 역량을 제공하는 종합 플랫폼이며, 후지앙왕샤오는 언어 학습에 중점을 둔 온라인 수업을 제공하고, 요우다오 프리미엄 코스는 AI 기술을 활용한 맞춤형 K12 교육을 제공하는 플랫폼이다. 연구 결과, 신동방온라인 및 요우다오 프리미엄 코스는 상대적으로 최고의 강사진과 높은 품질의 커리큘럼을 보유하고 있으며, 후지앙왕샤오의 강사진은 비교적 광범위하지만 커리큘럼의 품질은 보통정도로 나타났다. 기술 지원 및 사용자 경험 측면에서 신동방온라인, 요우다오 프리미엄 코스, 후지앙왕샤오는 모두 안정적이고 효율적인 기술 플랫폼을 갖추고 있어 원활한 온라인 학습 경험을 제공하고 있었다. 학습 커뮤니티 및 서비스 지원 측면에서 세 플랫폼은 모두 활발한 학습 커뮤니티를 구축했으며 교육생은 강사 및 기타 교육생과 상호 작용하고 학습 경험을 공유하며 학습에서 직면한 문제를 해결할 수 있었다. 그리고 3개의 플랫폼은 방과 후 교육, 과제 수정, 학습 계획 수립 등을 포함한 포괄적인 학습 지원 및 서비스를 제공하여 교육생이 더 잘 배우고 개선할 수 있도록 지원하고 있었다. 연구 결과는 온라인 교육 플랫폼의 운영자가 과정을 설계하고, 사용자 경험을 최적화함과 동시에 시장 경쟁력을 향상시키는 데 도움이 될 것이다.

주요 용어 : 중국 온라인 교육 플랫폼, 신동방온라인, 후지앙왕샤오, 요우다오 프리미엄 코스, 비교분석.

<sup>1</sup>03063 서울특별시 종로구 성균관로 25-2, 성균관대학교 교육학과 석사과정. E-mail: sunjingyu1999@naver.com  
<sup>2</sup>(교신저자) 03063 서울특별시 종로구 성균관로 25-2, 성균관대학교 교육학과 교수. E-mail: jyang@skku.edu

## 중국 ‘쌍일류’ 대학의 세계 대학 순위 변화와 영향요인 분석

단시기<sup>1</sup>, 양정호<sup>2</sup>

### 요약

중국의 ‘쌍일류’ 명문대학 정책이 지속적으로 추진됨에 따라 고등교육 발전의 질이 점차 중국 정부의 관심 초점이 되었다. 중요한 정량적 평가 도구로서 대학 순위는 대학의 발전 수준을 어느 정도 측정할 수 있다. 따라서 본 연구는 현재 유행하고 전 세계적으로 큰 영향을 미치는 QS(Quacquarelli Symonds), THE(Times Higher Education), ARWU(Academic Ranking of World Universities) 3대 세계 대학 순위를 매개로 하여 2020년부터 2023년까지 중국의 30개 ‘쌍일류’ 대학이 QS, THE, ARWU 순위에서의 순위 변동과 구체적인 평가지표 점수의 변동 상황을 연구하였다. 특히 중국의 ‘쌍일류’ 대학 순위 변동에 영향을 미치는 요인을 분석하고 3개 세계 순위 체계에서 순위 변동지수의 차이가 큰 학교를 분석하였다. 연구 결과에 따르면, 첫째, 중국의 30개 ‘쌍일류’ 대학이 3대 세계 순위 체계에서 상승세를 보이고 있었다. 둘째, 30개 ‘쌍일류’ 대학은 THE 및 ARWU 세계 대학 순위에서 조금 빠르게 상승하였지만 QS 순위에서는 상대적으로 약간 느리게 상승하는 것으로 나타났다. 동시에 순위 체계에 따라 순위 상승 및 하락에 영향을 미치는 요인이 다르다. QS 순위에서 상승한 23개 대학 대부분은 교원당 논문 피인용수 지표 점수가 향상된 영향이다. 교수진 학생 비율, 외국인 학생/교원 비율, 학계 및 고용주 평판 지표 점수가 하락하면서 나머지 7개 대학의 순위가 하락했다. THE 순위에서 순위가 상승한 23개 대학에는 논문 인용, 연구, 교육, 산업소득 4개 지표가 큰 역할을 했고, 교육 및 국제화 지표 점수 하락으로 2개 대학이 순위가 하락했다. ARWU 순위에서 28개 대학 모두 상승세를 보였으며, 고인용(Highly Cited) 과학자, N&S 논문, 1인당 학술평가 지표가 순위 상승에 영향을 미친 주요 요인이었다. 셋째, 세계 대학 순위에 따라 대학 발전의 영역에 대한 선호도가 다르기 때문에 평가 체계에 큰 차이가 있다. 이러한 차이로 인해 동일한 대학의 순위는 각 3개 대학 평가 순위 체계에 따라 차이가 있다. 연구에 따르면, 3개 순위에서 순위 변동지수 차이가 큰 학교는 일반적으로 과학 연구 지표에서 높은 점수로 나타났으나 글로벌 평판, 국제화 수준 지표에서는 부족하여 QS 순위가 일반적으로 하락한 반면 ARWU 순위는 일반적으로 상승하는 모습을 보여준다. 마지막으로 3대 대학 평가의 분석을 기반으로 중국 대학의 경쟁력 확보를 위한 제언도 제시하였다.

