



한국자료분석학회

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

문서번호 : KDAS-2403000

시행일자 : 2024. 12. 5.

(경유)

수 신 : 전 회원

참 조 :

선결			지	
	일자		시	
접 수	시간		결 재	
	번호			
담당자				

제목: 한국자료분석학회 2024년 동계 학술논문발표대회

- 귀 회원님의 무궁한 발전을 기원하며, 본 학회의 발전을 위한 회원 여러분의 관심과 노고에 깊이 감사를 드립니다.
- 본 학회 2024년도 동계 학술논문발표대회를 다음과 같이 개최함을 알려 드리오니 많은 참여 바랍니다.
 - 개최일자 : 2024년 12월 5일(목) ~ 6일(금)
 - 장 소 : 충남대학교(대전광역시 유성구 대학로 99)

한국자료분석학회장 한상태



PROCEEDINGS OF THE KOREAN
DATA ANALYSIS
SOCIETY

December 5-6, 2024

at

Department of Information Statistics,
Chungnam National University, Daejeon, Korea.

(<https://www.plus.cnu.ac.k>)

THE KOREAN DATA ANALYSIS SOCIETY Founded 1998

PROCEEDINGS OF THE KOREAN DATA ANALYSIS SOCIETY DECEMBER 6, 2024



한국자료분석학회 2024년도 동계 학술논문발표대회 준비위원회

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 중 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
	제 상 영	고려대학교 경제통계학과
	조 장 식	경성대학교 정보통계학과
	이 상 인	충남대학교 정보통계학과
위원	김 지 희	강원대학교 응급구조학과
	김 태 훈	경성대학교 경제금융물류학부
	김 성 환	건국대학교 응용통계학과
	김 양 진	숙명여자대학교 통계학과
	곽 일 엽	중앙대학교 통계학과
	박 만 식	성신여자대학교 수리통계데이터사이언스학부
	김 은 석	(주) GDS 컨설팅그룹 대표이사
	김 지 연	(주) 케이스탯리서치 대표이사
	이 은 지	충남대학교 정보통계학과

한국자료분석학회

2024년도 동계 학술논문발표대회

- 일시 : 2024년 12월 5일(목) ~ 12월 6일(금)
- 장소 : 충남대학교
- 공동주최 : 한국자료분석학회,
충남대학교(정보통계학과, 데이터사이언스 융합인재양성사업단)
- 주관 : 충남대학교 정보통계학과
- 후원 : 한국연구재단, (주)케이스탯리서치, 트랜드리서치, (주)GDS 컨설팅, 데이터솔루션,
나이스평가정보, JMP Korea, 한국건강증진개발원, 충남대학교

● 12월 5일 목요일

17:00 - 19:00 편집위원회

17:00 - 19:00 이사회

● 12월 6일 금요일

09:00 - 18:00 리셉션(융합교육혁신센터 컨벤션홀 1층)

포스터발표(9:00 - 16:30, 융합교육혁신센터 101호)

09:30 - 10:50 논문발표 A1 - A4(융합교육혁신센터 201, 202, 301, 303호)

11:10 - 12:10 튜토리얼(산학연교육연구관 DSC홀 315호)

좌장: 이성건 교수(충무이사, 성신여자대학교)

제목 : 분류문제의 이해와 활용

강연자 : 신승준 교수(고려대학교 통계학과)

12:10 - 12:50 총회(산학연교육연구관 DSC홀 315호)

진행: 이성건 교수(충무이사, 성신여자대학교)

12:50 - 14:00 중식시간 (인재개발원 제 2후생관:식당)

14:00 - 15:00 학회장 초청 강연(산학연교육연구관 DSC홀 315호)

좌장: 한상태 교수(학회장, 호서대학교)

제목 : AI시대 살아남기

강연자 : 원유집 교수(KAIST, 한국정보과학회 회장)

15:10 - 16:30 논문발표 B1 - B4(융합교육혁신센터 201, 202, 301, 303호)

16:40 - 18:00 논문발표 C1 - C4(융합교육혁신센터 201, 202, 301, 303호)

논문발표 A: 09:30 - 10:50

논문발표 A1(201호)

통계 (응용)

좌장 : 진서훈(고려대학교)

-
- 09:30 | 1. Autoencoder-Driven Feature Extraction in Raman Spectroscopy Analysis for
10:50 | Hydrocarbon Mixtures 1
양승지*(한양대학교 응용통계학과 및 기초과학융합연구소, 박사후 연구원),
송우석, 정희일(한양대학교 화학과 및 기초과학융합연구소)
2. 출원 특허 기반의 복합어 키워드를 활용한
디지털 헬스케어 분야 이머징 기술 트렌드 분석 3
이서엘*, 김민기(한국과학기술기획평가원), 정호현 (성신여자대학교
수리통계데이터사이언스학부), 홍미영 (한국과학기술기획평가원)
3. 투표의 역설적 측면에 관한 통계 교육용 고찰 5
김혁주*(원광대학교 빅데이터 · 금융통계학부)

논문발표 A2(301호)

통계 (방법론)

좌장 : 이은지(충남대학교)

-
- 09:30 | 1. Sparse Sufficient Dimension Reduction for Directional Regression 11
10:50 | 권가윤*(이화여자대학교 통계학과), 노기정(OCI Company R&D center),
김경원(이화여자대학교 통계학과)
2. Angle-based Deep Hierarchical Classification via Label Embedding 13
김혜연*, 신승준(고려대학교 통계학과)
3. On an exhaustive SDR methods for a Large Scale Dataset 15
류채현*, 김경원(이화여자대학교 통계학과)
4. Variable Selection for Individualized Quantile Treatment Effects 17
박서현*, 성민지, 최상범(고려대학교 통계학과)

논문발표 A3(202호)

농업/환경/기상/통계(응용)

좌장 : 최보승(고려대학교)

09:30	1. 중국 30개 성(省)을 탄소 배출을 군집단 분석	19
10:50	CHEN SHANGYU*, 박성용(중앙대학교 경제학부)	
	2. Change point detection을 이용한 대한민국 기온 변화 분석	21
	박현우*, 신승준(고려대학교 통계학과), 윤건수(고려대학교 행정학과)	
	3. XGBoost, LightGBM을 활용한 시정 예측 모델	23
	신봉균*(대진대학교 AI빅데이터학과), 서혜선(대진대학교 AI융합학부)	
	4. Bayesian Multiple Change-point Estimation for Hazard in Weibull Distributions	29
	주아림*(Department of Economics and Statistics, Korea University),	
	전수영(Division of Big Data Science, Korea University),	
	김재희(Department of Statistics, Duksung Women's University)	

논문발표 A4(303호)

통계 (일반)

좌장 : 신승준(고려대학교)

09:30	1. Role of Maternal Depression and Network Trajectories	
	on Internet Addiction in Korean Youth	31
10:50	김주안*(연세대학교 의학전산통계학협동과정), 이상진(서울대학교 통계학과),	
	김진현(연세대학교 예방의학교실), 하민진(연세대학교 보건정보통계학과)	
	2. Ridge regression splines with knot selection via ℓ_0 penalty	33
	이은지*, 정재환(충북대학교 통계학과)	
	3. Zero inflated high dimensional compositional data with DeepInsight	35
	이제석*, 김병원(경북대학교 통계학과)	
	4. Prediction of Delisted Companies Using PSM-Based Machine Learning and	
	Deep Learning Models	37
	김연주*, 김병원(경북대학교 통계학과)	

논문발표 B: 15:10 - 16:30

논문발표 B1(201호)

경제 / 경영

좌장 : 이재우(고려대학교)

-
- | | |
|---------------------|---|
| 15:10

16:30 | 1. 문화재정지출이 지역경제에 미치는 영향 분석 39
오주현*(가천대학교 경영학부) |
| | 2. 담배 제품별 교차 가격탄력성 분석 연구 41
이성규, 김진영(한국담배규제연구교육센터),
최은철*(한국해양대학교 국제무역경제학부) |
| | 3. DEVELOPMENT AND EVALUATION OF ARTIFICIAL INTELLIGENCE TECHNIQUES
TO REVOLUTIONIZE THE MARKETING STRATEGIES 43
Minho Sun, Seung Woo Kim, Jai Woo Lee*(고려대학교 빅데이터사이언스학부) |

논문발표 B2(301호)

교육/심리/사회

좌장 : 정혜원(충남대학교)

-
- | | |
|---------------------|--|
| 15:10

16:30 | 1. 한국의 인구 변화와 가족 관계: 노인 지원 및
지속 가능한 장기 요양 시스템을 위한 대안 45
이지나(고려대학교 사회복지학과), 황명진*(고려대학교 공공사회통일외교학부) |
| | 2. DEU 핵심역량과 학부교육 실태의 연계성 분석 47
이순화, 김경희(동의대학교 교육성과관리센터), 강주리(동의대학교 교수학습개발센터),
이영창(충남대학교 심리학과), 김보성*(동의대학교 교육성과관리센터) |
| | 3. 아동학대 발굴 체계 강화를 위한 아동 건강정보 연계 분석 49
김선월*(한국사회보장정보원), 박준범(순천향대학교 서울병원) |
| | 4. 건강보험청구자료 활용을 위한 환례정의 마련 연구 51
김연희*, 이돈형(한국건강증진개발원) |

논문발표 B3(202호)

통계 (차원축소)

좌장 : 이성건(성신여자대학교)

15:10	1. Applications of Response Dimension Reduction in Large p-Small n Problems 53
	김민지*, 유재근(이화여자대학교 통계학과)
16:30	2. Projective Resampling with the Minimum Average Variance Estimator: Sufficiently Reducing Dimension with Multivariate Response 55
	김서영*, 유재근(이화여자대학교 통계학과)
	3. Multivariate Response Directional Regression: A Projective Resampling Approach 57
	이아름*, 김경원(이화여자대학교 통계학과)
	4. Comparative Analysis of Likelihood-Based Dimension Reduction Methods 59
	유혜민*, 유재근(이화여자대학교 통계학과)

논문발표 B4(303호)

경영/경제/통계(응용)

좌장 : 연구필(호서대학교)

15:10	1. Applying quantile regression to determine the effects of household characteristics on household saving rates in Vietnam 61
	Thanh Hoa Bui*, Keunjae Lee(Department of Economics, Pusan National University)
16:30	2. A Comprehensive Software Reliability Growth Model Incorporating Dependent Failures and Uncertain Operating Environments 63
	Onon-Ujin Otgonbayar*, Kwang Yoon Song(Chosun University, Department of Computer Science and Statistics), Youn Su Kim(Institute of Well-Aging Medicare & CSU G-LAMP Project Group), In Hong Chang(Chosun University, Department of Computer Science and Statistics)
	3. Revisit Forward Rate Unbiasedness Hypothesis using Smooth Time-Varying Cointegration Regression Model with Unknown Structural Breaks 65
	Xiaoyu Kang*, Sung Y. Park(School of Economics, Chung-Ang University)
	4. Ethnic Minorities in Vietnam: A Study of the Factors Influencing the Living Standards (Preliminary) 67
	Nguyen Tuan Dat*, Keunjae Lee(Department of Economics, Pusan National University)

논문발표 C: 16:40 - 18:00

논문발표 C1(201호)

통계 (응용)

좌장 : 전세봄(목원대학교)

-
- 16:40 | 1. Estimation of weather impacts on mosquito abundance using Bayesian period-varying
18:00 | Poisson modeling with reduced-rank regression 69
이지은*, 김용구, 양승엽, 이준영, 오세민, 변성준, 김영호, 박예은,
김명진(경북대학교)
2. Exploring Regression Associations in Multidimensional Contingency Tables with an Ordinal Dependent Variable 71
Saebom Jeon*(Department of Marketing Bigdata, Mokwon University),
Daeyoung Kim(Department of Mathematics and Statistics, University of Massachusetts Amherst)
3. Developing a Carbon Emission Index:
A Framework Using ESG Reports and Text Mining 73
김성재*, 조용복(동아대학교 경영정보학과)

논문발표 C2(301호)

교육/심리/사회/통계(응용)

좌장 : 이상인(충남대학교)

-
- 16:40 | 1. 사이버범죄 데이터에 대한 텍스트마이닝 75
18:00 | 김근아*, 한상태(호서대학교 데이터사이언스)
2. 지각적으로 현저한 자극과 정서적으로 현저한 자극에 대한 기능적 뇌 네트워크의 자발적 반응:기능적 데이터분석을 통한 접근 77
정혜찬*, 정고은, 한석원(충남대학교 심리학과),
이은지(충남대학교 정보통계학과)
3. 기업 구성원의 직무만족도 예측변인 탐색:
머신러닝과 설명 가능한 인공지능(XAI)의 적용 79
장은아*, 정혜원(충남대학교 교육학과)
4. Model evaluation criterion for anti-cancer
drug sensitivity-specific gene network inference 81
오주이*, 박희원(성신여자대학교 통계학과)

논문발표 C3(202호)

통계 (딥러닝 응용)

좌장 : 양대원(충남대학교)

16:40	1. 딥러닝 기반 제네레이티브 모델을 활용한 함수형 데이터 합성	83
18:00		이동훈*, 송준(고려대학교 통계학과)
	2. A Deep Semi-Supervised Learning framework using Learnable Data Augmentation technique via Masked Autoencoders for tabular data	85
		이지혜*, 김동하(성신여자대학교 통계학과)
	3. Improving Outlier Detection by Optimizing the Inlier-Memorization Effect in Deep Generative Models	87
		조서영*(성신여자대학교 통계학과), 황재성(SK Telecom), 박관영, 김동하 (성신여자대학교 통계학과)
	4. 온센서 음성 분류를 위한 전이학습 기반의 신경망 구조 자동 탐색 기법	89
		전성재*(서울시립대학교 인공지능학과), 이영민(서강대학교 인공지능학과)

논문발표 C4(303호)

의료/간호/보건/통계(응용)

좌장 : 강현철(호서대학교)

16:40	1. Functional SKAT : 해마의 형태학적 표현형에 대한 유전적 변이 탐지 연구	91
18:00		임현정*, 이윤석, 이은지(충남대학교 통계학과)
	2. Improving Time Series Forecasting Accuracy with Increasing Weighted Loss Functions for Further Prediction	93
		전수민*, 송준(고려대학교 통계학과)
	3. Automatic Seed Word Selection for Topic Modeling	95
		정다현*(서울시립대학교 통계학과), 황정인(서울시립대학교 통계데이터사이언스학과), 최연진, 김윤영(서울시립대학교 통계학과)
	4. BERT-based graph level anomaly detection for molecular graphs	97
		김기원*(서울시립대학교 통계데이터사이언스학과), 진종준(서울시립대학교 통계학과)

1. 시계열 모형을 이용한 커피음료와 탄산음료 매출액 예측모형 개발	99
김민지*(호서대학교 데이터사이언스학과), 한상태(호서대학교 빅데이터 AI학부)	
2. Low-Dimension Adaptive Neural Network Regression with Directional Change Detection via Nuclear Norm Penalization	101
김용구*(고려대학교 통계학과), 정재환(충북대학교 통계학과), 구자용(고려대학교 통계학과), 박관영(성신여자대학교 통계학과)	
3. 웨이퍼 결함 패턴 분류를 위한 계층적 분류 모델과 특징 중요도 분석	103
김재운*(한양대학교 응용통계학과), 정재홍(한양대학교 수학과)	
4. 머신러닝 기반 안면인식 기술을 활용한 피부 시술의 안면 동안 효과에 대한 정량화 연구	105
장현규(이화여자대학교 국제대학원), 노민정*(부경대학교 경영학부)	
5. 침수 피해 위험 예측을 위한 모형 개발	107
박영재*(고려대학교 빅데이터사이언스학부), 안수경(고려대학교 국가통계전공), 최보승(고려대학교 빅데이터사이언스학부)	
6. Impact of Global Fertilizer Prices on Import Allocation	109
황보준성*, 노연수, 서동희(고려대학교 식품자원경제학과)	
7. B형 간염 환자의 간암 발병 위험 조기 판별 모형	111
서준호*, 남진현(고려대학교 빅데이터사이언스학부)	
8. 농산물 재배조사를 위한 표본설계	113
손창균*, 이단비(동국대학교 빅데이터·응용통계학과), 홍기학(동신대학교 컴퓨터학과), 이기성(우석대학교, 아동복지학과)	
9. 한국의 교통문화 설문응답에 대한 분석	115
신지성*(충남대학교 통계데이터사이언스학과), 양대원(충남대학교 정보통계학과)	
10. Testing for Bubbles in the Korean won - US dollar Exchange Rate	117
심성훈*(부경대학교 경제학과)	
11. 현대 학생들의 과도한 사회화 연구	121
오열*, 양정호(성균관대학교 교육학과)	
12. 간호대학생의 학업지연행동에 미치는 영향 요인	123
오윤정*, 이은미(남부대학교 간호학과)	
13. 중국대학생의 전공선택과 직업선택의 충돌 문제	127
오적*, 양정호(성균관대학교 교육학과)	
14. 한국형 심방세동 위험 점수 모형 KAF 개발 연구	131
윤지원*(고려대학교 응용통계학과), 남진현(고려대학교 빅데이터사이언스학부)	

15. 증화 직접질문 조건부 무관질문모형	133
이기성*(우석대학교 아동사회복지학부), 홍기학(동신대학교 컴퓨터학과), 손창균(동국대학교 빅데이터·응용통계학과), 박근화(한국문화관광연구원), 홍성준(마사회 말산업연구소)	
16. Comparison of algorithms for estimating regression splines using lasso penalty	137
이은지(충북대학교 통계학과), 박중범, 이동영*(충북대학교 정보통계학과), 정재환(충북대학교 정보통계학과)	
17. Dynamic Time Warping을 활용한 포트폴리오 최적화 방법	139
이석준*(한양대학교 응용통계학과), 정재홍(한양대학교 수학과)	
18. 소셜 빅데이터에 나타난 고독에 대한 인식 분석	141
전란영*(한국인문사회총연합회)	
19. 로버스트 왜도를 기반으로 한 이상치 탐지에 관한 연구.	
영향함수 산출 및 산업자료로의 응용	143
정운채*(충남대학교 정보통계학과), 정수빈(충남대학교 데이터사이언스학과), 박민수(충남대학교 데이터사이언스학과)	
20. 베이지안 기법을 적용한 당뇨병 및 당뇨합병증 예측 딥러닝 모델 설계	145
조영찬*, 박성희(국립금오공과대학교 일반대학원 수학과), 이경준(국립금오공과대학교 수리빅데이터학과)	
21. Llama-3 모델을 활용한 법정감염병 정보 챗봇 구현	147
주진범*, 형도훈(부경대학교 데이터공학과), 김창훈(부산대학교 의과대학), 노맹석, 문형빈(부경대학교 정보융합대학 빅데이터융합전공)	
22. Robust Slope Estimation in Progress Data from AI-Driven ALEKS:	
Application of Theil-Sen Estimator	149
왕훈이*, 한현석(전북대학교 심리학과)	
23. 효과적인 국문 판례문 요약을 위한 거대언어모형 비교연구	151
홍인석*, 최호식(서울시립대학교 대학원 도시빅데이터융합학과)	
24. 중국 고등교육 입학기회의 공평성 연구	153
GUO ZIFEI*(성균관대학교 교육사회학과)	
25. Statistical Inference Methods for Two-Parameter Exponential Distribution in Generalized Progressive Hybrid Censoring Schemes	159
Jihyeon Kim*, Yeongeun Hwang, Kyeongjun Lee (Department of Mathematics, Kumoh National Institute of Technology)	
26. Exponentiality Testing for Progressive Type-II Censored Data with a Local Linear Regression Estimator of Kullback-Leibler Information	161
SangHeon Jeong*, Kyeongjun Lee(Department of Mathematics, Kumoh National Institute of Technology)	

27. 현대화 교육에서의 AI 활용이 교사 교육 방식에 미치는 영향 탐구 163
 Yang Yaqing*(성균관대학교 교육사회학과)

28. 간호대학생의 자기주도학습, 학업적 자기효능감,
 비판적 사고경향 및 셀프리더십의 관계 167
 이은미*, 오윤정(남부대학교 간호학과)

29. 생존확률 외삽을 위한 최적 노츠 탐색 연구 169
 정완석*(고려대학교 응용계학과), 남진현(고려대학교 빅데이터사이언스학부)

30. 서울시 공공자전거 데이터 분석을 통한 도시 문제 개선방안 연구 171
 권선일*, 최호식(서울시립대학교 도시빅데이터융합학과)

31. 50세 이상 성인여성의 골관절염 유병률과 관련 요인 173
 전은미*(배재대학교 간호학과)

32. Investigating Meteorological and Environmental Effects
 on COVID-19 Transmission in South Korea 175
 이현민*(고려대학교 경제통계학과), 최보승(고려대학교 빅데이터사이언스학부)

33. 네트워크 모델링 접근에 기반한 확률적 전염병 모델 개발 177
 박경호*(고려대학교 응용통계학과), 최보승(고려대학교 빅데이터사이언스학부)

34. Hi-C 데이터를 활용한 희소 가중 그래프 기반 대장암 예측 모형에 관한 연구 179
 고민규*(충남대학교 통계데이터사이언스학과), 임지원(충남대학교 정보통계학과),
 장인수(충남대학교 통계데이터사이언스학과&한국생명공학연구원),
 박민수(충남대학교 통계데이터사이언스학과)

35. 메타데이터와 텍스트 데이터를 결합한 온라인 리뷰 유용성 예측 모델 및
 해석 가능성 연구 181
 김유영*, 최호식(서울시립대학교 도시빅데이터융합학과)

36. 학과목 학원 금지 정책이 중국 지방 한자녀 가정에 미치는 영향 분석 183
 도과우(TAO KEYU)*, 양정호(성균관대학교 교육사회학과)

37. 불법 온라인 성매매 광고 분류 및 자동 신고를 위한 다중 모달 시스템 개발 189
 박민희*(서울시립대학교 도시빅데이터융합학과), 박수민(연세대학교 통계데이터사이언스학과),
 윤성범, 김준철(AI 빅데이터랩, 서울연구원), 최호식(서울시립대학교 도시빅데이터융합학과),
 송경우(연세대학교 응용통계학과)

38. 동형암호를 통한 이중 데이터 결합 및 분석의 실용성 검토 191
 최민호*, 정효혁, 신성철((주)크립토패), 최호식(서울시립대학교 도시빅데이터융합학과)

39. Diversity in Integrated Interventional Radiology Residency Programs 193
 김명지*, 현진섭, 이은지(충남대학교 정보통계학과),
 Samir S. Jambhekar(The Ohio State University Wexner Medical Center Department of Radiology)

40. 인구고령화 시대에 중국 노인교육의 수요분석 195
등자운*, 양정호(성균관대학교 교육학과)

41. Predictive Model for Survivors of the Estonian Shipwreck 199
Xin Zan*, Jai Woo Lee(고려대학교 빅데이터사이언스학부)

42. 머신러닝 방법론을 결합한 혼합효과 모형의 성능 비교 연구 201
박정연*(충남대학교 통계데이터사이언스학과), 이상인(충남대학교 정보통계학과)

43. INTEGRATIVE MACHINE LEARNING TECHNIQUES FOR MODELLING COMPLEX SYSTEMS IN ALZHEIMER’S DISEASE 203
Juwon Kim*, Sangjin Kim, Jai Woo Lee(고려대학교 빅데이터사이언스학부)

44. AI-DRIVEN APPROACHES FOR ANALYSIS OF SOCIO-DEMOGRAPHIC AND HEALTH-RELATED FACTORS AFFECTING INDIVIDUAL DEPRESSION 205
Hyeonjun Nam*, Donggeun Kim, Jai Woo Lee(고려대학교 빅데이터사이언스학부)

45. AI-BASED RECOMMENDATION SYSTEM APPLIED IN CONSUMER BEHAVIOR ANALYSIS FOR A NEW AUTOMOBILE SALES MARKETING MODEL 207
Dongha Kim*, Juyong Ko, Jai Woo Lee(고려대학교 빅데이터사이언스학부)

46. PREDICTION OF FUTURE RISK TO SUICIDAL IDEATION BASED ON TEXT DATA FROM SOCIAL MEDIA WITH A MACHINE LEARNING APPROACH 209
Kin Lok Wong*, Jai Woo Lee(고려대학교 빅데이터사이언스학부)

47. Classification of Tourist Movement Patterns Using Mobile Communication Data 211
남성재*, 이홍재, 최보승(고려대학교 빅데이터사이언스학부)

48. AR 오차모형을 포함한 회귀모형을 통한
확진자수와 인구이동량의 연관에 대한 연구 213
이선재*, 최보승(고려대학교 빅데이터사이언스학부)

튜토리얼

(1) 발표자 : 신승준 (고려대학교 통계학과)

(2) 제목 : 분류문제의 이해와 활용

최근 인공지능의 발전으로 분류는 데이터 분석에서 가장 대중적인 문제가 되었으며, 많은 분야에서 널리 활용되고 있다. 본 튜토리얼에서는 다양한 분류 학습 기법을 소개하고, 이들의 차이점과 유사점을 체계적으로 살펴보고자 한다. 데이터 기반의 학습 방법은 크게 두 가지 방식으로 나눌 수 있는데, 이는 데이터를 확률변수로 보는가 (통계학), 아니면 숫자로 보는가 (공학)에 따라 구분된다. 데이터를 확률변수로 본다면, 데이터 기반의 학습은 데이터의 생성 분포를 추정하는 문제로 이해할 수 있다. 대표적인 분류학습의 예로 로지스틱 회귀분석, 선형판별분석 등을 들 수 있다. 반면, 데이터를 숫자로 이해한다면, 데이터의 패턴을 학습하는 문제로 이해할 수 있으며, 대표적인 예로는 Support Vector Machine, 인공신경망 등이 있다. 서로 달라 보이는 두 가지 방법은 오분류율을 최소화한다는 관점에서 그 유사성을 찾을 수 있으며, 이를 통해 데이터 분석에서 널리 활용되는 다양한 분류 학습 기법 간의 관계를 확인하고 그 이해의 폭을 넓힐 수 있다. 나아가, 오분류율을 최소화하는 것만으로 문제를 해결하지 못하는 경우가 실제 응용에서는 종종 발생한다. 오분류율이 아닌, 오분류 비용, ROC곡선 등의 대안적인 분류 성능의 기준을 활용하여 분류 문제를 해결하는 방법에 대해서도 살펴보고자 한다.

초청강연

(1) 발표자 : 원유집 교수 (KAIST 전기 및 전자공학부)

(2) 제목 : AI시대 살아남기

AI 기술의 여파가 핵폭풍을 능가한다. 지난 10월 26일 미국 조바이든 대통령이 AI 안보각서에 서명했다. 지난 9월 윤석열 정부는 국가 AI기술관련하여 기술개발뿐 아니라 법과 제도까지를 모두 관장하는 국가 AI위원회를 발족하였다. 작금 주요국의 AI기술 개발관련한 정책은 2차세계대전직후 핵무기 개발경쟁 초기의 상황과 비슷할 정도로 치열하고, 심각하다. AI시대에 기술적 우위를 선점하는 것은 국가, 기업뿐만아니라 개개인에게도 선택하는 필수사항이 되어버렸다. 본 강연에서는 현대 AI기술의 발전사, 현재 AI 기술이 직면하고 있는 문제인 에너지 벽, 그리고 AI기술에의 연구개발 투자 현황을 다룬다. 이를 기반으로 우리가 학계로서, 택할 수 있는 전략을 함께 고민해본다.

Autoencoder-Driven Feature Extraction in Raman Spectroscopy Analysis for Hydrocarbon Mixtures*

*Seung Jee Yang*¹, *Woosuk Sohng*², *Hoeil Chung*³

Abstract

Recent studies have applied autoencoders (AEs) in Raman spectroscopic analysis, demonstrating improved analytical performance, yet often without explaining the reasons behind this enhancement. In this work, we experimentally investigated the underlying reasons, focusing on the analysis of complex mixtures typical in industrial settings. Specifically, we examined the effectiveness of AE-based feature extraction for compositional analysis of 11 hydrocarbon mixtures and its potential application to vibrational spectroscopy in the petroleum and petrochemical industries. The results showed that AE-based features effectively capture highly overlapping or weak spectral signals, leading to enhanced accuracy in concentration predictions. This approach offers significant potential for real-world applications, improving the analysis of complex hydrocarbon mixtures in industrial contexts.

Keywords : Autoencoder, Feature extraction, Near-infrared spectroscopy, Compositional analysis.

*This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A06046728).

¹Postdoc, Department of Applied Statistics and Research Institute for Convergence of Basic Science, Hanyang University, Seoul 04763, Korea.

E-mail: okseungji@hanyang.ac.kr

²Ph.D, Department of Chemistry and Research Institute for Convergence of Basic Science, Hanyang University, Seoul 04763, Korea. E-mail: sohng@hanyang.ac.kr

³(Corresponding author) Professor, Department of Chemistry and Research Institute for Convergence of Basic Science, Hanyang University, Seoul 04763, Korea.

E-mail: hoeil@hanyang.ac.kr

출원 특허 기반의 복합어 키워드를 활용한 디지털 헬스케어 분야 이머징 기술 트렌드 분석*

이서엘¹, 김민기², 정호현³, 홍미영⁴

요 약

과학기술의 급속한 발전으로 효율적인 예산 배분을 위한 각 분야별 유망 기술 발굴 및 우선 순위 설정은 국가의 기술 경쟁력을 확보하는데 필수적이다. 기존 연구들은 다양한 머신러닝 기법 및 통계적 방법론을 활용하여 연구 동향을 파악하고 기술 추세를 분석해 왔다. 하지만 공개 시점이 늦은 데이터로 인해 최신 기술을 신속하게 포착하기 어렵고, 대부분 단일어 기반 분석이기 때문에 텍스트 데이터가 지닌 맥락을 충분히 반영할 수 없고, 회귀분석 기반 방법은 최근 급부상하는 이머징 토픽의 의미를 충분히 담아내지 못하는 등의 한계가 있었다. 이를 개선하기 위해 본 연구는 출원 특허 기반의 복합어 키워드를 활용한 디지털 헬스케어 분야의 이머징 기술 트렌드 분석을 제안하였다. 해당 방법은 출원 특허를 활용하여 기술 트렌드 분석에 있어 최신성을 확보하였고, 복합어 키워드 단위의 데이터를 활용함으로써 기존의 단일어 기반의 분석보다 더 많은 맥락적 의미를 반영하였다. 토픽 출현 지도(Topic Emergence Map, TEM)를 활용하여 중요도는 낮지만 높은 성장률을 가지는 토픽을 이머징 토픽으로 정의하고, 이를 통해 복합어 키워드가 가지는 충분한 의미와 해석이 가능한 수준에서의 이머징 기술을 도출하였다. 해당 방법은 디지털 헬스케어 분야 이외에도 다양한 분야에 적용 가능하며 추후 전략적 R&D 기획 및 자원 배분에 있어 신규 투자 영역 등 새로운 자원 수요 발굴에 활용성이 높을 것으로 기대된다.

주요 용어 : 유망 기술 분석, 텍스트 데이터, 이머징 토픽, 토픽 모델링, 토픽 출현 지도.

*본 연구는 한국과학기술기획평가원(KISTEP)의 기관과유사업으로 수행한 “근거 기반 R&D예산 배분·조정 지원 및 투자전략 연구(2024년)”의 일부 내용을 수정·보완한 것임(AV24070).

¹27740 충청북도 음성군 맹동면 원중로 1339, 한국과학기술기획평가원 사업조정평가본부 사업조정전략센터 위촉연구원. E-mail: lse1031@kistep.re.kr

²27740 충청북도 음성군 맹동면 원중로 1339, 한국과학기술기획평가원 사업조정평가본부 사업조정전략센터 연구위원. E-mail: mk@kistep.re.kr

³02844 서울특별시 성북구 보문로 34다길 2, 성신여자대학교 수리통계데이터사이언스학부 조교수, 데이터사이언스센터. E-mail: hhjung@sungshin.ac.kr

⁴(교신저자)27740 충청북도 음성군 맹동면 원중로 1339, 한국과학기술기획평가원 사업조정평가본부 사업조정전략센터 센터장. E-mail: myhong@kistep.re.kr

투표의 역설적 측면에 관한 통계 교육용 고찰

김혁주¹

요 약

투표에는 사람들이 깊이 생각하지 않고 간과하기 쉬운 역설적 측면들이 있어서, 투표 방식과 그에 따른 결정 방식을 어떻게 사용하느냐에 따라 결과가 크게 달라질 수 있다. 본 발표에서는 교양 수업 중 투표에 관해 학생들에게 얘기한 몇 가지 내용을 정리하였다.

주요 용어: 투표, 역설, 과반수제, 종다수제.

1. 서론

투표에는 여러 후보자나 여러 안 중 일부를 뽑기 위한 투표, 어떤 안건에 대한 찬반 투표 등이 있다. 투표란 단순히 표를 많이 얻은 쪽을 선택하는 것이라고 생각하기 쉽고 또 실제로 그렇게 처리되는 경우도 많지만, 깊이 생각하지 않고 간과하기 쉬운 역설적인 측면들이 있어서, 투표 방식과 그에 따른 결정 방식을 어떻게 사용하느냐에 따라 결과가 크게 달라질 수도 있다. 본 발표에서는 교양 수업 중 투표에 관해 학생들에게 얘기한 몇 가지 내용을 정리한다. 소개되는 사례 중에는 이미 알려진 내용도 있지만, 필자의 경험에 의해 처음 소개되는 사례도 있다.

2. 몇 가지 경우로 살펴본 역설적 상황

2.1. 경우 1(콩도르세의 역설)

A, B, C 3인이 과반수제에 의한 표결로 문제를 처리하는데, 세 가지 안이 후보로 올라왔다. 각 사람은 이 안들을 다음과 같이 평가하고 있다.

이 경우 각자 가장 좋아하는 안을 선택하는 투표를 한다면 결론이 나지 않으므로 표결을 다음 방식에 의해 진행하기로 하였다. 우선 두 안을 짝지어 표결에 부쳐 과반수의 지지를 얻은 안을 제 3의 안과 다시 짝지어 표결에 부친다. 그 결과 과반수의 지지를 얻은 안을 최종적으로 채택한다. 이렇게 하면 짝짓는 순서에 따라 최종적으로 결정되는 안이 달라진다. 이러한 경우 의사 진행 과정을 장악하고 있는 사람이 특정한 안에 유리한 결과가 나오도록 큰 영향력을 행사할 수 있다는 문제점이 있다.

2.2. 경우 2(심슨의 역설)

투표와 밀접한 관련이 있는 여론조사의 예를 생각해 보자. 어느 지역 주민들을 대상으로 두 중

¹54538 전북 익산시 익산대로 460, 원광대학교 빅데이터·금융통계학부 교수. E-mail: hjkim@wku.ac.kr

류의 안전에 대한 찬반 여론조사를 실시하여 남녀별로, 그리고 전체적으로 찬성률을 계산한 결과가 다음과 같다. 여기서 A안과 B안은 직접적으로 경쟁하는 안이 아니고, 별개의 종류의 안이다(예를 들면, A안은 도로 확장, B안은 도서관 건립).

남녀별로 보면 A안에 대한 찬성률이 모두 B안보다 높지만, 전체적으로 보면 B안에 대한 찬성률이 더 높다. A안에 대한 여론조사에서는 찬성률이 낮은 여자가 많이 응답했고, B안에 대한 여론조사에서는 찬성률이 상대적으로 높은 남자가 많이 응답했기 때문에 부분에 대한 분석 결과와 전체에 대한 분석 결과가 일치하지 않는 현상이 나타났다.

2.3. 경우 3(최선을 택하느냐, 최악을 먼저 떨어뜨리느냐)

여섯 사람이 함께 휴가를 어디로 갈 것인가를 놓고 투표로 결정하려 한다. 후보지는 세 곳(산, 바다, 놀이공원)이며, 각자 가장 가고 싶은 곳에 투표를 해서 가장 많은 표를 얻은 곳이 휴가지로 결정된다. 여섯 사람 각각의 좋아하는 순서가 다음과 같다고 하자.

투표 결과 바다 3표, 산 2표, 놀이공원 1표로 휴가지가 바다로 결정되었다. 그런데 바다는 가장 가기 싫어하는(제3 선택) 사람이 가장 많은 곳이기도 하다. 만약 가장 가기 싫은 곳에 투표하여 가장 많은 표가 나온 곳부터 탈락시키는 방식을 사용한다면 제일 먼저 탈락할 곳이 바다일 것이며, 이 경우에는 휴가지가 산으로 결정됐을 것이다.

Table 1. 세 가지 안에 대한 3인의 평가

	A	B	C
제1 선택	1안	2안	3안
제2 선택	2안	3안	1안
제3 선택	3안	1안	2안

Table 2. 두 가지 안에 대한 찬성률

A안	응답자 수	찬성자 수	찬성률
남자	50	25	50%
여자	100	30	30%
전체	150	55	37%
B안	응답자 수	찬성자 수	찬성률
남자	100	45	45%
여자	50	13	26%
전체	150	58	39%

Table 3. 휴가지에 대한 6인의 선호 순서

	A	B	C	D	E	F
제1 선택	바다	바다	바다	산	산	놀이공원
제2 선택	산	산	놀이공원	놀이공원	놀이공원	산
제3 선택	놀이공원	놀이공원	산	바다	바다	바다

Table 4. 6가지 가능한 선호 순서에 대한 유권자들의 비율

제1 선택	제2 선택	제3 선택	비율
A	B	C	20%
A	C	B	20%
B	A	C	7%
B	C	A	28%
C	A	B	5%
C	B	A	20%

2.4. 경우 4(후보 단일화 또는 결선투표를 한다면?)

세 후보가 출마한 선거에서 A, B, C 후보의 지지율이 각각 40%, 35%, 25%라고 하자. 일반적인 종다수제의 투표를 적용하면 A후보가 당선될 것이다. 그런데 이 세 후보를 선호하는 순서까지 생각해서 비율을 조사한다면 다음과 같은 결과가 나왔을 것이라고 가정하자.

B와 C가 후보 단일화를 한다고 가정하자. B가 단일후보가 된다면 A후보 45%, B후보 55% 득표로 B후보가 당선되며, C가 단일후보가 된다면 A후보 47%, B후보 53% 득표로 C후보가 당선되게 된다. 결선투표제를 적용하는 경우에는, 1차 투표에서 과반 득표자가 없으므로 1위인 A후보와 2위인 B후보가 결선투표에서 맞붙어 A후보 45%, B후보 55%의 득표율로 B후보가 최종 승자가 될 것이다.

2.5. 경우 5(1인2표제의 경우) (출처: Park (2012))

유권자가 세 개의 큰 집단 A, B, C로 나누어 있고, 각 집단에서 후보가 나왔다고 하자. 세 집단의 규모가 A집단이 41명, B집단이 39명, C집단이 20명이라고 하면, 단순한 다수결 투표로는 A집단의 후보가 가장 유리하다. 그러나 1인2표제를 적용한다면, A집단의 유권자는 일단 한 표는 자기 집단의 후보에게 던지지만, 다른 한 표는 B집단 후보를 견제해야 되므로 C에게 던질 수밖에 없다. B집단도 마찬가지로 한 표는 자기 집단의 후보에게, 다른 한 표는 A를 견제하려고 C에게 던지게 된다. 그러나 C집단 유권자는 A도 견제해야 되고 B도 견제해야 된다. 따라서 만약 C집단이 한 표는 자기 집단 후보에게 던지고, 다른 한 표는 정확히 반씩 나누어 A와 B집단 후보에게 던졌다고 하자. 그러면, A집단의 후보는 $41+20/2 = 51$ 표, B집단의 후보는 $39+20/2 = 49$ 표인데, C집단의 후보는 $41+39+20 = 100$ 표가 되어, 압도적인 차이로 C집단의 후보가 당선된다. 단순 다수결로는 풀찌일 후보가 당선되는 상황이 되는 것이다.