주요 용어 : 세계 대학 순위, ‘쌍일류(雙一流)’정책, 중국 대학, 순위 비교, QS, THE, ARWU.

## 중국인 유학생에 대한 한국학생의 인식 및 태도에 관한 연구

비시월<sup>1</sup>, 양정호<sup>2</sup>

### 요 약

중국인 유학생의 숫자가 늘면서 중국인 유학생에 대한 한국학생들의 태도도 주목을 받게 되었는데, 소셜미디어에 중국인 유학생에 대한 한국인 학생들의 부정적인 인식과 평가가 종종 올라와 사회적 관심을 끌기도 하였다. 일반적인 예상과 달리 한국학생들이 외국인과의 접촉 빈도가 많을수록 중국인 유학생에 대한 긍정적인 인식과 태도가 오히려 줄어드는 경향이 있다는 연구도 있는데, 이것은 단순히 접촉 빈도뿐 아니라 만남의 질도 중요하다는 것을 보여준다. 이러한 관점에서 본 연구에서는 ‘만남의 질에 영향을 미치는 요인은 무엇인가?’, ‘만남의 질을 긍정적으로 변화할 수 있는 방법은 무엇인가?’라는 연구 문제에 초점을 맞춰, 문헌 분석과 인터뷰를 통해 중국과 한국 학생 간 만남의 질이 중국인 유학생에 대한 한국인 학생의 인식과 태도에 영향을 미치는지, 만남의 질에 긍정적 변화로 영향을 미치는 요인은 무엇이며, 만남의 질을 향상시키기 위해서는 어떠한 지원과 개선안이 있는지 논의하였다. 이 연구는 한국 대학생 3명과 대학원생 4명 총 7명을 대상으로 인터뷰를 하였다. 연구 결과, 처음에는 한국인 학생이 중국인 유학생에 대해 부정적인 편견이 많이 있었지만, 만남의 질이 긍정적일수록 한국인 학생은 이러한 편견을 줄이고 중국인 유학생에 대한 부정적인 인식을 바꾸게 되었다. 그러나 만남의 질이 긍정적 영향 없이 단순히 만나는 횟수가 많아지는 것만으로는 한국인 학생이 중국인 유학생에 대한 부정적인 인식과 태도를 개선하는 데는 큰 영향을 미치지 않는 것으로 나타났다. 인터뷰를 통해 한국인 학생은 중국인 학생과의 만남에 긍정적 영향을 미치는 요인으로는 주로 중국인 유학생의 한국어 능력, 양국 학생 서로의 교류 태도(예컨대 수업 및 토론 참여도, 서로 이해, 존중, 수용 등)라고 생각하는 것을 확인할 수 있었다. 따라서 본 연구에서는 양국 학생 만남의 질을 높이고 중국인 유학생에 대한 긍정적인 인식과 태도를 촉진하는 지원 방안으로는 3가지를 제시하였다. 첫째, 중국인 유학생의 한국어 능력 향상 지원이 필요하다. 대학이나 어학원에 개설된 어학 수업은 유학생들의 글쓰기나 읽기 한국어 능력뿐만 아니라 1:1 레슨 및 도우미 활동을 통해 말하기 능력 향상에 집중할 필요가 있다. 둘째, 한국과 중국 학생들 간의 상호이해를 위해 다문화화를 위한 관련 과정을 개설할 필요가 있다. 셋째, 적극적인 수업 참여는 양국 학생들의 만남과 교류를 높이는 방법인데 수업 참여도를 높이기 위해 학생 스스로 수업에 적극적으로 참여하는 것 외에 다문화 역량을 갖춘 교수의 양성이 필요하고, 학교는 수업 참여도 비중이 높은 성적관리 시스템을 도입할 필요가 있다.

주요 용어 : 중국인 유학생, 만남의 질, 한국 유학, 질적 연구.

<sup>1</sup>서울특별시 종로구 성균관로 25-2, 성균관대학교 교육학과 석사과정. E-mail: bxy954546450@163.com

<sup>2</sup>(교신저자) 서울특별시 종로구 성균관로 25-2, 성균관대학교 교육학과 교수. E-mail: jyang@skku.edu

## 중국 대학생의 페미니즘 의식에 미치는 영향에 대한 연구

유예첸<sup>1</sup>, 양정호<sup>2</sup>

### 요약

본 논문은 중국의 성별교육센터에서 제공된 데이터를 바탕으로 중국대학의 페미니즘 현황을 분석하였으며, 연구 타당성을 위해 중국 대학생 50명을 대상으로 설문조사를 하였고, 대학 수준, 전공 그리고 학력별로 학생 4명 그리고 교수 2명 등 총 6명 인터뷰도 실시하였다. 데이터 분석 및 인터뷰자료를 통해 고등교육을 받은 중국 학생과 교수 대상으로 대학내 성차별대우와 성희롱에 대한 태도와 인식을 기초로 중국 대학생의 페미니즘 의식에 미치는 영향, 즉 대학교육을 통해 페미니즘 인식 변화에 영향을 줄 수 있는지, 어떤 방식으로 인식 변화에 영향을 미치는지 그리고 대학 수준에 따라 페미니즘 인식 변화 정도에 영향을 미치는지를 살펴보았다. 연구 결과는 다음과 같다. 1) 성별 교육센터에서 제공된 데이터에 따르면 69.3% 중국 대학생은 성희롱 및 성별 차별대우를 경험한 적이 있는 것으로 나타났다. 성희롱 및 성별 차별대우는 중국 대학에서도 보편적인 현상이라고 할 수 있다. 2) 중국 대학생 50명을 대상으로 한 본 설문조사에서도 83.5%가 페미니즘을 알게 되었다고 응답했다. 3) 또 인터뷰를 한 4명의 대학생과 2명의 교수도 대학에서 페미니즘을 구체적으로 알 수 있었다고 밝히고 있었다. 이들은 대학 수업보다도 TV 등 특정 프로그램 및 강연을 통해 페미니즘에 점차 알게 되는 과정을 겪었다고 하였다. 특히 인터뷰에 응한 2명 교수도 현재 중국 대학생들은 페미니즘에 대해 제대로 알기 위해 대학교육의 변화가 필요하다고 강조하였다. 4) 대학 수준에 따라 좋은 대학일수록 학생들은 사회적 남녀불평등에 대한 인식이 더 강한 부분이 있었으며, 대학 수준이 높은 학교일수록 여성권익보호단체가 더 많이 생기는 추세도 나타났다. 데이터 및 인터뷰자료를 분석해 보면, 대학은 대학생들에게 페미니즘 인식의 변화를 유도할 수 있는 다양한 여건을 제공한다는 것을 알 수 있다. 대학은 공간으로서 학생들에게 페미니즘 이해와 단체 활동의 기회를 일부 제공하지만, 현재 중국 대학은 강의를 통해 페미니즘이나 성별 평등 인식 변화를 유도하는 역할을 충분하게 하지 못하고 있는 것으로 나타났다. 페미니즘이 중국에서 비교적 늦게 알려지고 있고 사회 분위기도 보수적이고 폐쇄적인 부분이 있어 대학교육이 페미니즘 인식 변화에 충분하게 영향을 미치기 위해서는 대학 강의 개혁 등에 더 많은 노력을 기울일 필요가 있다.