2.6. 경우 6(불필요하고 불합리한 조 편성)

2014년 11월 JTBC의 <히든싱어>라는 TV 프로그램에서 <히든싱어 왕중왕전>을 진행하고 있다. 각자 원조 가수를 모창한 11명의 모창자들 중 최종 결승에 진출할 3명을 300명의 방청객들(패널 포함)의 투표로 뽑는다. 방청객들이 각각의 모창자에 대해 평가하는데, 모창자가 모창을 잘했다고 생각하면 버튼을 누르는 방식이다. 모창자들은 사전 추첨에 의해 3개 조로 나뉘었으며, 프로그램 제작진은 각 조의 최다득표자가 결승에 진출하도록 규정을 정했다. 투표 결과는 다음과 같고, 규정에 따라 모창자 2번, 6번, 9번이 결승에 진출했다.

Table 6. 모창자 11명의 득표 수

조	모창자	원조 가수	득표 수
A	1	이선희	206
	2	박현빈	269
	3	김태우(god)	96
	4	이적(패닉)	147
B	5	이재훈(쿨)	271
	6	환희(플라이투더스카이)	280
	7	윤종신	154
	8	태진아	165
C	9	이승환	237
	10	태연(소녀시대)	149
	11	인순이	104

이 방식에는 심각한 오류가 있다. 조를 나누는 것은 불필요하고 불합리하며, 전체 참가자들 중 득표순으로 1, 2, 3위(모창자 6번, 5번, 2번)를 결승에 진출시켜야 합리적이다. 실제로 이 프로그램에서 전체 득표 순위 4위인 모창자(9번)가 결승에 진출하는데도 2위인 모창자(5번)는 탈락할 상황이 발생하자, 제작진이 탈락자 8명에 대해 시청자들의 투표에 의해 1명을 뽑아 추가로 결승에 진출시키는 와일드카드 제도를 적용하여 모창자 5번이 추가로 결승에 진출했다.

2.7. 경우 7(경기 결과 예측 투표의 경우)

2002 한·일 월드컵을 앞둔 2002년 4월 KBS의 <비바 월드컵>이라는 TV 프로그램에서 시청자들의 투표에 의한 가상 월드컵이 진행되고 있다. 16강이 정해진 상황에서 대진은 다음과 같으며, 매주 8강 진출팀 두 팀씩을 투표에 의해 뽑는다.

이번 주는 A, B, C, D팀 중 어느 두 팀이 8강에 진출할지를 결정할 시점이다. 시청자들은 보기 ① A팀과 C팀, ② A팀과 D팀, ③ B팀과 C팀, ④ B팀과 D팀 중 한 가지에 투표를 하며, 가장 많은 표를 얻은 번호의 두 팀이 8강에 진출한다.

이 방식에는 문제점이 있다. 예를 들어 ①, ②, ③, ④에 투표한 시청자의 비율이 각각 35%, 10%, 25%, 30%라 하자. 그러면 A팀과 C팀이 8강에 진출하는 것으로 처리될 것이다. 그런데 따져 보면 A팀과 B팀 중 8강 진출팀을 A팀으로 예상한 시청자는 $35\%+10\%=45\%$ 이며, B팀으로 예상한 시청자는 $25\%+30\%=55\%$ 이다. 또한 C팀과 D팀 중 8강 진출팀을 C팀으로 예상한 시청자는 $35\%+25\%=60\%$ 이며, D팀으로 예상한 시청자는 $10\%+30\%=40\%$ 이다. 따라서 투표 결과가 이렇게 나온 경우에는 B팀과 C팀이 8강에 진출하는 것으로 처음부터 방식을 정했어야 한다.

일반화하여 설명하면 다음과 같다. ①, ②, ③, ④에 투표한 시청자의 비율을 각각 p, q, r, s 라 하자. 일반적인 처리 방식은 p, q, r, s 중 최대 비율이 선택한 번호의 두 팀이 8강에 진출하는 방식이지만 이 방식에는 오류가 있고, $\max(p+q, r+s)$ 와 $\max(p+r, q+s)$ 에 공통으로 들어 있는 문자(위의 경우는 r)에 해당하는 번호의 두 팀이 8강에 진출하도록 하는 방식이 합리적이다.

당시 이 프로그램을 시청하던 필자로부터 위와 같은 문제점을 지적하는 의견을 받은 제작진은 8강 선정 이후의 단계부터 방식을 수정하여 프로그램을 진행하였다.

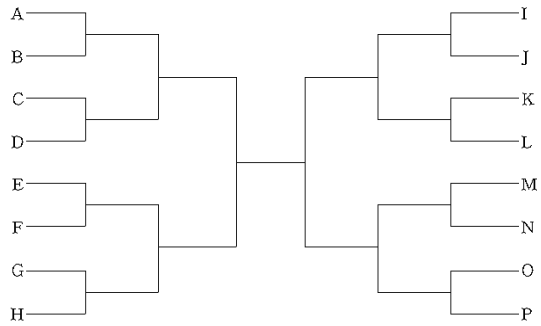


Figure 1. 가상 월드컵의 16강 대진표

References

Park, B. S. (2012), http://navercast.naver.com/contents.nhn?rid=22&contents_id=7827
<http://search.jtbc.co.kr/search/replay?term=히든싱어&type=vod&page=10>
<http://search.kbs.co.kr/main.html?keyword=비바+월드컵>

Sparse Sufficient Dimension Reduction for Directional Regression

*Gayun Kwon*¹, *Gijeong Noh*², *Kyongwon Kim*³

Abstract

Sufficient dimension reduction has emerged as a powerful tool for extracting meaningful information within high-dimensional datasets over the past few decades. These methods aim to reduce the complexity of data by focusing on its most informative components and this allows us to avoid ‘curse of dimensionality’. However, many sufficient dimension reduction methods have challenges because their outcome has the form of the linear combinations of the original predictors. This can make the interpretation of the extracted components quite difficult, particularly when working with a large number of variables. To address this issue, we introduce a sparse sufficient dimension reduction method for directional regression (Li and Wang, 2007). Our approach converts generalized eigendecomposition to regression type optimization problem with LASSO constraint and this promotes interpretability by generating sparse estimates. We demonstrate the efficacy of our approach by comparing it against existing methods such as non-sparse directional regression, sparse sliced inverse regression, and sliced average variance estimation through comprehensive numerical experiments. We further apply our method to two real-world datasets to present its practical value in extracting meaningful insights from complex data structures.

Keywords : Sufficient dimension reduction, Directional regression, LASSO, High-dimensional data, Multivariate analysis

References

- Li, B., & Wang, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*, 102(479), 997-1008.
- Li, L. (2007). Sparse sufficient dimension reduction. *Biometrika*, 94(3), 603-613.

¹Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemum-gu, Seoul, 03760, South Korea.. E-mail: gayunkwon.ewha@gmail.com

²R&D Center, OCI Company, 61, Sagimakgol-ro 62 beon-gil, Jungwon-gu, Seongnam, 13212, South Korea. E-mail: gjnoh@oci.co.kr

³(Corresponding author) Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemum-gu, Seoul, 03760, South Korea.. E-mail: kimk@ewha.ac.kr

Angle-based Deep Hierarchical Classification via Label Embedding

Hyeyeon Kim¹, Seung Jun Shin²

Abstract

Hierarchical classification problems frequently arise in various practical applications; however, most existing methods fail to fully utilize the hierarchical relationships among class labels. This study aims to enhance hierarchical classification accuracy by effectively integrating hierarchical information. We employ the label embedding technique proposed by Fan et al (2022)., which preserves hierarchical structures by embedding nodes into a low-dimensional space limited to $n_{\text{leaf}} - 1$ dimensions. Subsequently, we propose an angle-based deep neural network classifier that leverages all available data features through a top-down hierarchical classification approach. Experimental results on real-world datasets demonstrate that our proposed method achieves superior computational efficiency and classification accuracy compared to linear models, even without feature screening. These findings contribute to the development of scalable hierarchical classification models suitable for large-scale datasets.

Keywords : Hierarchical Classification, Label Embedding, Neural Network.

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

²(Corresponding Author)Associate Professor, Department of the Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: sjshin@korea.ac.kr

On an exhaustive SDR methods for a Large Scale Dataset

*Chaehyun Ryu*¹, *Kyongwon Kim*²

Abstract

Directional regression is an effective dimension reduction approach for capturing inherent characteristic in regression problems. We extend the idea of directional regression to handle massive datasets. In particular, this paper adopts a "divide and conquer" strategy, breaking down the dataset into manageable chunks, and subsequently merging the results based on the proximity between dimension reduction subspaces. We further harness the capabilities of capturing distance to significantly enhance computational efficiency and optimize memory usage. We demonstrate the competitiveness of our approach through a comprehensive numerical study as well as its application to a real world dataset. In both simulation and application, R packages "foreach" and "bigmemory" are utilized for optimizing the execution speed and managing memory when dealing with a massive dataset. The comparison between the proposed methodology, BIG-DR, and existing methods, namely BIG-SIR and BIG-SAVE, was conducted with a focus on computational speed and accuracy. The application of BIG-DR to real datasets demonstrates its practical applicability and versatility.

Keywords : Sufficient Dimension Reduction, massive dataset, parallel computing, sliced average variance estimate, directional regression.

¹Ph.D. Student, Department of Statistics, Ewha Womans University, Seoul 03760, South Korea.
E-mail: chryu23@gmail.com

²(Corresponding Author) Assistant Professor, Department of Statistics, Ewha Womans University, Seoul 03760, South Korea. E-mail: kimk@ewha.ac.kr

Variable Selection for Individualized Quantile Treatment Effects^{*}

Seohyeon Park¹, Minji Sung², Sangbum Chor³

Abstract

The primary objective of developing an optimal treatment policy is to ensure that patients receive the most suitable treatment at the appropriate time. Traditionally, research in this field has relied on mean-based models, such as linear regression and Cox models, to create personalized treatment strategies that aim to maximize average responses. However, a growing approach focuses on quantile-based methods, which prioritize maximizing conditional quantiles rather than average outcomes. In this paper, we propose a weighted quantile regression methodology to identify individualized treatment rules for observational studies. Our method addresses the challenge of misspecifying either the outcome or propensity score models, offering the advantage of double robustness. Additionally, we introduce an ‘Alternating direction method of multipliers (ADMM)’ algorithm tailored to high-dimensional observational data, utilizing a Broken Adaptive Ridge (BAR) penalty to identify key factors from extensive patient data. Through comprehensive simulation studies, we demonstrate the accuracy of our method in estimating decision-related regression parameters. Finally, we apply our methodology to data from the Prevention of Renal and Vascular End-stage Disease (PREVEND) study, illustrating its practical utility.

Keywords : Alternating direction method of multipliers, Broken adaptive ridge, Causal inference, Q-learning, Quantile regression.

^{*}This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00237435, 2022M3J6A1063595, 2022R1A2C1008514) and by grant from Korea University (K2305261).

¹Graduate Student, Department of Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: seohyeon20@korea.ac.kr

²Master’s graduate, Department of Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: minji913@korea.ac.kr

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

중국 30개 성(省)의 탄소 배출에 대한 군집단 분석

Shangyu Chen¹, 박성용²

요약

본 연구는 2000년부터 2021년까지 집계된 중국 30개 성의 탄소 배출량을 대상으로 그의 변동 및 원인을 분석하였다. 구체적으로 중국 각 성의 탄소 배출량의 유형에 따라 k-평균 군집 분석(k-means clustering analysis) 방법을 통해 각 성이 소속하는 탄소 배출 패턴을 정의 후 표본기간에서 해당 성의 탄소 배출량 동태적인 변동을 관찰하였다. 이 뿐만 아니라 본 연구는 마르코프 체인(Markov chain) 모형을 이용해서 이러한 변동에 대한 확률 전이행렬(probability transition matrix)을 제시하여 변동 원인도 규명하였다. 결과적으로는 광우동성, 저장성 등 지역은 청정 및 재생 에너지로 구성된 저(低) 탄소 배출 패턴으로 전환된 것으로 나타났다. 베이징과 충칭 등 지역은 소속한 패턴에 변동이 없는 것으로 관찰되었다. 반면 상하이, 푸젠성 등 지역은 탄소 배출량이 줄인 것으로 나타났으나 여전히 화석연료에 의존도가 높은 고(高) 탄소 배출 패턴에 벗어 나지 않는 것으로 확인되었다. 또한 본 연구에서는 다양한 경제적인 변수를 대입 후 이러한 변동 원인을 찾아내서 나타내는 시사점은 향후 탄소 배출량 변동 연구와 탄소 배출 저감 정책 입안에 도움이 될 것으로 사료된다.

주요 용어 : 중국 탄소 배출, k-평균 군집 분석, 탄소 배출 패턴, 마르코프 체인, 확률 전이행렬.

Change Point Detection을 이용한 대한민국 기온 변화 분석*

박현우¹, 신승준², 윤건수³

요약

전 세계적으로 지구 온난화는 가장 중요한 환경 문제 중 하나로 대두되고 있으며 특히 대한민국은 전 세계 평균보다 빠른 온도 상승을 겪고 있다. 이러한 기후 변화는 다양한 분야에 심각한 악영향을 미치고 있어 기후 변화에 적절히 대응하기 위한 연구의 필요성이 더욱 커지고 있다. 본 연구는 장기적인 온도 변화를 분석하고, 기온 변화가 어느 시점에서 어떤 양상으로 발생했는지 파악하기 위해 변화점 탐지 기법을 적용하였다. 이를 통해 대한민국의 장기적인 기온 변동 추세와 최근 기후 변화 경향을 파악하는 것이 본 연구의 목적이다. 연구를 위해 1969년부터 2023년까지의 한국 주요 8개 도시의 일일평균 기온 데이터를 사용하였으며 다양한 알고리즘을 적용하여 시계열 자료 내 변화점을 탐지 한 뒤 비교 분석하였다. 본 논문은 이러한 분석 결과를 바탕으로 한국의 기후 변화 양상에 대한 중요한 통찰을 제공하며, 지역 및 국가 차원의 적응 전략 수립의 필요성을 강조한다.

주요 용어 : 지구 온난화, 변화점 탐지, 시계열 자료.

*본 논문은 2022년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2022S1A5C2A03091302).

¹02841 대한민국 서울특별시 성북구 안암로 145 고려대학교 통계학과 박사과정.

E-mail: gusdn112000@korea.ac.kr

²(교신저자) 02841 대한민국 서울특별시 성북구 안암로 145 고려대학교 통계학과 교수.

E-mail: sjshin@korea.ac.kr

³02841 대한민국 서울특별시 성북구 안암로 145 고려대학교 행정학과 교수. E-mail: gsmagic@korea.ac.kr

XGBoost, LightGBM을 활용한 시정 예측 모델*

신봉균¹, 서혜선²

요약

본 연구는 기상 특성과 대기질 데이터를 통합한 시정(가시거리) 예측 모델에 관한 것이다. XGBoost(extreme gradient boosting)와 LightGBM(light gradient boosting) 기반의 머신러닝 앙상블(ensemble) 기법을 활용하여, 기상청의 중관기상관측 ASOS(automated synoptic observing system)와 에어코리아의 지상대기질 데이터(2018년-2021년)를 종합적으로 적용하였다. 연구의 분석 범위는 서울, 인천, 대전, 광주, 대구, 부산 등 6개 주요 도시로 설정하였다. 데이터 전처리 과정에서 결측치는 선형보간법과 인근 관측소 데이터를 활용하여 적용하였으며, 지면온도-대기온도 차이, 시간의 주기성을 고려한 변수들을 추가하였다. 시정 거리는 0~1km(class 0), 1km~5km(class 1), 5km 이상(class 2)으로 구분하였다. 시계열 데이터의 특성을 고려하여 2018년부터 2020년까지의 데이터를 훈련 데이터로, 2021년 데이터를 테스트 데이터로 사용하였으며, XGBoost와 LightGBM 모델의 앙상블을 통해 최종 예측 모델을 구현하였다. Permutation importance 기법을 통한 변수 선택과 계층적 이진 분류 모델 구조를 적용하였으며, 베이지안 최적화를 통해 모델을 최적화하였다. 최종 모델은 지역별로 0.24에서 0.51 사이의 CSI(critical success index) 점수를 달성하였으며, 이는 머신러닝 기법을 활용한 시정 예측의 실현 가능성과 잠재력을 입증하였다.

주요 용어 : XGBoost, LightGBM, ASOS, 대기질, 시정예측.

1. 서론

기상현상은 사회에 직간접적으로 영향을 미치는 중요한 요소이다. 특히 시정은 교통, 항공, 해상 분야에서 안전과 운영에 결정적인 영향을 미치는 핵심 요인이다. 시정 저하는 교통사고, 항공기 운항 지연 및 취소, 해상 사고 등 다양한 사회경제적 위험을 초래할 수 있어, 정확한 시정 예측은 매우 중요한 과제로 대두된다.

시정에 영향을 미치는 요인은 복잡적이며, 안개, 강수, 대기질 등 다양한 기상 및 환경 조건에 의해 결정된다. 특히 대기 중 수증기 응결, 미세먼지 농도, 기온과 습도 변화 등은 시정에 직접적인 영향을 미치는 주요 변수들이다.

이러한 배경에서 본 연구는 중관기상관측(ASOS) 데이터와 지상 대기질 데이터를 종합적으로 분석하여, 보다 정확하고 신뢰성 높은 시정 예측 모델을 개발하고자 한다. 머신러닝 기반의 앙상블 기법을 활용함으로써, 기존 예측 방식의 한계를 극복하고 시정 변화의 복잡성을 더욱 정밀하게 진단하고자 한다.

¹11159 대한민국 경기도 포천시 호국로 1007, 대전대학교 AI 빅데이터학과 학부재학생. E-mail: 20191624@daejin.ac.kr
²(교신저자)11159 대한민국 경기도 포천시 호국로 1007, 대전대학교 AI융합학부 부교수. E-mail: jako403@daejin.ac.kr

2. 연구 방법론

2.1. 데이터 수집

본 연구에서는 시정 예측을 위해 기상청의 종관기상관측(ASOS)과 에어코리아의 지상대기질 데이터(2018년-2021년)를 수집하였다. 대기질 데이터의 관측 지점은 전국 660개 지점이며, 종관기상관측 데이터의 관측 지점은 98개 지점이 있다. 연구 대상 지역으로는 서울, 인천, 대전, 광주, 대구, 부산 등 6개 주요 도시를 선정하였다.

2.2. 데이터 전처리

대기질 데이터의 관측 지점은 전국 660개 지점이며, 종관기상관측 데이터의 관측 지점은 98개 지점이다. 기상청과 에어코리아는 독립적으로 데이터를 수집하기 때문에 두 데이터의 관측 지점이 상이하여 데이터 병합이 필요하였다.

ASOS 데이터(00-23시)와 대기질 데이터(01-24시)의 시간 체계를 통일하였으며, 데이터 병합을 위해 종관기상관측 데이터를 메인 데이터로 설정하고, 각 관측 지점에서 지리적으로 가장 가까운 대기질 데이터 관측 지점을 하버사인(haversine) 공식을 활용하여 매핑하였다. 최종적으로 서울, 인천, 대전, 광주, 대구, 부산 등 6개 주요 도시의 데이터를 분석 대상으로 선정하였다.

결측치 처리는 데이터의 시계열적 특성을 고려하여 단계적 접근 방식을 채택하였다. 1시간 단위의 짧은 결측 구간에는 선형보간법을, 장기간 결측 구간에는 인근 관측소의 데이터를 활용하였다. 이상치의 경우, 기상청 데이터는 이미 품질관리가 수행된 상태였으나, 일부 지역의 데이터에서 발견된 비정상적 관측값은 제거 처리하였다.

또한 모델의 예측 성능 향상을 위해 세 가지 주요 변수를 생성하였다. 복사안개 발생 가능성을 고려한 지면온도-대기온도 차이 변수, 시간의 주기성을 반영하기 위하여 변환된 시간 변수, 그리고 시정 거리를 0~1km(class 0), 1km~5km(class 1), 5km 이상(class 2)으로 구분한 class 변수를 추가하였다.

2.3 연구 모형

본 연구에서는 XGBoost와 LightGBM 두 가지 부스팅 계열의 모델을 사용하였다. 부스팅은 여러 개의 약한 학습기(weak learner)를 순차적으로 학습시켜 강한 학습기(strong learner)를 만드는 앙상블 기법이다, XGBoost는 의사결정나무를 기반으로 하여 이전 모델의 오차를 보완하면서 점진적으로 예측 성능을 향상시키는 알고리즘이며, LightGBM은 XGBoost와 유사한 원리로 작동하지만 트리 생성 방식을 최적화하여 더 빠른 학습 속도를 제공한다. 또한, XGBoost와 LightGBM 모두 결측치에 대해 학습이 가능하다.