주요 용어 : 중국대학, 페미니즘, 대학교육, 성별평등, 복합연구.

## 기술수용모델 적용을 위한 새로운 패러다임

고서정<sup>1</sup>

### 요약

기술수용모델(TAM)은 정보통신기술의 채택 행동을 설명하기 위해 설계된 이후 많은 연구에서 활용되어 높은 설명력을 입증하였고 사회의 새로운 양상으로 대두되었다. TAM은 모델의 도입, 검증, 확장, 정교화를 거치며 발전했고, 관련 변수의 확장과 다른 이론들과의 통합을 통해 그 이론적 깊이와 범위를 넓혀왔다. 주요 변수로는 시스템 품질, 사회적 압력, 자기 효능감 등이 있으며, 연구 대상은 PC, 스마트폰, 인터넷 뱅킹 등으로 다양하다. 초기에는 비자발적 상황에서 조직의 정보통신기술 채택 요인을 찾는 것이 주요 관심사였다. 그러나 웹 서비스와 스마트폰의 보급으로 자발적 채택이 점점 중요해졌고, 이는 상업적, 엔터테인먼트, 자기 개발 등 다양한 목적으로 수용되었다. 디지털 기기의 확산과 AI 서비스는 비자발적 기술 채택 문제를 야기할 수 있으며, 고령층과 사회적 약자의 디지털 소외 문제 또한 고려해야 한다. 정보통신기술의 빠른 확산과 발전의 특성상 개인의 의사결정 문제와 기대 가치에 대한 포괄적이고도 종합적인 접근이 중요하다.

주요 용어 : 기술수용모델, TAM, 정보통신기술, 기술 채택 문제.

### 1. 서론

Davis(1989)가 제안한 기술수용모델(1989)은 정보통신기술 수용 관련 수용자의 수용 행동을 설명하기 위해 고안되고 제안된 이래 많은 후속 연구들이 수행되어 왔으며, 상당한 설명력을 보여왔다. 기술수용모델은 지난 30여년 동안 모델 도입, 모델 확인, 모델 확장, 모델 정교화 등으로 발전 과정을 설명하기도 하고, 독립변수 등 관련 변수의 확장을 도모하거나, 기술 준비도 등 다른 이론들과의 통합을 통한 이론의 폭과 깊이를 넓히려는 시도가 꾸준히 있어 왔다.

변수확장 관련하여 독립변수로는 시스템적 특성 요인으로 시스템 품질, 적합성, 상대적 이점, 결과설명력, 접근성을 채용하며, 사회적 특성요인으로 사회적 압력, 타인 이용도를 활용하고, 개인적 특성 요인으로 이용 경험, 자기효능감, 개인 혁신성, 교육훈련, 지식, 감정 등을 적용하고 있다(유재현·박철, 2010). 또한, 매개변수로서 신뢰, 유의성, 위험, 몰입, 가치 등을 도입하고 있으며, 조절 변수로 문화, 인구통계, 개인 특성 등을 활용하고 있다(유재현·박철, 2010).

<sup>1</sup>10540 경기도 고양시 덕양구 항공대학로 76, 한국항공대학교 경영대학 경영학과 박사과정.  
E-mail: angelsgo74@gmail.com

## 2. 본론

기술수용모델의 연구 대상을 살펴보면, 제품 분야에서 PC, 스프레드 시트, PDA, MP3, 스마트폰을 대상으로 하고, 서비스 분야로서는 웹서비스, 전자메일, 인터넷쇼핑, 이러닝, 인터넷뱅킹, 온라인 게임, 커뮤니티서비스, DMB 서비스, IPTV, 모바일 커머스, 모바일뱅킹, 모바일 인터넷, 모바일게임 등을 다루고 있다(유재현·박철, 2010).

기술수용모델의 초기 관심사는 조직에서 정보통신기술을 수용하는데 있어 영향을 미치는 요인들을 찾아보고자 하는 것이었으며, 주로 비자발적 상황을 전제로 하고 있었다. 즉, 조직내에서 팀워크와 효율성을 높이고 생산성 제고를 위해 정보통신기술을 도입 및 보급하고자 하는데 조직구성원의 수용과 활용을 위하여 필요한 요인이 무엇인지 파악하고자 하는 것이 주된 동기였다. 따라서 기술수용모델의 도입, 확장기에는 비자발적 수용을 염두에 둔 연구들이 대세를 이루고 있었다.

그러나 웹서비스가 대중화되고 전자상거래가 보편화되면서 기술수용모델은 자발적 수용을 넘어 적극적 수용을 염두에 두어야 하는 시대적 전환을 맞게 된다. 더구나 스마트폰의 보급률이 거의 전국민적으로 이루어지면서 모바일시대가 도래하게 된다.

이러한 정보통신기기의 발달과 데이터 전송 기술의 획기적 발전은 모바일커머스 등 상업적인 측면 뿐 아니라 DMB서비스 등 엔터테인먼트 분야에 커다란 영향을 미치게 되며, 연락도구 및 정보전달 도구로서 커뮤니티서비스 역할도 수행하게 된다. 인터넷을 통하여 자신의 학습활동을 보강할 수 있음은 물론 메타버스 등은 새로운 경험이나 실험 등을 가능하게 하고 있다. SNS의 발달과 보급은 정보통신기기의 새로운 역할의 장을 열고 있으니 단순한 정보교환이나 메시지전달에서 나아가 자신의 이미지전달, 자신의 외적세계를 구축하는 도구로서의 기능을 수행하고 있는 것이다.

이러한 급격한 정보통신기기의 발달을 구분하여 보자면 다음과 같이 나누어 볼 수 있을 것이다.

비자발적 수용 -> 자발적 수용(상거래, 스마트폰 사용) -> 엔터테인먼트 목적 수용 -> 자아개발을 위한 수용 -> 자신의 이미지 구축을 위한 수용 -> 자아실현을 위한 수용

이러한 정보통신기기의 자발적 수용 나아가 적극적 수용 관련 간과하지 말아야 할 부분이 사회적으로 대두되고 있다. 그것은 생산성 향상, 인건비 절약, 고용관계에서 초래될 수 있는 문제점 등을 줄이기 위한 방안의 하나로 급속히 도입되고 있는 키오스크 등 디지털기기의 보급, AI 등에 의한 챗봇 서비스 등과 관련한 비자발적 기술수용 문제인 것이다.

특히 주로 고령층이나 사회적 약자층 등이 겪을 수밖에 없는 디지털 문맹에 기한 어려움에 대한 고려가 필요하다. 정보통신기술은 하나의 궤적으로 진입하여 전개하게 되면 그 방향으로 더욱 급속도로 발전하는 속성이 있게 되므로 현재의 상태에서 어려움이나 불편함을 느낀다면 그 어려움이나 불편함은 점점 더 커질 수 밖에 없는 상황으로 내몰릴 수 밖에 없다는 점에서 현재 시점에서 이 문제를 반드시 짚고 넘어 가야할 것이다.

## 3. 결론

정보통신기술의 빠른 발전에 따른 디지털 소외 문제를 해결하기 위해서는 보다 포괄적인 접근이 필요하다. 고령층과 사회적 약자를 대상으로 디지털 교육 프로그램을 강화하고, 접근성을 높이는 정책적 지원이 매우 중요하다. 기술 수용에 대한 저항을 줄이기 위해 사용자 친화적인 인터페이스

와 지원 시스템을 도입해야 한다. AI와 디지털 기기의 확산으로 발생하는 비자발적 수용 문제를 해결하기 위해 법적 규제와 보호 장치도 마련해야 한다. 미래에 더욱 빠르게 확산할 정보통신 기술의 지속적인 발전을 통해 기술수용모델에 대한 연구 또한 꾸준히 개선하고 발전시켜야 할 것이다.