Figure 1과 같이 hm(상대습도), low_cloudbase(최저 운고), solarRad(일사량) 와 같은 변수들의 시정 등급별 데이터 분포 분석 결과, class 2(시정: 5000m 이상)에서만 뚜렷한 차이가 확인되어 계층적인 분류 모델 구조를 채택하였다. Figure 2와 같이 지역별로 첫 번째 단계에서는 class 2와 나머지 클래스를 구분하고, 두 번째 단계에서 남은 클래스들을 분류하는 방식으로 구성하였다. 이러한 계층적 구조는 분류의 정확성을 높이는 데 효과적이었다.

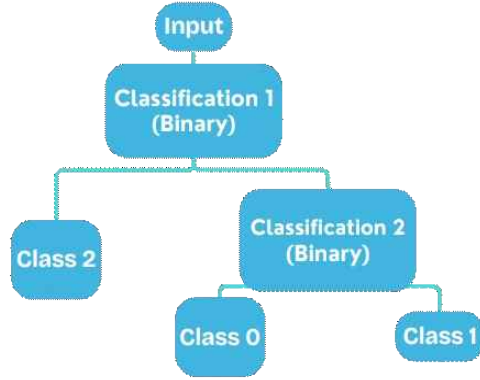


Figure 1. Box plots of meteorological variables

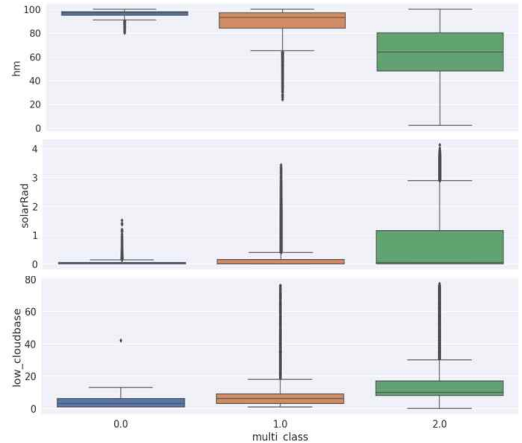


Figure 2. Hierarchical binary classification model structure

3. 실증 분석

3.1. 데이터 분할 및 학습

모델 학습을 위해 시계열 데이터의 특성을 고려하여 Table 1과 같이 2018년부터 2020년까지의 데이터를 훈련 데이터로, 2021년 데이터를 테스트 데이터로 사용하였다. 각 fold에서 XGBoost와 LightGBM 모델을 독립적으로 학습한 후, 두 모델의 예측 결과를 소프트 보팅(soft voting) 방식으로 앙상블하여 최종 예측 모델을 구현하였다.

Table 1. Three-fold cross-validation splits for temporal data spanning 2018-2020

Fold	Train	Validation
1 Fold	2018, 2019	2020
2 Fold	2018, 2020	2019
3 Fold	2019, 2020	2018

3.2. 모델 최적화 및 결과

XGBoost와 LightGBM 모델 각각에 대해 베이저안 최적화(Bayesian Optimization) 기법을 적용하여 최적의 파라미터 조합을 탐색하였다. Table 2와 같이 광주 지역을 제외한 나머지 지역에서는 최적화 이후 F1 score가 감소하지만 재현율(recall)이 증가한 것을 볼 수 있다. 이는 시정 예측에서 중요한 낮은 시정(5km 미만) 상황에 대한 탐지 능력이 향상되었음을 의미한다. Table 3과 같이 지역별로 모델의 최종 성능은 0.24에서 0.51 사이의 CSI 점수를 달성하였다. 이는 Table 4에서 확인할 수 있는 기상청(korea meteorological administration)의 황사/연무 예측 모델들과 비교했을 때 경쟁력 있는 수준임을 알 수 있다. 기상청의 KIM-ADAM3 모델이 보여주는 0.23의 CSI 점수와 UM-ADAM3 모델의 0.41 CSI 점수를 고려할 때, 본 연구의 모델이 달성한 성능은 현업 모델과 비교하여 실용화 가능성을 보여주었다고 할 수 있다.

Table 2. Comparison of F1, recall scores before and after optimization by region

Region	Classification 1				Classification 2			
	pre_optimization		post_optimization		pre_optimization		post_optimization	
	F1	recall	F1	recall	F1	recall	F1	recall
Seoul	0.566	0.814	0.539	0.868				
Incheon	0.716	0.787	0.706	0.875				
Dajeon	0.6	0.686	0.595	0.824				
Daegu	0.465	0.52	0.428	0.597	0.966	0.993	0.966	0.993
Gwangju	0.662	0.623	0.689	0.761				
Busan	0.713	0.77	0.709	0.843				

Table 3. Comparison of CSI and accuracy by region

Region	CSI	Accuracy
Seoul	0.359	0.915
Incheon	0.243	0.878
Dajeon	0.401	0.926
Daegu	0.276	0.982
Gwangju	0.511	0.946
Busan	0.494	0.973

Table 4. KMA yellow dust and haze prediction models

	KIM-ADAM3	UM-ADAM3
accuracy	0.96	0.98
CSI	0.23	0.41

4. 결론

본 연구를 통해 몇 가지 주목할 만한 결과를 확인하였다. 첫째, 계층적 이진 분류 접근 방식은 시정 거리의 불균형한 클래스 분포 문제를 효과적으로 해결하였다. 특히 5km 이상의 시정 거리(class 2)를 먼저 분류한 후, 나머지 클래스를 구분하는 방식이 전반적인 예측 성능 향상에 기여하였다. 둘째, 베이지안 최적화를 통한 모델 튜닝 과정에서 대부분의 지역에서 F1 score는 다소 감소했으나 재현율이 크게 향상되었다. 이는 낮은 시정 상황에 대한 탐지 능력이 개선되었음을 의미하며, 실제 운영 환경에서 더욱 유용하게 활용될 수 있음을 시사한다. 광주(CSI: 0.511)와 부산(CSI: 0.494) 지역에서 상대적으로 높은 예측 성능을 보인 반면, 인천(CSI: 0.243)과 대구(CSI: 0.276) 지역에서는 다소 낮은 성능을 보였다. 이러한 지역별 성능 차이는 각 지역의 지형적 특성, 기상 조건의 복잡성, 그리고 대기질 관측소의 분포 등 다양한 요인에 기인한 것으로 분석된다. 본 연구에서 구현된 시정 예측 모델은 기존의 통계적 접근이나 수치 예보 모델의 한계를 극복하고, 머신러닝 기반의 새로운 가능성을 제시하였다. 특히 XGBoost와 LightGBM의 앙상블 접근은 개별 모델의 장점을 결합하여 더욱 안정적이고 정확한 예측을 가능하게 하였다. 이는 교통, 항공, 해상 분야의 안전 성과 운영 효율성 향상에 실질적으로 기여할 수 있을 것으로 기대된다. 향후 연구에서는 더 많은 기상 변수의 도입, 시공간적 특성을 고려한 심층 학습 모델의 적용, 그리고 실시간 예측 시스템으로의 확장 등을 통해 모델의 성능을 더욱 개선할 수 있을 것으로 기대된다.

Visibility Prediction Model Using XGBoost and LightGBM

*BongGyun Shin*¹, *Hyesun Sul*²

Abstract

This study developed a visibility prediction model integrating meteorological characteristics and air quality data. Using machine learning ensemble techniques based on XGBoost (eXtreme Gradient Boosting) and LightGBM (Light Gradient Boosting Machine), we comprehensively analyzed data from the Korea Meteorological Administration's ASOS (Automated Synoptic Observing System) and Air Korea's ground-level air quality measurements (2018-2021). The primary methodology was implemented on data from six major cities: Seoul, Incheon, Daejeon, Gwangju, Daegu, and Busan. During data preprocessing, missing values were addressed using linear interpolation and data from nearby observation stations. Additional variables were reflected considering the difference between ground and air temperatures, as well as time periodicity. Visibility distances were categorized into three classes: 0-1km (class 0), 1-5km (class 1), and above 5km (class 2). Considering the time series characteristics, data from 2018 to 2020 were used as training data, while 2021 data served as test data. The final prediction model was constructed through an ensemble of XGBoost and LightGBM models. The methodology incorporated variable selection through permutation importance technique and implemented a hierarchical binary classification model structure, optimized through Bayesian optimization. The final model achieved CSI (Critical Success Index) scores ranging from 0.24 to 0.51 across different regions, demonstrating the feasibility and potential of machine learning techniques in visibility prediction.

Keywords : XGBoost, LightGBM, ASOS, air quality, Visibility Prediction.

¹Undergraduate student, Department of AI & Big Data, Daejin University, 1007 Hoguk-ro, Pocheon-si, Gyeonggi-do 11159, Republic of Korea. E-mail: 20191624@daejin.ac.kr

²(Corresponding Author) Associate Professor, Department of Artificial Intelligence Convergence Engineering, Daejin University, 1007 Hoguk-ro, Pocheon-si, Gyeonggi-do 11159, Republic of Korea. E-mail: jako403@daejin.ac.kr

Bayesian Multiple Change-point Estimation for Hazard in Weibull Distributions

*Ah-Rim Joo*¹, *Sooyoung Cheon*², *Jaehee Kim*³

Abstract

A Bayesian approach to the problem of hazard change with unknown multiple change-points is developed using informative and noninformative priors for survival data. Due to their flexibility, Weibull distributions are suitable for parametric modeling of time-to-event data and they have been widely used in scientific applications such as in survival analysis, reliability and industrial engineering, hydrology and etc.. For the Weibull distribution, piecewise constant hazard is considered with change-point estimation. The two-parameter Weibull distribution is a particularly difficult with related prior distribution because it requires a two-dimensional joint prior for Weibull parameters. The Stochastic Gradient Langevin Dynamics is implemented for rapid and efficient search for change-points with the posterior distributions. A major feature of the proposed approach is that there is no necessity for the number of change-points resulting hazard function. The performance of the proposed estimator is checked via simulation. As a real data application, Leukemia data and Diabetic data are analyzed.

Keywords : weibull distribution; hazard' multiple change-point; SGLD; survival analysis.

¹Department of Economics and Statics, Korea University.

²Division of Big Data Science, Korea University.

³Department of Statistics, Duksung Women's University.

Role of Maternal Depression and Network Trajectories on Internet Addiction in Korean Youth^{*}

Juan Kim¹, Sangjin Lee², Jinhyun Kim³, Min Jin Ha⁴

Abstract

Internet addiction (IA) in children and adolescents has become a growing concern, particularly in South Korea, where internet usage rates are highest globally. Previous studies have identified a relationship between parental mental health and IA in children, but the long-term effects of maternal depression on IA have not been fully explored. This study aims to investigate the association between maternal depression, both during the peripartum period and in subsequent years, and IA in children using data from the Panel Study on Korean Children (PSKC), a nationally representative longitudinal survey. The study also examines how other family and child-related factors, such as parents' employment status and psychiatric symptoms, interact with maternal depression to influence IA. The analysis included 652 participants from the PSKC. In PSKC, IA was assessed using the Korean Internet Addiction Scale (K-scale), while maternal depression was measured using the Kessler Psychological Distress Scale (K6). Data from 2008, 2017, and 2019 were analyzed using longitudinal network modeling and generalized estimating equations (GEE). The results showed that children of mothers who experienced peripartum depression had a higher risk of developing IA compared to children whose mothers did not have peripartum depression. Furthermore, children whose mothers were currently depressed also had a higher risk of IA. Network modeling revealed that maternal depression and IA were interconnected over time, with maternal depression in 2017 predicting higher IA in children by 2019. These findings highlight the significant role of maternal mental health in shaping children's behavioral outcomes, particularly regarding their internet usage. The persistence of IA across years suggests that early intervention for maternal depression could mitigate the long-term risk of IA in children. This study emphasizes the importance of comprehensive mental health assessments for both mothers and children to prevent the onset and progression of IA.

Keywords: Network modeling, Dependency network, longitudinal study.

^{*}This work was supported by National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) [No. 2022R1A2C1091488]

¹Department of Biostatistics and Computing, Yonsei University Graduate School. zxs132@yuhs.ac

²Department of Statistics, Seoul National University, Seoul, Republic of Korea.

³Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea.

⁴Assistant professor, Department of Health Informatics and Biostatistics, Graduate School of Public Health, Yonsei University. E-mail: mjha@yuhs.ac

Ridge regression splines with knot selection via ℓ_0 penalty*

*Eun-Ji Lee*¹, *Jae-Hwan Jhong*²

Abstract

The Ridge regression model that considers the ℓ_2 -norm penalty is a model that can alleviate multicollinearity problems by slightly reducing all coefficients in high-dimensional data. When the ℓ_0 -norm penalty is considered in a linear regression model, variable selection is explicitly performed to build a model using the fewest variables, and a sparse model is created to facilitate interpretation and prevent overfitting. From a nonparametric regression perspective, the knot selection problem can be solved using a norm penalty with a regression model using a combined truncated power spline. We propose estimator based ℓ_2 -norm and ℓ_0 -norm penalties. We estimate the coefficients of the regression model with a coordinate descent algorithm. We evaluate the performance of ℓ_2 -norm and ℓ_0 -norm penalty-based estimation through three simulations and we present numerical studies through real data analysis.

Keywords : Best subset selection; coordinate descent algorithm; knot selection; ridge penalty; truncated power splines.

*This research was supported by Chungbuk National University Korea National University Development Project (2023).

¹MS-Ph.D. integrated course student, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea. E-mail: ejkrsu97@gmail.com

²(Corresponding Author) Associate Professor, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea. E-mail: jjh25@chungbuk.ac.kr

Zero inflated high dimensional compositional data with DeepInsight^{*}

Jeseok Lee¹, Byungwon Kim²

Abstract

Through the Human Microbiome Project, research on human-associated microbiomes has been conducted in various fields. New sequencing techniques such as Next Generation Sequencing (NGS) and High-Throughput Sequencing (HTS) have enabled the inclusion of a wide range of features of the microbiome. These advancements have also contributed to the development of numerical proxies like Operational Taxonomic Units (OTUs) and Amplicon Sequence Variants (ASVs). Studies involving such microbiome data often encounter zero-inflated and high-dimensional problems. Based on the need to address these two issues and the recent emphasis on compositional interpretation of microbiome data, we conducted our research. To solve the zero-inflated problem in compositional microbiome data, we transformed the data onto the surface of a hypersphere using a square root transformation. Then, to solve the high-dimensional problem, we modified DeepInsight, an image-generating method using Convolutional Neural Networks (CNNs), to fit the hypersphere space. Furthermore, to resolve the common issue of distinguishing between true zero values and fake zero values in zero-inflated images, we added a very small value to the true zero values. We validated our approach using pediatric inflammatory bowel disease (IBD) fecal sample data and achieved an area under the receiver operating characteristic curve (AUC) value of 0.847, which is higher than the previous study's result of 0.83.

Keywords : Non-euclidean data, Deep Learning, Microbiome data.

^{*}This work was in part supported by the Korea Research Foundations, Korea, under grant RS-2023-00213626.

¹Doctor program, Department of Statistics, Kyungpook National University, Daegu 41566, Korea.

²(Corresponding Author) Assistant professor, Department of Statistics, Kyungpook National University, Daegu 41566, Korea. E-mail: byungwonkim@knu.ac.kr

Prediction of Delisted Companies Using PSM-Based Machine Learning and Deep Learning Models *

Yeonju Kim¹, Byungwon Kim²

Abstract

The recent increase in delisted companies has raised concerns and caused losses for investors, creating a growing need for more sophisticated models to predict delisted and listed companies. However, domestic studies mainly rely on simple statistical methods, such as logistic regression or t-tests, and often use artificial or random matching methods. These approaches fail to ensure statistically meaningful matching and do not adequately address the issue of data imbalance. This study moves beyond simple statistical analysis by applying Propensity Score Matching (PSM) to achieve statistically significant matching, and leveraging machine learning and deep learning models to improve the accuracy in predictions of delisted companies. The analysis used financial data from companies delisted from the KOSDAQ and KOSPI between 2005 and 2019, excluding voluntarily delisted firms. The results show that the PSM-applied data achieves higher F1 scores compared to non-matched data, and deep learning and some machine learning models outperformed traditional logistic regression. This study demonstrates the potential for applying advanced techniques in delisting prediction, allowing for more precise analysis of complex and imbalanced data.

Keywords : Financial data analysis, Propensity Score Matching, Delisting Prediction, Deep learning, Machine learning.

*This work was in part supported by the Korea Research Foundations, Korea, under grant RS-2023-00213626.

¹Master program, Department of Statistics, Kyungpook National University, Daegu 41566, Korea.

²(Corresponding Author) Assistant professor, Department of Statistics, Kyungpook National University, Daegu 41566, Korea. E-mail: byungwonkim@knu.ac.kr

문화재정지출이 지역경제에 미치는 영향 분석

오주현¹

요약

본 연구는 한국의 지역별 문화재정지출이 지역경제에 미치는 영향을 분석한다. 한국의 지역별 문화재정지출이 지역경제에 미치는 영향을 분석한 연구는 미흡한 실정이며, 지역경제의 이질성을 고려하여 문화재정지출의 효과를 분석한 연구는 이루어지지 않았다. 따라서, 본 연구는 패널 분위회귀모형(panel quantile regression model)을 이용하여 지역 소득 및 특성별 문화재정지출의 효과성을 검증한다는 점에서 선행연구와 차별성이 있다. 특히 설명변수의 내생성을 고려하여 문화재정지출과 경제 성장 간의 관계를 보다 엄밀하게 규명하고자 한다. 본 연구는 지역 경제의 이질성을 반영함으로써 문화재정지출이 지역 경제 성장에 미치는 다양한 효과를 세밀하게 분석하며, 이를 통해 정책적 의사결정에 필요한 실증적 근거를 제시한다는 점에서 의의가 있다.

주요 용어 : 지역경제, 문화재정지출, 이질성, 분위수회귀모형, 문화격차.

담배 제품별 교차 가격탄력성 분석

이성규¹, 김진영², 최은철³

요약

최근 다양한 담배제품의 출시와 함께 흡연자들은 쉐련, 쉐련형 전자담배, 액상형 전자담배 등 여러 제품을 상황에 따라 선택하며 사용하는 경향이 뚜렷해지고 있다. 이러한 변화는 니코틴 섭취 패턴을 복잡하게 만들며, 총 니코틴 노출량 증가로 이어지고 있다. 특히 금연을 목표로 하기 보다는 덜 유해하다고 여겨지는 제품으로 전환하는 것이 더 쉬운 선택으로 자리 잡으면서, 흡연자들의 제품 선택 행태가 국민 건강 보호를 위한 정부의 보건정책에 부정적인 영향을 미칠 가능성이 제기된다. 더불어, 담배제품의 다양화는 차등 과세에 대한 논의를 심화시키고 있다. 각 제품의 유해성을 기준으로 차별적인 세금 부과 필요성이 담배업계를 중심으로 요구되고 있으며, 특히 합성니코틴을 사용하는 액상형 전자담배의 경우 담배사업법상 담배로 분류되지 않아 현행 법에 따라 제세부담금이 부과되지 않고 있다. 이로 인해 규제의 공백이 발생하고 있으며, 합성니코틴 제품의 시장 유통과 사용이 빠르게 확산되고 있다. 이에 따라, 변화하는 담배시장과 복합적인 흡연 행태에 맞춘 합리적인 과세 방안을 마련하는 것이 시급하다. 이에 본 연구는 우선적으로 쉐련, 쉐련형 전자담배, 액상형 전자담배 간의 교차 가격 탄력성을 실증적으로 분석한다.

주요 용어 : 교차 가격탄력성, 수요 분석, 흡연.

¹13451 경기도 성남시 수정구 옛골로 34, 한국담배규제연구교육센터 센터장. Email: wwwvince77@gmail.com

²13451 경기도 성남시 수정구 옛골로 34, 한국담배규제연구교육센터 선임연구위원. Email: jy9651@daum.net

³(교신저자) 49112 부산광역시 영도구 태종로 727, 한국해양대학교 국제무역경제학부 조교수.

Email: ecchoi@kmou.ac.kr

Development and evaluation of Artificial Intelligence Techniques to Revolutionize the Marketing Strategies^{*}

Minho Sun¹, Seung Woo Kim¹, Jai Woo Lee²

Artificial intelligence techniques are innovatively transforming the field of business, attracting growing interest from researchers. In this study, we review marketing strategies and develop artificial intelligence techniques to build large-scale data analysis for marketing communications. We provide an overview of artificial intelligence techniques to process large-scale marketing data, presenting flexible big data analytics with promising predictive performance. An integrative analytic framework is devised to extract insights from high-dimensional network data to represent inferential analyses. The aim of this study is to build more accessible network analysis based on fundamental data requirements with large business communication networks to validate network measures and theories in computation that have been utilized in marketing studies. The proposed methods can be applied to analyze complex and dynamic associations of marketing big data with various features.

Keywords : Artificial Intelligence, Big Data, Large-Scale Computing, Network Analysis, Marketing.

^{*}This article is financially supported by the 2024 College of Public Policy at Korea University.

¹Department of Big Data Science, Korea University, Sejong.

²(Corresponding Author) Department of Big Data Science, Korea University, Sejong.

Email: jaiwoolee@korea.ac.kr

인구 변화와 가족 관계의 변화: 노인 지원 및 장기 요양 시스템을 위한 대안

이지나¹, 황명진²

요 약

한국은 급격한 인구 고령화와 낮은 출산율로 인해 가족 구조와 가치관에 큰 변화를 겪고 있다. 이러한 변화는 노인 지원 및 장기 요양 시스템에 막대한 영향을 미치고 있으며, 이에 따라 가족 중심의 전통적 돌봄 방식은 한계를 보이고 있다. 본 연구는 가족의 역할 변화와 현대 사회에서 개인주의 가치관이 확산되는 상황 속에서, 한국 사회가 어떻게 새로운 돌봄 체계를 모색할 수 있을지를 고찰한다. 또한 커뮤니티 기반 돌봄 모델과 사회보장 시스템 강화를 포함한 지속 가능한 돌봄 체계의 필요성을 논의한다. 황명진(2020), 이지나와 황명진(2022), 박소영과 황명진(2022) 등 선행 연구들을 바탕으로, 고령화 사회에서의 돌봄 정책과 가족 구조 변화에 대한 분석을 통해 보다 효과적인 노인 지원 대안을 제시하고자 한다.

주요 용어: 인구 변화, 가족 관계, 노인 돌봄, 장기 요양, 사회 보장

Reference

- 강동주, 김윤섭, 황명진. (2023). 장기요양시설 입소자의 시설 적응 영향요인 연구. *공공사회연구*, 2023(8), 105.
- 유염향, 최승주, 이지나. (2023). 코로나 19 로 인한 중국인 유학생의 학업 적응과정 탐색에 관한 질적 연구. *공공사회연구*, 2023(5), 108.
- 박소영, 황명진. (2022). 직장 기혼 여성의 출산 의향에 영향을 미치는 요인 연구. *공공사회연구*, 2022(11), 52.
- 이지나, 황명진. (2022). 사회복지 담론으로서의 기본소득. *공공사회연구*, 2022(5), 875.
- 황명진. (2020). 인구고령 사회의 노인복지와 커뮤니티 케어. *공공사회연구*, 2020(5), 3, 837.
- 도지원, 황명진. (2020). 제 3 국 출생 북한이탈청소년의 가족경험 변화에 관한 질적 사례연구. *사회과학연구*, 2020(1), 700.
- 이강수, 황명진. (2021). 커뮤니티케어 담론으로서의 강점관점 연구. *공공사회연구*, 2021(2), 565.
- 황명진. (2017). 고령화 시대 중장년 및 노년층의 의료비 지출 결정 요인. *공공사회연구*, 2017(5), 117.
- 박소영, 황명진. (2017). 한국 청년층의 주택구입 의향에 대한 영향요인 연구. *사회과학연구*, 2017(7), 124.
- 황명진. (2015). 한국의 공적연금제도과 노인 소득 유지. *한국사회학회 사회학대회 논문집*, 2015(12), 54.
- Nobutane Hanayama, 황명진. (2014). 한국과 일본의 여성 노동 참여 비교 연구. *한국인구학회지*, 2014(6).

¹고려대학교 사회복지학과

²고려대학교 공공사회통일외교학부

DEU 핵심역량과 학부교육 실태의 연계성 분석*

김보성¹, 이순화², 김경희³, 강주리⁴, 이영창⁵

요약

본 연구는 대학의 고유한 교육 방향을 위해 설정한 핵심역량이 역량기반 교육과정을 이수한 학생들의 학부교육 실태 결과와 어떠한 관계를 맺고 있는지 살펴보기 위해서 핵심역량 수준과 학부교육 실태 결과와의 연계성을 분석하였다. 이를 위해, D 대학 학생을 대상으로 DEU-CCDS (핵심역량 척도) 문항과 K-NSSE(학부교육 실태조사) 문항을 활용하여 응답하도록 하였다. 그 결과, D 대학의 재학생들의 핵심역량 수준과 이들의 학부교육 실태 수준은 전반적으로 유의한 관계를 보이는 것으로 나타났다. 한편, IPA 분석을 통해 학부교육 실태의 영역에 따라 지속-개선 영역의 핵심역량 구성요소를 도출한 결과, 학부교육 실태 수준을 잘 설명하는 핵심역량 구성요소들은 유지강화 및 점진개선 영역에 주로 분포하였으며, 일부는 중점개선 영역에 분포하였다. 이는 학부교육의 질적 수준을 유지하기 위해서 적절한 핵심역량 수준을 강화할 수 있도록 하는 교과 및 비교과 활동의 지속적인 개발 및 개선이 필요함을 시사한다.

주요 용어 : 핵심역량, 학부교육 실태, DEU-CCDS, K-NSSE, IPA.

*본 연구는 ‘학생역량 관리를 위한 재학생 역량 관련 평가 결과의 연계성 분석 연구보고서’의 내용을 일부 축약한 것임.

¹(교신저자) 47340 부산광역시 부산진구 엄광로 176, 동의대학교 철학인문교양학부 교수 및 교육혁신처 교육성과관리센터 소장. E-mail: bkim@deu.ac.kr

²47340 부산광역시 부산진구 엄광로 176, 동의대학교 교육혁신처 교육성과관리센터 조교수.
E-mail: naturalhwa@deu.ac.kr

³47340 부산광역시 부산진구 엄광로 176, 동의대학교 교육혁신처 교육성과관리센터 조교수.
E-mail: keung87@deu.ac.kr

⁴47340 부산광역시 부산진구 엄광로 176, 동의대학교 교육혁신처 교수학습개발센터 소장.
E-mail: misohanip@deu.ac.kr

⁵34134 대전광역시 유성구 대학로 99, 충남대학교 심리학과 강사. E-mail: kafka13@cnu.ac.kr

아동학대 발굴 체계 강화를 위한 아동 건강정보 연계 분석

김선월¹, 박준범²

요 약

우리나라는 의무적으로 국민건강보험에 가입해야 하는 시스템을 갖추고 있다. 세계적으로 높은 도입률을 보이는 전자의무기록(EMR) 시스템을 기반으로 5천 명 이상 국민의 의료 이용 명세가 축적되고 있어 자료가 갖는 대표성이 매우 높다. 또한 국민 1인당 의사 외래 진료 횟수가 17.5회(OECD 평균 6.4회)로 매우 높고, 국가 건강검진 수검률도 높은 수준으로 건강 수준을 관찰하는 데 우수한 자료원으로 평가받는다. 2023년 아동학대 신고접수 48,522건 중 94.3%인 45,771건은 아동학대 의심사례로 나타났다. 이는 매년 증가하는 추세를 보이고 있다. 아동학대는 시간이 지날수록 악화되는 양상을 빚기 때문에 조기 발견이 가장 중요한데, 현재의 발굴체계에는 많은 한계가 있다. 본 연구에서는 아동의 영유아 검진 정보를 포함한 건강정보를 연계하여 아동학대의 위험도 분석을 수행하였다. 아동학대 판정 경험이 있는 학대판정군에서는 1세의 뇌손상, 아동청소년의 행동 및 정서, 정신지체, 성폭력, 섭식장애 등과 관련된 진단의 비율이 대조군에 비해 다소 높게 나타났다. 영유아 검진에서는 학대판정군의 신체측정 및 발달평가에서의 심화평가 판정 비율이 대조군에 비해 다소 높게 나타났다. 심화평가 판정 횟수가 늘어날수록 학대판정군의 비율은 증가하였다. 건강보험 자격 및 보험료, 진료정보 등의 일부 변수들을 이용한 로지스틱 회귀 분류에서 학대아동 분류의 precision과 recall은 0.67과 0.59로 나타났다. 영유아검진의 검진차수별 정밀평가 필요여부와 필요 판정 횟수를 이용한 로지스틱회귀 분류에서 precision과 recall이 0.63과 0.71로 나타났다. 분석 설계와 분석법 활용 한계로 인해 분류모델의 정확도가 높은 수준은 아니지만, 아동 건강정보 활용의 가능성을 확인한 최초의 연구로서 의미가 있다.

¹(교신저자) 04933 서울 광진구 능동로 400, 한국사회보장정보원 사회보장정보연구소 연구위원.

E-mail: sunworl@gmail.com

² 04401 서울 용산구 대사관로 59, 순천향대학병원 응급의학과 부교수. E-mail: jesumania@gmail.com

건강보험청구자료 활용을 위한 환례정의 마련 연구

김연희¹, 이돈형¹

요약

우리나라는 급속한 인구고령화와 질병부담 증가로 건강증진의 필요성이 커지고 있으며, 이를 해결하기 위한 정책적 대응이 요구되는 상황입니다. 특히 제5차 국민건강증진종합계획(HP2030)은 건강수명의 연장과 건강형평성 제고를 목표로 하고 있으며, 이를 실현하기 위한 중요지표로 건강수명이 산출되고 있습니다. 건강수명 산출의 중요한 기초자료인 건강보험청구자료는 유병률 및 발생률을 정확히 파악하는 것이 중요합니다. 본 연구의 목적으로 건강보험청구자료를 활용하여 질환별 유병자 및 발생자를 파악하기 위한 환례정의를 제시하는 것이며, 이를 통해 건강수명 산출의 신뢰성을 높이고 질병부담 분석에 기여하고자 합니다. 본 연구의 방법으로 260개 질환을 대상으로 문헌고찰과 전문가 조사를 통해 환례정의 기준을 설정하고, 전문가 의견을 바탕으로 각 질환에 적합한 임상적 기준을 도출한 후 이를 건강보험청구자료에 적용 가능한 방식으로 정리합니다. 연구를 통해 제시된 환례정의는 건강보험 데이터를 활용한 질병분석의 정확도를 높이고, 건강수명 산출 및 질병부담 연구 분석을 위한 기준을 마련하고, 향후 관련 정책의 개선 및 후속 연구에 기초 자료로서 중요한 역할을 할 것입니다.

주요 용어 : 건강수명, 환례정의, 국민건강증진종합계획.

Applications of Response Dimension Reduction in Large p-Small n Problems^{*}

*Minjee Kim*¹, *Jae Keun Yoo*²

Abstract

The goal of this paper is to show how multivariate regression analysis with high-dimensional responses is facilitated by the response dimension reduction. Multivariate regression, characterized by multi-dimensional response variables, is increasingly prevalent across diverse fields such as repeated measures, longitudinal studies, and functional data analysis. One of the key challenges in analyzing such data is managing the response dimensions, which can complicate the analysis due to an exponential increase in the number of parameters. Although response dimension reduction methods are developed, there is no practically useful illustration for various types of data such as so-called large p-small n data. This paper aims to fill this gap by showcasing how response dimension reduction can enhance the analysis of high-dimensional response data, thereby providing significant assistance to statistical practitioners and contributing to advancements in multiple scientific domains.

Keywords : High-dimensional data analysis; Large p-small n data; Model-based reduction; Multivariate regression; Response dimension reduction.

^{*}For Jae Keun Yoo and Minjee Kim, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (RS-2023-00240564 and RS-2023-00217022)

¹Department of Statistics, Ewha Womans University.

²(Corresponding Author) Professor, Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea. E-mail: peter.yoo@ewha.ac.kr

Projective Resampling with the Minimum Average Variance Estimator: Sufficiently Reducing Dimension with Multivariate Response

Seoyoung Kim¹, Jae Keun Yoo¹

Abstract

This paper addresses the challenge of dimension reduction in datasets where both the predictor and response variables are multivariate. We introduce a novel approach that integrates the Minimum Average Variance Estimator (MAVE) with the Projective Resampling (PR) method to render its applicability in multivariate response settings. By applying MAVE within the framework of Projective Resampling, we achieve an efficient dimension reduction, ensuring the exhaustive recovery of the central mean subspaces and eradicating strong assumptions on the distribution of predictors. The proposed methodology is validated through a series of simulations to evaluate its effectiveness and robustness in various multivariate contexts. Further refinements were made to improve the computational performance as well.

Keywords : Dimension Reduction , Multivariate nonlinear regression, Central mean subspace, Average derivative estimation.

¹Department of Statistics, Ewha Womans University.

Multivariate Response Directional Regression: A Projective Resampling Approach

이아름¹, 김경원²

Abstract

In high-dimensional data analysis, directional regression is a tool for implementing linear sufficient dimension reduction by extracting core information while reducing data complexity. However, extending sufficient dimension reduction techniques to handle multivariate response data remains a relatively challenging task. Building upon the framework proposed by Li et al. (2008), this paper introduces a novel approach that integrates directional regression with a projective resampling method to tackle multivariate response regression problem. Our proposed methodology not only enhances the accuracy of recovering dimension reduction subspaces but also offers greater flexibility across diverse datasets. We validate the effectiveness of our approach through extensive simulation studies and real-world data analyses. This work provides a powerful alternative for multivariate response dimension reduction, with broad applicability in fields requiring high-dimensional data modeling.

Keywords : sufficient dimension reduction, directional regression, projective resampling, high dimensional data analysis, multivariate analysis.

¹Department of Statistics, Ewha Womans University, 52, Ewhayodae-gil, Seodaemun-gu, Seoul, 03760, South Korea. E-mail: 232stg19@ewha.ac.kr

²(Corresponding Author) Professor, Department of Statistics, Ewha Womans University, 52, Ewhayodae-gil, Seodaemun-gu, Seoul, 03760, South Korea. E-mail: kimk@ewha.ac.kr

Comparative Analysis of Likelihood-Based Dimension

Reduction Methods

Hyemin Yoo¹, Jae Keun Yoo¹

Abstract

To overcome the limitations of widespread dimension reduction methods such as Sliced Inverse Regression (SIR) and Sliced Average Variance Estimates (SAVE), many alternative methods have been proposed. In this paper, we compare two such methods likelihood acquired directions (LAD) and unstructured principal fitted components (UPFC)—both of which are likelihood-based sufficient dimension reduction techniques. While LAD and UPFC share foundational similarities in their approach to dimension reduction, they diverge significantly in the specifics of their assumptions and implementation about the underlying data structure.

¹Department of Statistics, Ewha Womans University.

Applying quantile regression to determine the effects of household characteristics on household saving rates in Vietnam

Thanh Hoa Bui¹, Keunjae Lee²

Abstract

Household savings are a crucial source of funding for investments that drive long-term economic growth, particularly in developing countries like Vietnam. In Vietnam, which consists of six distinct socio-economic regions, household savings also help buffer against income fluctuations and provide financial security. To understand the factors influencing household saving rates, the study employs a microeconomic approach using data from the 2018 Vietnam Household Living Standard Survey. Conditional quantile regression methods are applied to identify the determinants of household saving rates across Vietnam's six socio-economic regions, exploring regional heterogeneity in saving behaviors. Key findings show that household income positively affects saving rates in all regions, while education level, household size, and location generally have negative effects. The influence of income is strongest in the Northern Midlands and weakest in the Southeast. Wage income impacts saving more than farm income in certain regions, and dependency ratios show varying significance depending on region and saving level. Ethnic minority households tend to have higher saving rates than Kinh households, except in the Southeast and Mekong River Delta. These insights provide guidance for regional and national policymakers to design targeted strategies that encourage household savings and support socioeconomic development in Vietnam.

Keywords : Household saving rates, Vietnam, quantile regression, microeconomic approach.

¹46241 Busan Geumjeong-gu, Busandaehak-ro 63beon-gil 2, Pusan National University, Department of Economics, Graduate student. E-mail: buithanhhoa@pusan.ac.kr

²46241 Busan Geumjeong-gu, Busandaehak-ro 63beon-gil 2, Pusan National University, Department of Economics, Professor. E-mail: kjlee@pusan.ac.kr

A Comprehensive Software Reliability Growth Model Incorporating Dependent Failures and Uncertain Operating Environments^{*}

Onon-Ujin Otgonbayar¹, In Hong Chang², Kwang Yoon Song³, Youn Su Kim⁴

Abstract

As the reliance on software systems continues to expand across various fields, the demand for highly reliable software has become increasingly critical. In response to this, numerous statistical models have been developed to estimate software reliability. This study introduces a software reliability growth model that goes beyond traditional assumptions by addressing the influence of dependent failures that emerge under uncertain operating environments. Unlike previous approaches, which often presume that testing environments closely resemble operating environments, this model recognizes the diverse and dynamic settings in which software may ultimately be deployed. These environmental variations can significantly impact software failure rates and, consequently, overall reliability. By incorporating the effects of uncertain and varied conditions into the model, a more realistic reliability estimation is achieved. Twelve criteria are assessed by aligning actual failure data with the eleven models, which includes existing models and the one proposed in this study. Consequently, the model introduced in this study demonstrates its strengths over traditional approaches and its practical applicability in improving software reliability assessments.

Keywords : Software reliability growth model, Nonhomogeneous Poisson process, Dependent failure, Uncertain operating environment.

^{*}This research was supported by Global-Learning & Academic research institution for Master's·PhD students, and Postdocs(LAMP) Program of the National Research Foundation of Korea(NRF) grant funded by the Ministry of Education(No. RS-2023-00285353).

^{*}This study was supported by research fund from Chosun University, 2024.

¹Graduate student, Department of Computer Science and Statistics, Chosun University, Gwangju, 61452, Korea. E-mail: ujin7709@chosun.kr

²(Corresponding Author)Professor, Department of Computer Science and Statistics, Chosun University, Gwangju, 61452, Korea. E-mail: ihchang@chosun.ac.kr

³Assistant Professor, Department of Computer Science and Statistics and institute of Well-Aging Medicare & CSU G-LAMP Project Group, Chosun University, Gwangju, 61452, Korea.
E-mail: ccssig@chosun.ac.kr

⁴Post-Doc, Institute of Well-Aging Medicare & CSU G-LAMP Project Group, Chosun University, Gwangju, 61452, Korea. E-mail: mk92315@chosun.ac.kr

Revisit Forward Rate Unbiasedness Hypothesis using Smooth Time-Varying Cointegration Regression Model with Unknown Structural Breaks

*Xiaoyu Kang*¹, *Sung Y. Park*²

Abstract

A substantial body of prior research has focused on examining the fixed relationship between spot and forward exchange rates in order to test the forward rate unbiasedness hypothesis (FRUH). However, this assumption is quite strong, as there are numerous factors, such as macroeconomic shocks, that have the potential to disrupt the fixed relationship over the specified time horizon. This paper applies a smooth time-varying cointegration regression model with unknown structural breaks for the purpose of analyzing the aforementioned relationship across sixteen foreign exchange markets, including both advanced and emerging countries. We propose a model in which the functional parameters are specified by a well-known orthonormal series. Although cointegration theory is well-suited to elucidating long-run dynamics, it is possible that structural breaks may occur even within a smooth time-varying relationship. To address this issue, we have identified and tested the structural breaks in the considered model. The results demonstrate that the relationships are indeed time-varying and exhibit structural breaks. Furthermore, the FRUH, but also the foreign premium, are maintained over specific periods. Additionally, the results differ between countries, offering valuable insights for policymakers.