## References

- Jang, Y. H., Chun, S. Y., Lee, S., K. (2011). A Study on Consumer Acceptance of New Financial Products, *Journal of the Korean Data Analysis Society*, 13(2), 979-988. (in Korean).
- Kim, B. (2019). Understanding key antecedents of SNS addiction: focused on flow and trust in members, *Journal of the Korean Data Analysis Society*, 21(6), 3155-3168. (in Korean).
- Lee, C. (2017). An analysis of determinants of purchase of accident coverage using structural equation model, *Journal of the Korean Data Analysis Society*, 19(2), 827-837. (in Korean).
- Lee, Y. J., Jung, H. H. (2023) Keyword Analysis of Twitter Data on New Digital Technology through Topic Modeling and ERGM, 26(6), 2093-2107, (in Korean). <https://doi.org/10.37727/jkdas.2023.25.6.2093>
- Moon, H. Y., Kim, Y. S. (2023). Multiple mediating effects of SSTQUAL and satisfaction in the relationship between technology anxiety and behaviour intention toward SST, *Journal of the Korean Data Analysis Society*, 25(6), 2325-2339. (in Korean). <https://doi.org/10.37727/jkdas.2023.25.6.2325>
- Park, K. N. (2005). A Study on the Application of TAM for the Promotion of User s Intention to Utilize the CSS, *Journal of the Korean Data Analysis Society*, 7(2), 549-564. (in Korean).
- You, J. H., Park, C. (2010). A Comprehensive Review of Technology Acceptance Model Researches, *Entrue Journal of Information Technology*, 9(2), 31-50. (in Korean).

## 트렌드리서치는?

국내&해외 마케팅 및 사회/여론 조사 경력이 20년 넘는 전문가가 직접 연구에 참여해 통찰력 있는 INSIGHT 제공  
트렌드리서치의 약 72만 ACCESS PANEL을 통한 정확한 조사 수행이 가능함



마케팅 전문가와  
사회조사 전문가의 결합



자사 패널을 직접 보유  
(720,000명)



별도의 정성조사 전문가 활용

## 주요 서비스



온라인/모바일 조사

전문인력을 통해 국내 온라인 및 모바일 조사를 진행하고 있으며 Quick survey, Mobile Diary 등의 조사 수행



마케팅/소비자 조사

마케팅 의사 결정을 위한 시장 및 소비자 정보 제공



학술/연구/석박사 논문 조사

학술/연구/석박사 논문의 경우 사회기여 차원으로 일반 조사와는 차별화된 단계(다운된 단계)로 제공

## T-Panel 현황

### 해외조사

해외 제휴 네트워크를 활용하여 신속한 조사 진행이 가능합니다.

### 특정 포털 사이트 회원이 아닙니다

조사참여 횟수, 조사 참여일, 참여조사 종류 및 응답의 성실도를 체계적으로 관리합니다.

### 대표성이 있습니다

샘플링을 통해 추출된 표본들에게만 이메일, 문자, 알림톡을 발송하여 조사를 진행합니다.

### 분석 서비스

조사 결과의 통계 검정 결과를 제공하여 분석의 신뢰성을 제고합니다.

Available  
Panels  
720,000



## A faster, more productive AI and analytics platform

What if you could make decisions more decisively?  
And your team could work more productively?

You can with SAS® Viya®. SAS is committed to creating technology that is not only collaborative, powerful and intuitive, but also ethical, equitable and sustainable. Together, we can build a better, more productive future for all.

[sas.com/viya](https://sas.com/viya)



### SAS®Ondemand for Academics

- SAS 클라우드 기반 소프트웨어를 무료로 체험해보십시오.
- 먼저 SAS Profile를 생성하고 Sign Up 해주십시오.

## 데이터솔루션 | 前 SPSS Korea

데이터솔루션은 Digital Transformation Journey를 함께 하는 Cloud와 Data 기반의 Digital Innovation Provider입니다. 데이터솔루션은 Data와 Application Lifecycle 전 과정에 최적화된 Total Service와 솔루션을 보유하고 있으며, 특히 Cloud, IT Modernization, Analytics, Intelligence Service를 통해 고객에게 새로운 Insight와 차별화된 Value를 제공하고, 고객의 비즈니스 성공을 위한 Digital Innovation을 함께 합니다.



### 주요 분석 사업 영역

SPSS Korea를 전신으로 한 Data & Analytics 부문에서는 통계소프트웨어/ML/AI 솔루션 공급 및 데이터분석 컨설팅 사업을 수행하며, 고객의 데이터 기반 의사결정을 지원하고 있습니다. 30여 년간 제조, 금융, 유통, 통신 등 다양한 산업 군에서 경험 및 지속적인 연구개발을 통해 통계분석에서 ML/AI로 진화하는 시장의 변화를 이끌고 있습니다.