Keywords : Spot and Forward Exchange Rate; Forward Rate Unbiasedness Hypothesis; Time-varying Cointegration Regression; Structural break.

¹Ph.D. Candidate, School of Economics, Chung-Ang University, 84 Heukseok-Ro Dongjak-Gu, Seoul, 06974, Republic of Korea. E-mail: kxy8166@cau.ac.kr

²(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

hnic Minorities in Vietnam: A Study of the Factors InfluencingtheLivingStandards(Preliminary)

*Nguyen Tuan Dat*¹, *KeunjaeLee*²

Abstract

The authors use the Vietnamese Household Living Standards Survey 2020 dataset (VHLSS) and model regression to study the factors affecting the living standards of ethnic minorities in Vietnam. Income is determined to have the greatest influence on ethnic minorities' living standards, while health insurance has the opposite effect. This conclusion serves as the foundation for the authors to develop a number of recommendations and solutions for stakeholders to ensure and improve the living conditions of ethnic minorities in Vietnam.

Keywords : ethnic minorities; living standards; linear regression.

¹Ph.D. Student, Department of Economics, Pusan National University, Busan, Korea.

E-mail: nguyendat.tuan@pusan.ac.kr

²(Corresponding Author) Professor, Department of Economics, Pusan National University, Busan, Korea.

Estimation of weather impacts on mosquito abundance using Bayesian period-varying Poisson modeling with reduced-rank regression*

*Jieun Lee*¹, *Myungjin Kim*^{1,2}, *Yongku Kim*^{1,2}, *Seung Yeop Yang*^{1,3}, *Jun-Yeong Lee*^{1,4}, *Semin Oh*¹, *Seongjun Byun*¹, *Young Ho Kim*⁵, *Ye Eun Park*⁵

Abstract

Poisson regression models have been classical tools for estimating and predicting species abundance. As the changes in weather conditions become more frequent and dramatic, however, Poisson models that can reflect varying impacts of climate variables with respect to time are in need. In this paper, we expand the application domain of Poisson regression models by developing a period-varying Poisson regression framework that can explain the varying impacts of covariates on the outcome over the course of time. The framework of period-varying Poisson regression modeling uses the notion of reduced-rank regression that improves the accuracy of statistical inference by allowing the regression coefficient vectors to share information with each other. To enable statistical inference under the low-rank constraint, we incorporate the weighted Bayesian bootstrap into our framework. Results from a simulation study are provided to demonstrate the performance of our proposed method. Finally, the period-varying Poisson regression model is applied to investigate the varying climate effect on the mosquito population in a rural city in South Korea.

Keywords : Bayesian inference, Climate impacts on ecosystems, Maximum A Posteriori (MAP) estimation, Row-rank constraints, Weighted Bayesian bootstrap.

*This research was supported by the Basic Science Research Program and the Global - Learning & Academic research institution for Master's, PhD students, and Postdocs (G-LAMP) Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. RS-2023-00248103, No. RS-2023-00301914).

¹KNU G-LAMP Project Group, KNU Institute of Basic Sciences, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 10587, Republic of Korea.

²Department of Statistics, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 10587, Korea.

³Department of Mathematics, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 10587, Korea.

⁴School of Life Sciences, BK21 FOUR KNU Creative BioResearch Group, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 10587, Korea.

⁵Department of Ecological Science, College of Ecology & Environmental Science, Kyungpook National University, 2559 Gyeongsang-daero, Sangju-si, Gyeongsangbuk-do 37224, Korea.

Exploring Regression Associations in Multidimensional Contingency Tables with an Ordinal Dependent Variable*

Saebom Jeon¹, Daeyoung Kim²

Abstract

We present a data-analytic framework designed to reveal the associations within firefighters' PTSD (Post-Traumatic Stress Disorder) survey data, structured in a multidimensional contingency table. The primary aim of this study is to explore possible regression relationships between firefighters' PTSD levels, an ordinal response variable, and a comprehensive set of categorical risk factors. These risk factors include demographic, job-related, and mental health factors, which we analyze both in terms of their joint/marginal relationships and conditional influences on PTSD patterns. To achieve these insights, our approach makes use of two established data-driven methods: the newly developed non-model-based regression association measure, CCRAM (Checkerboard Copula Regression Association Measure), and resampling techniques such as bootstrap and permutation methods. The empirical findings suggest that this framework is an effective exploratory tool for identifying and exploring potential regression associations in complex, multidimensional categorical data. It also provides a basis for identifying promising candidate hypotheses for further confirmatory analysis.

Keywords : Multidimensional contingency tables, ordinal categorical variables, PTSD, resampling methods.

*This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2022R1F1A1065520 to Jeon).

¹35349 Department of Marketing Bigdata, Mokwon University, Daejeon, South Korea.

E-mail: alwaysns@mokwon.ac.kr

²17035 Department of Mathematics and Statistics, University of Massachusetts Amherst, MA, USA.

E-mail: dkim1@umass.edu

Developing a Carbon Emission Index: A Framework Using ESG Reports and Text Mining

Sung Jae Kim¹, Youngbok Chor²

Abstract

As the demand for corporate accountability in carbon emissions grows, the need for standardized metrics to evaluate progress toward carbon neutrality has become critical. This study aims to develop a comprehensive carbon emission index based on annual Environmental, Social, and Governance (ESG) reports published by corporations. The proposed index consists of seven key components, each assigned a reference score derived through a systematic approach. Relevant sentences associated with each component are extracted from ESG reports using Euclidean distance-based similarity measures. Sentiment analysis is then applied to the extracted sentences to generate sentiment scores for each component. These sentiment-driven scores, along with corporate ESG ratings, are utilized to construct a predictive model for the carbon neutrality index. The study's findings provide a robust framework for measuring corporate efforts in achieving carbon neutrality, offering insights into the alignment between ESG reporting and sustainable practices. This research contributes to the development of standardized tools for assessing carbon emissions and supports evidence-based policymaking for sustainable development.

Keywords : Carbon Emission Index, ESG Reports, Euclidean Distance, Sentiment Analysis, Text Mining, Carbon Neutrality, Sustainability Metrics, Corporate ESG Ratings, Sustainable Development, Predictive Modeling

¹Research Professor, Department of Management Information Systems College of Business Administration Dong-A University 225 Gudeogro, Seo-gu, Busan, Republic of Korea, 49236.

E-mail: sung0704@dau.ac.kr

²(Corresponding Author) Assistant Professor, Department of Management Information Systems College of Business Administration Dong-A University 225 Gudeogro, Seo-gu, Busan, Republic of Korea, 49236.

E-mail: ybcho@dau.ac.kr

사이버범죄 데이터에 대한 텍스트마이닝*

김근아¹, 한상태²

요약

시대의 변화에 따라 불특정 다수를 인제든 만날 수 있는 사이버 공간이 일상에 자리 잡으며 사이버범죄라는 새로운 범죄 유형을 만들어냈다. 본 연구에서는 빅데이터 분석 플랫폼 중 하나인 Textom을 이용하여 사이버범죄의 여러 유형 중 보이스피싱, 스미싱, 로맨스스캠, 몸캠피싱에 대한 총 21,831개의 지식 IN 본문 글을 기반으로 CONCOR 분석, 감성분석을 진행하였다. 감성분석 결과, 보이스피싱은 ‘괜찮다’, ‘정확하다’ 등 피해자를 안심시키기 위한 단어를 주로 사용했다. 스미싱은 거부감, 분노와 같은 부정적 단어들 많이 나타났으며 자책하기보단 문자 내용 공유를 통해 또 다른 피해를 막거나 더욱 조심하려는 행동 양식이 파악된다. 앞선 두 범죄와 달리 몸캠피싱, 로맨스스캠은 상대적으로 데이터가 적음에도 감성어휘사전의 단어를 많이 포함하고 있다. 이는 상황 설명을 위해 본문에 작성한 피해 과정에서 감정적 교류가 많았으며 작성자가 감정적으로 가장 많이 시달린 것으로 판단된다. 상당수의 글이 도움을 청하는 내용임을 통해 다른 범행에 비해 주변에 밝히거나 도움을 받기에 상대적으로 어려워 익명성을 빌린 것으로 보인다. 따라서 피해자들에 대한 더 많은 보호와 관심이 필요하다고 판단된다. 또한, 파이썬 프로그램을 이용하여 CNN, LSTM, DistilBERT 모델로 보이스피싱 판별 모델의 성능을 비교해 보았다. 분석 결과를 비교해 보면 문맥에 따라 다른 의미를 띄는 패턴 학습과 순차적 의존성 학습에 강한 LSTM 모델이 가장 우수한 결과를 보였다.

주요 용어 : 텍스트마이닝, 딥러닝, 사이버범죄, 보이스피싱.

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

¹31499 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 데이터사이언스학과 석사과정.
E-mail: kimguena_23@naver.com

²(교신저자)31499 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 빅데이터AI학부 교수.
E-mail: sthan@hoseo.edu

지각적으로 현저한 자극과 정서적으로 현저한 자극에 대한 기능적 뇌 네트워크의 자발적 반응: 기능적 데이터분석을 통한 접근*

정해찬¹, 정고은¹, 이은지², 한석원²

요 약

본 연구는 외인적, 내인적 주의 제어의 신경 메커니즘을 탐구하고, fMRI 데이터 분석에서 기능적 데이터 분석(FDA, Functional Data Analysis)의 적용 가능성을 검증하였다. 본 연구의 실험에서 주의 제어 과정을 관찰하기 위해 과제와 맥락적으로 무관한 방해자극을 사용했다. 참가자들은 연속적으로 변화하는 숫자들 중에서 한글 글자를 탐지하는 과제를 수행하는 동안, 화면 중앙에 비정서적 또는 정서적 방해자극이 지속적으로 제시되었다. BOLD (Blood oxygen level dependent) 신호의 시간적 연속성을 보존함과 노이즈 관리를 위해 B-spline 평활화 기법을 도입하였으며, 연속적인 데이터의 시계열적 변화를 통계적으로 평가하고자 반복측정 기능적 분산분석(repeated measures functional ANOVA)을 적용하였다. 연구 결과, 다중수요 네트워크(MDN, Multiple Demand Network), 기본모드 네트워크(DMN, Default Mode Network), 현저성 네트워크(SN, Salience Network)에서 각각의 기능적 특성을 반영하는 차별적인 활성화 패턴이 관찰되었다. MDN은 주의 통제, DMN은 내적 정보 처리, SN은 자극 감지와 평가에서 특징적인 활성화를 보였으며 세 네트워크 모두에서 비정서적 자극보다 정서적 자극에 대해 더 강한 활성화가 확인되었다. 이러한 결과는 두 가지 측면에서 중요한 의미를 가진다. 첫째, 정서적 자극과 비정서적 자극이 주의 메커니즘에 미치는 차별적 영향을 신경활동 수준에서 규명하였다. 둘째, FDA 기반 분석을 통해 BOLD 신호의 연속적 특성을 보존하며 뇌 네트워크의 역동적 활성화 패턴을 정밀하게 포착할 수 있음을 입증하였다. 이는 향후 실험 기반 fMRI 연구에서 시계열 데이터 분석의 새로운 가능성을 제시한다.

주요 용어 : fMRI, B-spline 평활화, 반복측정 기능적 분산분석.

*이 논문은 한국정부가 지원하는 한국연구재단의 지원으로 수행되었음.

(NRF-2022R1F1A1063125, RF-2022M3J6A1084843).

¹34134 대전광역시 유성구 대학로 99 충남대학교 심리학과. E-mail: hczzang71@naver.com

¹54896 전북특별자치도 전주시 덕진구 백제대로 567 전북대학교 심리학과 교수. E-mail: jungke1225@gmail.com

²(교신저자) 34134 대전광역시 유성구 대학로 99 충남대학교 정보통계학과 교수. E-mail: eunjee.cnu@gamil.com

²(교신저자) 34134 대전광역시 유성구 대학로 99 충남대학교 심리학과 교수. E-mail: suk.w.han@gmail

기업 구성원의 직무만족도 예측변인 탐색: 머신러닝과 설명 가능한 인공지능(XAI)의 적용*

장은아¹, 정혜원²

요 약

본 연구는 기업 구성원의 직무만족도와 관련된 주요 예측변인을 탐색하기 위하여 수행되었다. 이를 위하여 한국직업능력연구원에서 조사한 인적자본기업패널Ⅱ(Human Capital Corporate Panel Ⅱ, HCCPⅡ)의 3차년도(2022년) 근로자 설문 자료(n=9,512)를 사용하였고, 기업 구성원의 직무만족도를 가장 잘 예측하는 머신러닝 모델을 선택하기 위하여 랜덤포레스트(random forest, RF)와 익스트림 그라디언트 부스팅(eXtreme Gradient Boosting, XGBoost)의 예측성능(root mean squared error, mean absolute error, R^2)을 비교하였으며, 주요 예측변인을 도출하고 예측변인과 정신건강과의 경향성을 시각적으로 확인하고자 설명 가능한 인공지능 기법 중 하나인 SHAP(SHapley Addictive exPlanations) 알고리즘을 사용하였다. 주요 결과는 다음과 같다. 첫째, RF 대비 XGBoost의 예측성능이 상대적으로 우수한 것으로 나타났다. 둘째, 기업 구성원의 직무만족도에 대한 상위 중요도 예측변인을 탐색한 결과, 개인 특성 영역에서는 조직몰입, 이직의도, 직무스트레스 문항이 중요도가 높은 것으로 나타났으며, 조직몰입은 직무만족도와 정적인 경향을 보인 데 반해, 이직의도와 직무스트레스는 부적인 경향성을 보이는 것으로 나타났다. 기업 특성 영역에서는 지식의 상호보완성, 조직 내 민주성, 인재우대 정책, 커뮤니케이션과 신뢰관계, 기업문화(혁신문화, 관계문화) 변인이 도출되었으며, 기업 특성 변인의 경우 직무만족도와 모두 정적인 경향을 보이는 것으로 확인되었다. 주요 결과를 토대로 기업 구성원의 직무만족도 증진을 위한 정책 마련의 시사점을 제시하였다.

주요 용어 : XGBoost, random forest, SHAP, 직무만족도, HCCPⅡ.

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

¹34134 대전광역시 유성구 대학로99, 충남대학교 교육학과 박사과정. E-mail: rhcsla456@naver.com

²(교신저자) 34134 대전광역시 유성구 대학로99, 충남대학교 교육학과 교수. E-mail: chw7@cnu.ac.kr

Model evaluation criterion for anti-cancer drug sensitivity-specific gene network inference

오주이¹, 박희원¹

Abstract

The gene network analysis is essential for uncovering disease mechanisms, because the complex mechanism involved in perturbation of molecular interplays rather than abnormalities of a single gene. The linear regression model and the L1-type regularization methods are commonly used to infer gene network, but it provides averaged results across samples, failing to reveal the cell line-specific gene regulatory system. So, we consider the varying coefficient model with the kernel-based L1-type regularization that enable us to reveal cell line characteristic-specific gene networks. In the L1-type regularized regression modeling, the hyperparameter selection is crucial, because it strongly influences both results of gene selection and network estimation. While Cross-Validation (CV) is widely used to select hyperparameters, it leads to computationally expensive procedures, and provides an averaged model evaluation result. To settle on this limitation, we propose a novel model evaluation criterion, called sample-specific Generalized Cross-Validation (ss-GCV). We demonstrate through Monte Carlo simulations that our strategy provides effectiveness results for sample-specific gene network analysis. Our method also applied to cytarabine (i.e., anti-cancer drug of AML) sensitivity-specific gene network analysis and reveal the cytarabine sensitivity-specific molecular interplays. We expect that our strategy will be a useful tool for cell line-specific gene network analysis and provide crucial insights to precision anti-cancer therapy.

Keywords : sample-specific Generalized Cross-Validation (ss-GCV), varying coefficient model, gene network, cell line-specific analysis.

¹Department of Statistics, Sungshin Women's University, Seoul, Republic of Korea.

딥러닝 기반 제네레이티브 모델을 활용한 함수형 데이터 합성

이동훈¹, 송준²

요약

데이터 부족과 보안 문제로 인해 합성 데이터는 다양한 분야에서 중요한 역할을 한다. 최근에는 생성적 적대 신경망(Generative Adversarial Network, GAN), 변분 오토인코더(Variational Autoencoder, VAE), 확산 모델(Diffusion Model) 등 딥러닝 기반 생성 모델을 통해 비선형적 관계를 파악하고 정교한 데이터 합성이 가능해졌다. 그러나 이러한 모델들은 주로 표형 데이터(Tabular Data)에 최적화되어 있어, 시간적 연관성이 높고 연속적인 고차원 함수형 데이터(Functional Data)에 직접적으로 적용하기 어렵다. 본 연구에서는 함수형 데이터의 합성을 위해 적응형 기저(Adaptive Basis)를 활용한 차원 축소 기법과 생성적 적대 신경망 및 확산 모델을 결합한 새로운 방법을 제안한다. 이를 통해 주어진 데이터를 효과적으로 표현하는 기저를 학습하고, 데이터 간의 연관성을 반영한 합성 데이터를 생성할 수 있었다. 실험 결과, 기존의 표형 데이터 합성 모델에 비해 함수형 데이터의 특성을 잘 보존하는 매끄러운 데이터를 생성하였으며, 합성 데이터를 활용한 회귀 분석에서도 향상된 성능을 확인하였다.

주요용어: 데이터 합성, 함수형 데이터, 생성형 모델, 적응형 기저

¹02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 석사과정. E-mail: ydh8206@korea.ac.kr

²(교신저자) 02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 부교수. E-mail: junsong@korea.ac.kr

A Deep Semi-supervised Learning Framework using Learnable Data Augmentation technique via Masked Autoencoders for tabular data

Jihye Lee¹, Dongha Kim²

Abstract

Semi-supervised learning(SSL) is a method for learning a powerful prediction model by leveraging both labeled and unlabeled data. The central idea of SSL is to propagate the label information from labeled data to the unlabeled data, effectively increasing the amount of labeled data as learning proceeds. To achieve this, data augmentation techniques, such as rotation and reflection in image domains, are commonly used. These augmentations have significantly enhanced performance in deep prediction models in SSL regime for image data. However, in contrast to image data, effective and desirable augmentation techniques for tabular data still remain limited, constraining the development of deep learning-based SSL methods in this domain. In this study, we introduce a novel transformer-based framework for SSL tasks in tabular data. Our method is an end-to-end type that integrates learnable data augmentation techniques with a prediction model within a single architecture. We modify masked autoencoders (MAE) by adding a prediction model to the encoder, creating a unified structure. By combining reconstruction error and FixMatch-like loss function, our framework implicitly generates augmented data while simultaneously achieving highly accurate prediction model. To empirically validate the superiority of our method, we compare its performance to other state-of-the-art competitors and also traditional machine learning approaches, and find that our method achieves the highest test AUC scores over all evaluated tabular datasets, even with a small amount of labels.

Keywords : Deep Semi-supervised Learning, Learnable Data Augmentation, Masked Autoencoders.

¹02844 Bomun-ro, 34 Da-gil 2, Seongbuk-gu, Seoul, Korea, Master's Student, Department of Statistics, Sungshin Women's University. E-mail: jihyelee987@gmail.com

²(Corresponding Author) 02844 Bomun-ro, 34 Da-gil 2, Seongbuk-gu, Seoul, Korea, Assistant Professor, Department of Mathematics, Statistics and Data Science, Sungshin Women's University. E-mail: dongha0718@sungshin.ac.kr

Improving Outlier Detection by Optimizing the Inlier-Memorization Effect in Deep Generative Models

*Seoyoung Cho*¹, *Jaeesung Hwang*², *Kwan-Young Baek*³, *Dongha Kim*⁴

Abstract

Outlier detection (OD) identifies unusual data points, or outliers, by learning patterns within typical observations, known as inliers. Recently, a robust unsupervised OD (UOD) approach was introduced, leveraging the inlier-memorization (IM) effect, a phenomenon in which deep generative models tend to memorize inliers before outliers during the early stages of training. This study aims to enhance UOD by leveraging this effect. Observing that the IM effect strengthens when fewer outliers are present, we propose a method that excludes outliers from mini-batches during loss calculation. Our approach, Adaptive Loss Truncation with Batch Increment (ALTBI), combines two techniques: gradually increasing mini-batch size and applying an adaptive threshold for truncated loss. We theoretically demonstrate that these strategies enhance the IM effect by filtering out outliers from the loss function. Additionally, ALTBI incorporates an ensemble method and achieves state-of-the-art performance across extensive datasets, all with lower computational demands. It also maintains effectiveness when integrated with privacy-preserving techniques.