Analytics Consulting	MAaaS (Managed Analytics as a Service)	Intelligence Solution	AI/BigData Education
<ul style="list-style-type: none"> <li>· 수요 예측</li> <li>· 리스크 예측</li> <li>· 상품 추천</li> <li>· 품질 분석 및 예지 보전</li> <li>· 이미지 분석</li> <li>· 텍스트 분석</li> <li>· 고객 분석</li> </ul>	<ul style="list-style-type: none"> <li>· MAaaS PASP</li> <li>· PoC</li> <li>· Pro</li> <li>· Maintenance</li> <li>· MAaaS by AI Model</li> </ul>	<ul style="list-style-type: none"> <li>· KoreaPlus Statistics (Embaded on SPSS)</li> <li>· KNIME</li> <li>· IBM CP4D</li> <li>· Market Mind</li> <li>· Brightics AI</li> <li>· AI.NER</li> </ul>	<ul style="list-style-type: none"> <li>· 빅데이터리닝센터</li> <li>· SPSS/Python/R 교육 연간 80여개 온/오프라인 강의제공</li> </ul>

### 데이터솔루션 분석 사례

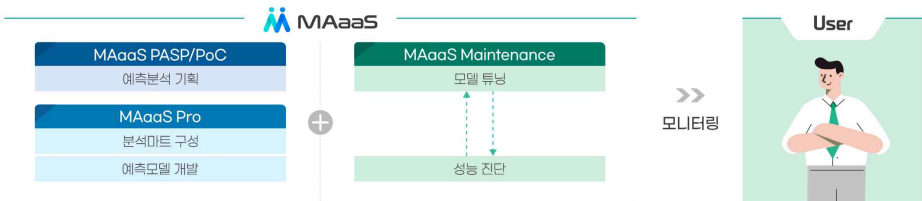
<b>수요 예측</b> <b>제조</b>   수요예측 분석 시스템 구축 <b>유통</b>   판매량 예측 모델 개발 <b>물류</b>   물동량 예측 시스템 구축 <b>공공</b>   열 수요 예측 모델 개발	<b>텍스트 분석</b> <b>통신</b>   VoC 분석 <b>금융</b>   텍스트 분석을 활용한 평판 분석 <b>유통</b>   텍스트 분석을 통한 마케팅 <b>증권</b>   챗봇 메시지 분석
<b>상품 추천</b> <b>유통</b>   AI 상품 추천 프로그램 구축 <b>유통</b>   온라인 쇼핑을 개인화 상품 추천 모델 개발 <b>유통</b>   개인화 추천 서비스 모델 구축 <b>금융</b>   고객관리 시스템을 통한 상품 추천	<b>품질 분석 및 예지보전</b> <b>제조</b>   취약 부위 예측 모델 개발 <b>제조</b>   특정 공정이상 예측 분석 <b>제조</b>   공정 최적화 및 설비 고장 모델 개발 <b>화학</b>   수율 예측 시뮬레이터 개발
<b>이미지 분석</b> <b>제조</b>   이미지 데이터 기반의 양불 판정 <b>제조</b>   비전 검사를 통한 양불 판정	

### AI 예측서비스, 이제 구독하세요!

No 플랫폼   No 데이터수입   No 데이터/가이언트스트   No 유지관리

#### MAaaS | Managed Analytics as a Service

MAaaS는 AI 예측모델 기획부터 유지관리까지의 Data Life-cycle에서 필요한 컨설팅을 제공하는 데이터솔루션의 AI 예측분석 토탈 서비스입니다. 이제는 분석 예측 결과를 구독으로 제공받으세요.



## Company : 회사소개 및 연혁

2003년 이래, 데이터 가공/분석 및 유통의 신시장 개척에 주력하였습니다.



<p><b>회사명</b> ㈜지디에스컨설팅그룹</p> <p><b>설립일</b> 2015년 2월</p> <p><b>대표이사</b> 김은석</p> <p><b>업종</b> 데이터베이스, 소프트웨어 자문, 개발 및 공급, 전자상거래업</p> <p><b>주소</b> 서울 성동구 광나루로 228, 9층 (성수동2가, 렉스모터스)</p> <p><b>연락처</b> Tel: 02-2135-8895(대표) Fax: 02-2135-8896</p>	<p><b>2018.07</b> 고려대 '데이터 사이언스 센터' 업무 제휴 (교육 및 인턴십)</p> <p><b>2017.07</b> KB국민카드 '빅데이터 센터' 업무 제휴 (카드 데이터 상품개발)</p> <p><b>2016.09</b> KT '빅데이터 지원단' 업무 제휴 (빅데이터 가공, 중계, 서비스 분야)</p> <p><b>2016.08</b> 데이터 거래 중개 사업 개시 (미래창조과학부 ICT 기금사업) 부실연구소 '한국공공정보개발원' 설립 (산기협 2016113799호)</p> <p><b>2016.07</b> 빅데이터 분석 사업 개시 (지속가능한 빅데이터 사업, 경기도)</p> <p><b>2015.04</b> 국가 DB 구축 사업 개시 (유동인구 실사 DB 구축, NIA)</p> <p><b>2015.02</b> ㈜지디에스컨설팅그룹 설립 (㈜지디에스케이 內 컨설팅 사업본부의 법인 분할)</p> <p><b>2004.10</b> 벤처기업 등록 (041134221-1-01033호)</p> <p><b>2003.02</b> ㈜지디에스케이 설립 (대표이사 김은석 차)</p>
---	---

## Business : 사업 영역

데이터 활용 및 유통의 일관된 서비스 제공!!

사업영역	빅데이터 분석 컨설팅	데이터 가공 <small>자세히 보기 &gt;</small>	빅데이터 플랫폼 구축 <small>자세히 보기 &gt;</small>	데이터 판매/유통 <small>자세히 보기 &gt;</small>
상품 및 서비스	<ul style="list-style-type: none"> <li>· 데이터 관련 ISP 및 계획수립</li> <li>· 주요 빅데이터 (통신, 카드) 공급 및 활용</li> <li>· 공공 데이터 활용 서비스</li> <li>· 데이터 분석 및 시각화 서비스</li> </ul>	<ul style="list-style-type: none"> <li>· 공공기관 정보의 정제, 가공 및 제공</li> <li>· 통신, 카드 등 빅데이터 정제, 결합 및 제공</li> <li>· 비정형 데이터 가공 및 제공</li> </ul>	<ul style="list-style-type: none"> <li>· 통계분석 도구 (알트릭스) 판매</li> <li>· 데이터 용복합 및 통합 시스템 구축</li> <li>· 빅데이터 분석 시스템 구축</li> </ul>	<ul style="list-style-type: none"> <li>· 데이터 사용영역별 데이터 판매</li> <li>· 업무 영역별 맞춤형 데이터 구축 및 제공</li> <li>· 주제별 맞춤형 데이터 제공</li> </ul>
사업사례	<ul style="list-style-type: none"> <li>· 국방기술품질원, 빅데이터 시범사업 (17-)</li> <li>· 경기도, 지속가능한 빅데이터 분석 (16)</li> <li>· 서울시, GDP 분석 모델링 구축 (16)</li> <li>· 건강보험공단, 빅데이터노안질원 모델 구축 (14)</li> <li>· 서울시, 유동인구 DB 구축 및 모델링 (15)</li> </ul>	<ul style="list-style-type: none"> <li>· 삼성전자, 법인영업 데이터 제공 (15-)</li> <li>· 한샘, 가구판매 고객 데이터 제공 (15-)</li> <li>· 코웨이, 고객정보 관리 데이터 제공 (06-)</li> <li>· KT, 지사 관리 데이터 제공 (10-)</li> <li>· 기업은행, 영업점 관리 데이터 제공 (10-)</li> </ul>	<ul style="list-style-type: none"> <li>· 창원시, 빅데이터 분석시스템 구축 (17)</li> <li>· KT, SI 조류독감 시각화 시스템 구축 (16)</li> <li>· 경상북도, 통계정보시스템 구축 (15)</li> <li>· KT, 지사 관리 데이터 제공 (10-현재)</li> <li>· 서울시, 도시통계 시스템 구축 (14, 15)</li> </ul>	<ul style="list-style-type: none"> <li>· 한국정보화진흥원, 데이터 거래 중개 플랫폼 구축 (16)</li> <li>· 한국정보화진흥원, 소상공인 데이터 제공 플랫폼 (17)</li> <li>· 경상북도, 사학스인구축을위한 데이터 (16)</li> <li>· 코웨이, 고객정보관리를위한 데이터 (16)외다수</li> </ul>