Keywords : Outlier Detection, Deep Generative Models, IM Effect

¹02844 Bomun-ro, 34 Da-gil 2, Seongbuk-gu, Seoul, Korea, Master's Student, Department of Statistics, Sungshin Women's University. E-mail: katesycho@gmail.com

²Manager, SK Telecom, 264, Pangyo-ro, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea. E-mail: postechiminuru@gmail.com

³02844 Bomun-ro, 34 Da-gil 2, Seongbuk-gu, Seoul, Korea, Assistant Professor, Department of Mathematics, Statistics and Data Science, Sungshin Women's University. E-mail: kybak@sungshin.ac.kr

⁴(corresponding author) 02844 Bomun-ro, 34 Da-gil 2, Seongbuk-gu, Seoul, Korea, Assistant Professor, Department of Mathematics, Statistics and Data Science, Sungshin Women's University. E-mail: dongha0718@sungshin.ac.kr

온센서 음성 분류를 위한 전이학습 기반의 신경망 구조 자동 탐색 기법*

전성재¹, 이영민²

요약

마이크로컨트롤러(MCU)에서 신경망을 추론하는 연구가 제안되고 있지만, 메모리 자원의 한계로 인해 MCU를 대상으로 하는 신경망 구조의 설계는 여전히 어려운 작업이다. MCU의 제약 조건을 만족하기 위한 이전 연구에서는, 단일 슈퍼넷(Supernet)의 학습만으로 다양한 MCU 제약 조건에서 신경망을 획득할 수 있는 원샷(One-Shot) 신경망 구조 자동 탐색 기법이 주목을 받았다. 그러나, MCU를 대상으로 음성 분류를 위해 설계된 슈퍼넷은 제안된 사례가 적으며, 소규모 데이터셋으로 학습이 어려운 슈퍼넷의 한계로 인해 대부분 사람이 직접 설계한 신경망 모델보다 낮은 성능을 획득하였다. 사람이 직접 설계하는 방법은 검증을 위해 매번 반복적인 가지치기(Pruning)와 미세조정을 요구하여 효율적이지 않으므로, 적은 비용으로 MCU에서 높은 성능을 달성하는 신경망을 획득할 수 있는 다른 해결책이 필요하다. 본 연구에서는 해결책으로 소규모 음성 데이터셋에서도 높은 성능을 갖는 슈퍼넷을 학습할 수 있는 초소형 신경망 구조 자동 탐색 기법을 제안한다. 우선 높은 정확도를 달성하기 위해서, 대규모 이미지 데이터셋인 ImageNet으로 사전학습한 슈퍼넷을 전이학습(Transfer Learning)하여 음성 분류 슈퍼넷을 획득하였다. 전이학습 과정에서 입력으로 단일 차원의 스펙트로그램(Spectrogram)으로 변환한 음성 데이터를 사용하고, 슈퍼넷의 입력 레이어 가중치를 3차원에서 단일 차원으로 병합하는 방법을 통해, 입력 레이어의 추론 과정에서 발생하는 SRAM 메모리 병목 문제를 개선하는 동시에 연산량을 감소시켰다. ESC-50 음성 데이터셋으로 전이학습한 슈퍼넷을 탐색한 결과로, SRAM 1MB, Flash 2MB의 MCU 제약 조건에서 선행 합성곱 신경망(CNN) 모델의 성능을 능가하는 Top-1 90.8% 정확도의 신경망을 획득할 수 있었다.

주요 용어: 신경망 구조 자동 탐색 기법, 마이크로컨트롤러, 음성 분류, 전이학습.

*이 연구는 2022년 정부(방위사업청)의 재원으로 국방과학연구소의 지원을 받아 수행된 미래도전국방기술 연구개발사업(No.915062201)

¹02504 서울 동대문구 서울시립대로 163, 서울시립대학교 인공지능학과 석사과정.

E-mail: erectbranch21@uos.ac.kr

²(교신저자) 04107 서울 마포구 백범로 35, 서강대학교 인공지능학과 교수. E-mail: ymyi@sogang.ac.kr

functional Sequence Kernel Association Test (fSKAT) for Genetic Variant identification in hippocampal radial distances^{*}

Hyeonjeong Lim¹, Yoonseok Lee², Eunjee Lee³

Abstract

Imaging genetics has advanced to uncover genetic variants associated with imaging phenotypes characterized by imaging data, which can be considered as objective phenotypes of genetic diseases. We focus on the hippocampal radial distance of Alzheimer's Disease Neuroimaging Initiative (ADNI) subjects as an imaging phenotype. In this study, we propose a novel approach, the functional sequence kernel association test (fSKAT), to test the effects of genetic variants on an imaging phenotype that can be represented by functional data. As a functional phenotype, we constructed hippocampal radial distance curves derived from structural magnetic resonance imaging data, defined by the distances from the centroid of the hippocampus to its surface at multiple radial points. Next, to address the infinite dimensionality of the data and extract important features, we employed functional principal component analysis for basis expansion of the functional phenotype. Finally, fSKAT was used to incorporate sets of non-linear single-nucleotide polymorphism (SNP) effects on the functional phenotypes, leveraging functional principal component scores. This study applied the proposed method to genotypes in chromosome 19 of ADNI subjects, including cognitively normal, mild cognitive impairment (MCI), and Alzheimer's disease patients, to identify significant genetic variants. Our results demonstrated superior statistical power and computational efficiency compared to other competing approaches that do not account for the functional nature of imaging data.

Keywords : functional data analysis, Imaging genetics, genetic variants, radial distances, phenotypes, kernel machine 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. NRF2022M3J6A1084843, No. NRF-2021R1C1C1013936)

¹Department of statistics and data science, Chungnam National University, hyeonjeongisabel@naver.com

²Department of statistics and data science, Chungnam National University, toatoal123@gmail.com

³Department of statistics and data science, Chungnam National University, eunjee.cnu@gmail.com

Improving Time Series Forecasting Accuracy with Increasing Weighted Loss Functions for Further Prediction

Sumin Jeon¹, Jun Song²

Abstract

Time series forecasting is a crucial task in multivariate time series analysis. With the recent demonstration of the superior performance of Transformer models in time series forecasting, this study selects the PatchTST model, a Transformer variant specifically suited for multivariate time series analysis. However, the loss function commonly employed in traditional Transformers, Mean Squared Error (MSE), has limitations, particularly in its tendency to exhibit larger discrepancies between predictions and actual values for more recent time points. To address this, this paper proposes a Weighted Sum Loss Function to enhance the predictive accuracy of time series analysis using Transformers. The proposed loss function assigns progressively increasing weights to recent time points in various functional forms, giving the highest weight to the most recent observations, thereby enabling more precise time series forecasting. The proposed model was evaluated using simulation data generated from stochastic processes and hourly measured ozone data from the United States. Results demonstrate that the PatchTST model with the Weighted Sum Loss Function outperforms existing models in predictive accuracy.

Keywords : Time series forecasting, PatchTST, Increasing Weighted Loss Functions.

¹MS student, Department of Statistics, College of Political Science and Economics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail : wjstnals0224@korea.ac.kr

²(Corresponding Author) Professor, Department of Statistics, College of Political Science and Economics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail : junsong@korea.ac.kr

Automatic Seed Word Selection for Topic Modeling

Dahyun Jeong¹, Jeongin Hwang², Yunjin Chor³, Yoon-Yeong Kim⁴

Abstract

Topic modeling has emerged as a prominent technique for uncovering latent semantic topics from a corpus. While fully unsupervised topic models have been widely employed in topic modeling tasks, they often tend to identify common or dominant topics within documents, potentially overlooking minor topics. Recently, seed-guided topic modeling has gained attention by incorporating ‘seed words’ to help uncover both dominant and previously overlooked minor topics. However, in most existing studies, seed words are chosen in supervised settings, often using label-dependent metrics or manual selection. In real-world applications, where data is typically unstructured, these methods are difficult to scale, and reliance on domain expertise can introduce potential bias into the seed words. Notably, the selection of seed words in unsupervised settings remains underexplored. To address these limitations, we propose an automated seed word selection process that discovers sets of words that are diverse and cohesive based on inter-word relationships. Next, we provide an instantiation of the proposed process as an algorithm called SeedCapture, which utilizes co-occurrence to capture meaningful associations. The proposed SeedCapture has several benefits: it requires minimal parameters for tuning, making it efficient to use. In addition, the algorithm is highly adaptable and can be seamlessly integrated into the seed word selection process of traditional seed-guided topic models. Through quantitative and qualitative analyses, we demonstrate the effectiveness of our process across various corpora and seed-guided topic models.

keywords : seed-guided topic modeling, seed words, automatic seed word selection.

¹Graduate Student, Department of Statistics, University of Seoul. E-mail: ekgus638@uos.ac.kr

²Graduate Student, Department of Statistical Data Science, University of Seoul.

³Professor, Department of Statistics, University of Seoul.

⁴(Corresponding Author) Professor, Department of Statistics, University of Seoul. E-mail: yykim@uos.ac.kr

BERT-based graph level anomaly detection for molecular graphs

Kione Kim¹, Jong-June Jeon²

Abstract

In recent years, there has been a growing interest in graph level anomaly detection, which is the detection of anomalous graphs that exhibit unusual features and topology compared to the majority of normal graphs. In this paper, we present a BERT-based learning framework for detecting local and global anomalies. Our approach uses BERT to learn a latent vector and learns to minimize the difference between the reconstructed node attributes based on the latent vector and the original node attributes. The difference between the original and reconstructed node attributes and the distance between CLS tokens and cluster centers can be utilized to compute an anomaly score to represent the degree of anomaly in each graph. To evaluate the performance of our method, we conducted experiments using the molecular graphs benchmark dataset to demonstrate its effectiveness in detecting anomalous graphs.

keywords : Graph Anomaly Detection, BERT for Graph Representation, Molecular Graphs.

¹Graduate Student, Department of Statistical Data Science, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul 02504, Korea. E-mail: rldnjs1616@uos.ac.kr

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

시계열 모형을 이용한 커피음료와 탄산음료 매출액 예측모형 개발¹

김민지², 한상태³

요 약

건강을 중요시하는 소비층이 증가하면서 디카페인, 제로 슈거 제품이 출시됨에 따라 커피음료와 탄산음료의 매출이 증가하고 있다. 본 연구는 (주)마켓링크에서 수집된 고수요 유통사 POS 인덱스 데이터 자료를 이용하여, 2021년 1월부터 2023년 12월까지 수집된 유통채널별 카테고리 판매 데이터를 이용한 주요 관심 품목에 대한 매출 예측 모형을 개발하였다. 특히 전체채널, 대형마트, 편의점에 대한 커피음료와 탄산음료의 매출액을 중심으로 분석을 진행하였다. 본 연구에서는 시계열 분석 기법 중 ARIMA모형과 Holt-윈터스 지수평활법을 이용했으며, 모형비교를 위해 AIC와 SBC 지표를 활용하였다. 매출 예측을 진행한 유통채널 모두 계절성을 띄고 있으며, 분산의 안정성을 위해 추가적으로 차분을 진행하여 분석을 진행하였다. 분석 결과, AIC와 SBC 값이 가장 낮은 상수항이 제거된 ARIMA모형이 최적모형으로 선택되었다. 그러나, 본 연구는 코로나19 직후로 경제적 변화 같은 외부요인이 매출에 영향을 미쳤을 가능성을 가지고 있어, 이러한 요인을 반영한 추가적인 연구가 필요하다. 또한 본 연구는 특정 기간과 음료 카테고리에 한정되어 있어, 향후 연구에서는 더 다양한 음료 카테고리 및 장기적인 데이터를 이용하여 장기 매출 예측에 대한 연구가 필요하다는 한계점을 제시했다.

주요용어 : ARIAM, 매출액 예측 모형

¹이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2022M3J6A1063595).

²31499 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 데이터사이언스학과 석사과정.
E-mail: kminji_23@naver.com

³(교신저자)31499 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 빅데이터AI학부 교수.
E-mail: sthan@hoseo.edu

Low-Dimension Adaptive Neural Network Regression with Directional Change Detection via Nuclear Norm Penalization

Yongku Kim¹, Jae-Hwan Jhong², Ja-Yong Koo³, Kwan-Young Bak⁴

Abstract

This paper reports on our study of a regularized neural network regression method that adapts to the low-dimensional directional structure of a function. The key innovation lies in identifying direction vectors where the function exhibits significant variation and conducting estimation within this reduced-dimensional space. This is achieved through a regularization scheme that controls the nuclear norm of the weights matrix. The nuclear norm penalization approach effectively reduces its rank and allows the model to discover the principal subspace where the function varies. Additionally, an ℓ_1 penalty is imposed to introduce node-level sparsity, which further enhances dimension reduction and improves estimation efficiency. The combination of the nuclear norm and ℓ_1 norm penalties results in a low-dimensional network structure that strikes a good balance in the bias-variance trade-off and improves estimator performance. An efficient and stable implementation scheme is designed based on the alternating direction method of multipliers and the Levenberg-Marquardt algorithm. The stability of the algorithm is enhanced by using a B-spline activation function with compact support and an initialization strategy based on model-based sliced inverse regression. Numerical experiments on simulated and benchmark datasets demonstrate that our method outperforms several popular machine learning and neural network regression techniques.

Keywords : Curse of dimensionality, Directional information, Multivariate regression, Neural network, Nuclear norm penalization.

¹MS student, Department of Statistics, College of Political Science and Economics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, South Korea. E-mail : rladydm3194@korea.ac.kr

²Professor, Department of Information Statistics, College of Natural Sciences, Chungbuk National University, Chungdae-ro 1, Cheongju, Chungbuk, 28644, South Korea. E-mail : jjh25@chungbuk.ac.kr

³Professor, Department of Statistics, College of Political Science and Economics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, South Korea. E-mail : jykoo@korea.ac.kr

⁴(Corresponding Author) Professor, Department of Statistics, School of Mathematics, Statistics and Data Science, Sungshin Women's University, Bomun-ro 34da-gil, Seongbuk-gu, Seoul, 02844, South Korea. E-mail : kybak@sungshin.ac.kr

웨이퍼 결함 패턴 분류를 위한 계층적 분류 모델과 특징 중요도 분석

김재윤¹, 정재홍²

요약

반도체 산업 발전과 복잡해지는 제조 공정으로 인하여 보다 정밀한 결함 패턴 분류 방법의 필요성이 증가하고 있다. 하지만 많은 선행 연구들은 결함 패턴의 분류 정확도에 중점을 두어 왔으나, 분류 모델이 어떤 특징을 바탕으로 결함 패턴을 분류하는지에 대한 체계적인 분석은 상대적으로 부족한 상황이다. 이에 따라 본 연구에서는 계층적 분류와 Shapley value (SHAP)을 활용한 특징 중요도 분석을 통해 해석 가능한 결함 패턴 분류 방법을 제안한다. 본 연구에서는 WM811-K 자료의 9가지 결함 패턴을 5개의 그룹으로 재구성하여 계층적 분류 기법을 적용함으로써 그룹 간 분류와 그룹 내 분류를 수행하였다. 실험에 사용한 모델은 Random Forest, XGBoost, SVM, KNN이며, 평가지표로는 Accuracy, F1-score, AUC를 사용하여 비교하였다. 실험 결과, 계층적 분류를 적용한 XGBoost가 가장 높은 성능을 보였다. 이어서 SHAP을 활용하여 XGBoost에서 각 분류 단계에서 어떤 특징들이 중요한 영향을 미치는지 살펴보았다.

주요 용어 : 웨이퍼 결함 패턴, 계층적 분류, SHAP.

¹04763 서울시 성동구 왕십리로 222, 한양대학교 응용통계학과 석사과정. E-mail: neop0801@hanyang.ac.kr

²(교신저자) 04763 서울시 성동구 왕십리로 222, 한양대학교 수학과 부교수. E-mail: jaehongjeong@hanyang.ac.kr

머신러닝 기반 안면인식 기술을 활용한 피부 시술의 안면 동안 효과에 대한 정량화 연구

장현규¹, 노민정²

요 약

본 연구는 머신러닝 기반의 안면인식 알고리즘을 사용하여 미용 시술을 받은 2만여 명의 안면 이미지를 분석하였다. 알고리즘을 통해 측정된 나이와 특정 피부 상태(예: 다크서클, 여드름)를 바탕으로 피부 시술이 안면 동안 효과에 미치는 영향을 조사하였다. 베이지안 분산분석(Bayesian ANOVA)을 이용한 매개효과 분석 결과, 피부 시술은 색소 침착, 다크서클, 여드름과 같은 피부 상태를 개선하여 전체적으로 더 젊어 보이는 외관을 만들어 주는 것으로 나타났다. 주요 시술 방식의 효능 또한 분석하였으며, 그 결과 수술적 안면 리프팅, 실 리프팅, 자가 지방 이식, 고강도 집속 초음파(HIFU)가 안면 동안 효과에 유의미한 영향을 미치는 것으로 밝혀졌다. 수술적 안면 리프팅은 수혜자가 약 3.4년 더 젊어 보이는 결과를 낳았고, 실 리프팅은 약 1년, 자가 지방이식은 약 6개월 정도 젊어 보이게 하는 효과를 낳았다. 반면 지방 감소, 고주파(RF), 레이저 기반 시술은 동안 효과에 유의미한 영향을 미치지 못하였다.

주요 용어: 인공지능, 안면인식, 베이지안 분산분석 (Bayesian ANOVA), 안면 리프팅, 피부 변색, 다크서클, 여드름.

¹03760 서울특별시 서대문구 이화여대길 52 이화여자대학교, 이화여자대학교 국제대학원 조교수.

E-mail: hyunhyu.jang@ewha.ac.kr

²(교신저자) 48513 부산광역시 남구 용소로 45 부경대학교 대연캠퍼스, 부경대학교 경영학부 부교수.

E-mail: mroh@pknu.ac.kr; 08826 서울시 관악구 관악로 1 서울대학교 경영대학, 서울대학교 경영학과 객원부교수. E-mail: pagani@snu.ac.kr

침수 피해 위험 예측을 위한 모형 개발

박영재¹, 안수경², 최보승³

요약

전 세계적인 기후온난화 따른 집중호우가 증가하고 있으며 이에 따른 침수 피해의 빈도와 규모가 증가하고 있다. 이에 따라 침수 피해 위험 예측 연구는 더욱 중요하고 필수적인 분야로 부각되고 있다. 본 논문에서는 다양한 다변량 통계 기법을 사용하여 침수 피해 지역을 예측하고자 하는 새로운 모형을 구축하고자 한다. ‘지번별 토지 특성’ 데이터를 이용해 경기도 지역을 대상으로 침수 위험을 모형화 하였다. 또한, 침수 취약성이 높은 지하 공간과 모형 검증을 위해 침수 피해 변수를 추가로 사용하였다. 변수는 직접 위험 요인과 간접 위험 요인으로 구분했으며, 인자 분석과 정준 상관 분석을 통해 변수의 가중치를 반영한 직접 취약성 지수와 간접 취약성 지수를 산출하였다. 두 지수를 결합하여 최종 침수 취약성 지수를 산출하였고 취약성 지수의 활용성을 높이기 위하여 위험 수준에 따라 Stanine 점수로 재구분 하였다. 실제 침수여부를 전혀 사용하지 않고 구축된 침수 취약성 지수의 예측능력을 확인하기 위하여 과거 침수 피해 데이터를 이용하여 유의미한 예측능력을 확인하였으며, 모형의 확장성을 검증하기 위하여 전북특별자치도의 데이터에 적용하여 예측 결과를 확인하였다. 이를 통해 본 연구의 모형이 다양한 지역에서 적용 가능함을 입증하였다. 따라서 본 연구는 침수 피해 데이터가 존재하지 않는 지역에서도 침수 피해 위험을 효과적으로 예측하고 예방할 수 있는 방법론을 제시한다.

주요 용어 : 정준 상관 분석, 침수 피해 예측, 침수 취약성 지수, Stanine 점수.

¹30019 세종시 세종로 2511, 고려대학교 빅데이터사이언스학부. E-mail: pey5060@korea.ac.kr

²30019 세종시 세종로 2511, 고려대학교 경제통계학부 국가통계전공. E-mail: dkstnrud109@korea.ac.kr

³30019 세종시 세종로 2511, 고려대학교 빅데이터사이언스학부. E-mail: cbskust@korea.ac.kr

Impact of Global Fertilizer Prices on Import Allocation

*Joonsung Whangbo*¹, *Yeon Soo Noh*², and *Dong Hee Suh*³

Abstract

Climate change and global political instability have increased fertilizer price volatility, pushing prices to record highs in 2022. Rising global prices have driven up import costs, resulting in higher farming costs, reduced farmer incomes, and weakened food security. Given agriculture's heavy reliance on imported fertilizers, these price surges may pose significant threats to the sector. Addressing the lack of past studies on the economic impacts of global fertilizer price fluctuations, this study examines the effects of global fertilizer price hikes on import demand and provides insights for optimal import strategies by addressing key risk and structural factors in the fertilizer market.

Keywords : fertilizer, nitrogen, import demand, price risk.

¹Department of Food and Resource Economics, Korea University, Seoul 02841, Korea. Graduate Student.
E-mail: fangbojordan@korea.ac.kr

²Department of Food and Resource Economics, Korea University, Seoul 02841, Korea. Graduate Student.
E-mail: noh4990@korea.ac.kr

³(Corresponding Author) 145 Anam-ro, Seongbuk-gu, Department of Food and Resource Economics, Korea University, Seoul 02841, Korea. Associate Professor. E-mail: dhsuh@korea.ac.kr

B형 간염 환자의 간암 발병 위험 조기 판별 모형

서준호¹, 남진현²

요약

간세포암(Hepatocellular carcinoma, HCC)은 높은 사망률을 보이는 질환으로, 그 주요 원인 중 하나는 만성 B형 간염이다. 만성 B형 간염 환자에서 간암 발생을 조기에 예측하는 것은 중요한 임상적 과제로 본 연구에서는 이러한 간암 발생 및 합병증을 예측하기 위한 모형을 제시하고자 한다. 본 연구는 건강보험심사평가원의 건강보험 청구 데이터를 기반으로 만성 B형 간염 환자의 간암 합병증 예측 모형을 개발하였고, 주요 예측 인자로는 진단 상병 자료를 활용하였다. Random forest, penalized logistic regression, 그리고 extreme gradient boosting 방법을 활용하여 예측 성능을 비교하였다. 환자 군 간 동질성을 확보하기 위해 성향 점수(propensity score) 매칭을 적용하고, 클래스간 불균형 문제를 해결하기 위해 부트스트랩과 언더 샘플링 기법을 적용하였다. 비교 모형 중 random forest 모형이 가장 우수한 성능을 보였다. 본 연구는 만성 B형 간염 환자의 간암 합병증을 조기에 예측할 수 있는 기계학습 기반 예측 모형을 제시하였으며, 예측 모형을 위해 진단 상병만을 활용하였다. 따라서, 이 외에도 시술 및 처방약 정보를 추가로 반영한다면 모형의 성능을 더욱 향상시킬 수 있을 것으로 기대된다.

주요 용어: B형 간염, 간암, 기계학습, 발병 예측, 주요 질병 확인.

¹30019 세종특별자치시 세종로 2511, 고려대학교 일반대학원 빅데이터사이언스학과 박사과정.

E-mail: tjwnsg0223@korea.ac.kr

²(교신 저자) 세종특별자치시 세종로 2511, 고려대학교 빅데이터사이언스학부 부교수.

E-mail: jinhnam@korea.ac.kr

농산물 재배조사를 위한 표본설계*

손창균¹, 홍기학², 이기성³

요 약

본 논문에서는 2021년 시범조사를 바탕으로 2022년부터 실시하고 있는 농진청의 “수요자 맞춤형 품종육성지원 기술개발을 위한 작목 품종별 재배현황 통계구축연구”의 일환으로 보리 등 11종의 식량작물에 대해 농가별 재배면적과 재배농가에 대한 통계를 과학적으로 산출하여 농가에 보급중인 작물품종에 대한 기초자료를 생성하기 위한 표본설계내용을 제시하였다. 2024년 전체 조사대상 11개 작목 중에서 주요 식량작물인 6종의 작물에 대해 2023년 기준 농업 경영체 데이터베이스를 모집단으로 고려하여, 각 작물별 재배농가 규모를 기반으로 주요 작물생산지역을 선정하고, 주요 산지별로 재배면적에 따라 층화하였다. 각 작물별로 재배 면적에 대해 전수 층과 표본 층으로 구분하였으며, 표본규모는 약 1,500개 농가를 비례배분, 네이만배분, 멉등배분 등을 고려하여 층별로 기대목표오차 값이 안정적으로 나타난 멉등배분을 적용하여 표본 농가를 배분하였다. 1차 년도와 비교하여 변경된 사항은 세부 품종별 추정을 안정적으로 수행하기 위해 표본규모를 확대하였으며, 주요 생산지역과 전수층의 규모를 조정하였다. 또한 가중치 조정과정에서는 사후 층화 가중치 조정과정에서 설계단계에서 파악되지 않은 농가가 표본조사과정에서 새롭게 수집된 농가정보를 반영하기 위해 가중치를 부여하는 방법을 제안하였다.

주요 용어 : 농경제DB, 네이만 배분, 멉등 배분, 가중치 조정, 사후층화조정.

*이 논문은 농진청과제 “수요자맞춤형품종육성지원 기술개발을 위한 작목, 품종별 재배현황 통계구축 개선방안연구”를 위한 표본설계방안을 정리 요약한 내용임.

¹(교신저자) 38066 경북 경주시 동대로 123 동국대학교 빅데이터·응용통계학과 교수.

E-mail: ckson85@dongguk.ac.kr

²58245 전남 나주시 건재로 185 동신대학교 컴퓨터학과 교수. E-mail: khong@dsu.ac.kr

³55338 전북 완주군 삼례읍 삼례로 443 우석대학교 아동사회복지학부 교수. E-mail : gisung@woosuk.ac.kr

한국의 교통문화 설문응답에 대한 분석*

신지성¹, 양대원²

요약

경찰청에서 제공한 자료에 따르면 우리나라의 교통사고로 인한 부상자와 사망자는 2014년부터 해가 지날수록 줄어드는 경향을 보이고 있다. 하지만 경제협력개발기구(OECD) 국제교통포럼(ITF)에서 국제도로교통사고 데이터베이스(IRTAD) 워킹그룹에 참여하는 34개 회원국의 교통사고를 비교 및 분석한 '도로 안전 연간 리포트'에서 우리나라의 전반적인 교통안전 수준은 매년 OECD 회원국 중 하위권에 위치하고 있다.³ (<https://www.kyeonggi.com/article/20230109580237>) 교통사고의 특성상 대부분의 사고가 잠깐의 실수 또는 안 좋은 운전습관에 의해서 발생하기 때문에 올바른 교통문화를 지키는게 중요하다. 본 연구에서는 Microdata Integrated Service(MDIS)에서 제공하는 운전습관에 대한 교통문화실태조사 중에서 운전자들의 설문응답을 기록한 데이터와 실제 관찰한 데이터의 차이점을 비교해보고 운전습관을 예측할 수 있는 요인들에 대해서 여러가지 분석방법들로 분석할 것이다. 추가로 설문응답 데이터와의 차이점과 MAR상태로 가정한 NA값에 대한 분석을 통해 데이터의 특징에 대해서도 다룰 것이다.

주요 용어 : 교통문화, 설문응답.

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

¹(교신저자) 34134 대전광역시 유성구 대학로 99, 충남대학교 통계데이터사이언스학과 석사과정. E-mail: seong413@naver.com

²34134 대전광역시 유성구 대학로 99, 충남대학교 정보통계학과 교수. E-mail: summet73y.da@gmail.com

³지윤석, 2023.01.09. 세계에서 우리나라 교통안전 수준은, <https://www.kyeonggi.com/article/20230109580237>

Testing for Bubbles in the Korean won - US dollar Exchange Rate

*Sunghoon Sim*¹

Abstract

Our empirical findings from SADF and GSADF tests indicate that the explosive behavior in the nominal won-dollar exchange rate can be attributed to the movements in both the relative prices of traded and non-traded goods. Graphical results also show evidence of three speculative bubbles. The first bubble is related to the Korean currency crisis in 1997, the second bubble coincides with the period of world financial crisis in 2008, and the third bubble occurs around the year of 2022 from which period the interest disparity between Korea the US started to expand.

Keywords : Exchange rate, traded goods, non-trade goods, GSADF test.

1. Introduction

Economic theory has postulated that exchange rates paly crucial roles in determining international trade and, thus, economic growth. Financial theory posits that there is a nexus between exchange rates, inflation and interest rate differentials, which contribute to the direction of capital and asset flows between countries, and thus determine asset prices. Therefore, the fact that exchange rates can freely float, no doubt, increases speculation in asset prices, increasing volatility in the financial markets. The existence of rational bubbles suggests that nominal exchange rate and prices are not related in the long run. Traditional theory attributes fluctuations in real exchange rates to change in relative price of non-traded goods.

Following the breakdown of the Bretton-Woods system of fixed exchange rates in the early 1970s, major developed countries switched from a fixed into a floating exchange rate regime. Korea also adopted fully free-floating exchange rate system after 1997 economic crisis. History has witnessed many episodes of crises in the won-dollar exchange market, such as the 1997 Asian crisis, the 2008 financial crisis, and the strong depreciation in recent few years. Dramatic depreciation of the won-dollar rate during these crisis periods has puzzled practitioners as well as researchers. Some economists conjecture that speculative bubbles were driving the market during these periods.

¹Professor, Department of Economics, Pukyong National University, Yongso-ro 45, Nam-gu, Busan, Korea.
E-mail: shsim@pknu.ac.kr

Recently, various new tests have been developed to detect speculative bubbles in asset prices. We employ the sequential unit root tests proposed by Phillips et al. (2011, 2015), which are based on the type of indirect stationarity tests initiated by Diba and Grossman (1984) and Hamilton and Whiteman (1985). These indirect tests have the advantage of detecting speculative bubbles despite a potential misspecification of the market fundamental process. With these backdrops, this paper explores the existence of multiple bubbles in Korean won-US dollar.

2. Data and Empirical Results

The data used in this study consist of monthly bilateral exchange rates between Korea and the United States from January 1995 through August 2023. The data on nominal exchange rates were obtained from the Bank of Korea. The nominal exchange rate series were converted to natural logarithms.

Following Engel (1999), the ratio of the exchange rate to the traded goods fundamental is given by,

$$s_t^{r,T} = s_t + \log(PPI_t^*) - \log(PPI_t), \quad (1)$$

where s_t is the logarithms of won/dollar exchange rate, PPI is the producer price index and the asterisk represents the foreign counterpart. Equation (1) measures the direct relative effect of traded goods on exchange rate. On the other hand, equation (2) shows the effect of non-traded goods on exchange rate.

$$s_t^{r,N} = s_t + \log(PPI_t^*) - \log(PPI_t) + \log(CPI_t^*) - \log(CPI_t), \quad (2)$$

where CPI is the consumer price index and the asterisk represents the foreign counterpart.

Table 1 displays the SADF and GSADF test statistics for the logarithm of nominal exchange rate between Korea and the United States. The results of SADF and GSADF tests show that the null hypothesis of no explosive behavior in the won-dollar exchange rate is rejected at 1% significant level.

Table 1 also provides the results from SADF and GSADF procedures for exchange rate relative price of non-traded goods ratio. The right-tailed ADF tests fail to reject the null of no speculative bubbles in the relative prices non-traded goods at the 1% significant level. Likewise, the results of Table 1 indicates that there exist the bubbles in the relative prices of traded goods at the 1% significant level. In sum, our empirical results of SASF and GSADF tests presents evidence of multiple bubbles in the nominal won-dollar exchange rate; driven by both the relative prices of non-traded and traded goods between Korea and the United States.

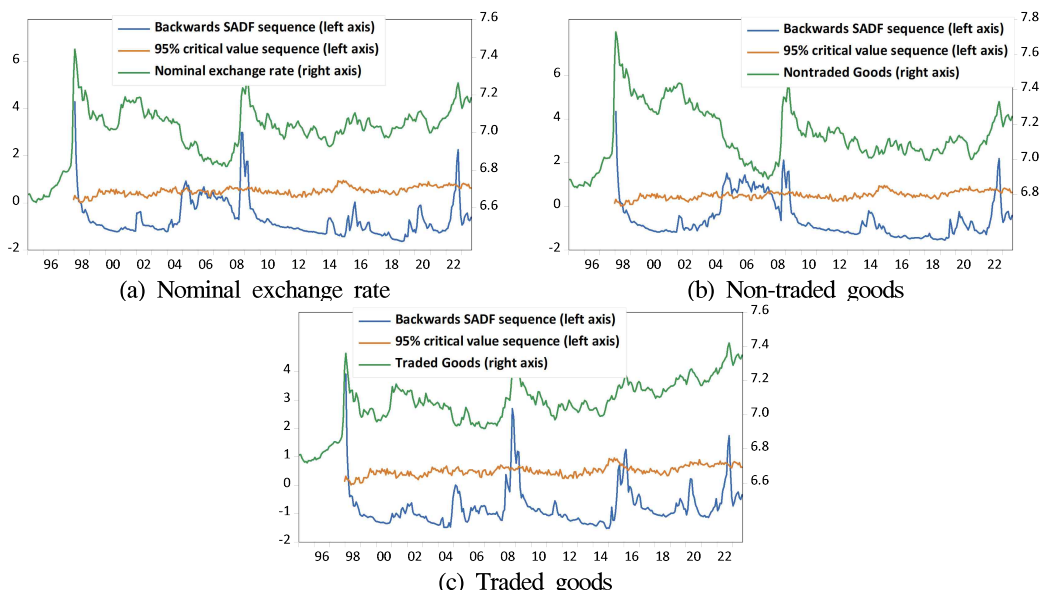
Once we have established the existence of bubbles in the exchange rate series, the next step is to identify the locations of bubble periods graphically using BSADF procedure. The

graph (a) of Figure 1 plots the GSADF statistics for the nominal won-dollar exchange rate. The figure shows evidence of multiple bubbles in the nominal exchange rate between Korea and the US. The first bubble occurred in end of 1997. This period corresponds to the currency crisis in Korea and to the period of managed floating exchange rate regime in Korea. The second bubble occurred around the year of 2008 which period corresponds to the world financial crisis and to the period of fully floating exchange rate regime in Korea. The third bubble occurred in 2022 from which period the interest disparity between Korea and the US started to largely increase.

Table 1. Results of right-tailed tests

Variable	SADF	GSADF
	Test statistics	Test statistics
s_t	4.2971***	4.2971***
$s_t^{r,N}$	4.3413***	4.3413***
$s_t^{r,T}$	3.9151***	3.9151***
Critical values (%)		
99	1.9580	2.6006
95	1.5116	2.0675
90	1.1636	1.9488

Note: *** indicates 1% significance level.



Note: The solid (or blue) lines denote the backward SADF (BSADF) sequences (left axis basis) and the dotted (or red) lines denote 95% critical value sequences (left axis basis). Top dashed (or green) lines indicate the three exchange rates (right axis basis).

Figure 1. Results of the GSADF date-stamping procedure for exchange rates

The graph (b) of Figure 1 displays graphically the results of the GSADF test for real exchange rate of non-traded goods. The graph is similar to the GSADF result of nominal exchange rate, showing the multiple bubbles with the additional bubble between 2005 and 2007. The graph (c) of Figure 1 also provides similar bubble pattern to the nominal exchange rate with exception of another short bubble between 2015 and 2016.

3. Conclusion

In this paper we provide new evidence casting doubt on the bubble hypothesis in the nominal won-dollar exchange rate by employing recent sequential ADF tests developed by Phillips et al. (2011, 2015).

Our empirical findings from SADF and GSADF tests indicate that the explosive behavior in the nominal won-dollar exchange rate can be attributed to the movements in both the relative prices of traded and non-traded goods. This result is not consistent with the traditional theories, which stresses that movements in relative price of non-traded goods alone are responsible for the fluctuations in the nominal exchange rates. Graphical results also show evidence of three speculative bubbles. The first bubble is related to the Korean currency crisis in 1997, the second bubble coincides with the period of world financial crisis in 2008, and the third bubble occurred around the year of 2022 from which period the interest disparity between Korean and the US started to expand.

The results of this study underscore the important roles of underlying fundamentals in determining rational bubbles in asset prices. The existence of rational bubble is likely lead to the redistribution of wealth so that the investors also should be cognizant of the existence of rational bubbles in won-dollar exchange rate. Moreover, asset bubbles can have negative financial and real consequences. Thus, the policymakers need to be aware of the early warning sings of speculative bubbles so that they can mitigate the reverse effect of bubbles by implementing appropriate monetary and fiscal policies.

References

- Diba, B., Grossman, H. (1984). Rational bubbles in the price of gold. NBER Working Paper No. 1300.
- Engel, C. (1999). Accounting for US real exchange rate changes. *Journal of Political Economy*, 107 (3), 507 - 538.
- Hamilton, J., Whiteman, C. (1985). The observable implications of self-fulfilling expectations. *Journal of Monetary Economics*, 16 (3), 353 - 373.
- Phillips, P. C. B., Shi, S., Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500, *International Economic Review*, 56, 1043-1078.
- Phillips, P. C. B., Wu, Y., Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset value?, *International Economic Review*, 52, 201-226.

현대 학생들의 과도한 사회화 연구

오열¹,

요약

시대의 급속한 발전과 변화에 따라 사람들의 물질적, 정신적 생활 수준이 크게 향상되었다. 동시에 사회의 가치관 시스템에도 큰 변화가 일어나고 있다. 오늘날 초등학생의 사회화 상황은 큰 변화를 겪었고 일부 초등학생은 점점 더 어른스러워지고 있으며 많은 성인화를 볼 수 있으며 과도한 사회화의 징후가 나타나고 성인화 경향이 점점 더 심각해지고 있다. 초등학교 단계의 사회화 발전은 개인의 사회화 발전의 기초이며 미래의 개인의 사회화 발전에 중요한 영향을 미친다. 본 연구에서는 교육사회학의 관점에서 문헌연구법, 설문조사법, 인터뷰법을 이용하여 초등학생의 과도한 사회화를 종합적으로 이해하고 연구를 수행하였다. 먼저 사회화 내용에 초등학생의 생활기능, 생활목표, 가치관, 행동규범, 사회적 역할 등 5가지 차원에 대한 설문지를 작성하여 초등학생의 사회화 실태를 조사하였다. 조사 결과 현재 일부 초등학생은 과도한 사회화 상황을 가지고 있으며, 성별과 가정 배경이 다른 초등학생은 사회화 상황에 확실히 차이가 있는 것으로 나타났다. 그다음 사회적 역할 이론, 참조 집단 이론, 사회 학습 이론 및 조사, 인터뷰 자료를 결합하여 초등학생의 과도한 사회화의 원인을 분석한다. 따라서 초등학생의 과도한 사회화가 초등학생 본인과 사회, 국가의 미래 발전에 부정적인 영향을 미친다는 견해를 제시하여 더 많은 사람들이 초등학생의 과도한 사회화 현상에 주목하기를 바란다. 초등학생의 과도한 사회화를 연구하는 것은 중요한 현실적 의의가 있다: 1. 현재 어린이와 어린이의 교육 환경에 존재하는 몇 가지 단점을 개선하는 데 도움이 된다. 2. 교육 분야의 다른 유사한 문제에 대한 참조 및 참조를 제공한다. 3. 현재 사회 발전의 위기와 숨겨진 위험을 깊이 이해하고 우리 사회의 정신 문명 프로세스를 개선하는 데 도움이 된다. 논문은 주로 다음을 포함한다: 1. 서론: 이 부분은 주로 주제 선정 이유, 연구 의의, 핵심 개념의 정의, 연구 검토 및 연구 설계를 포함한다. 2. 현황분석: 데이터 분석을 이용하여 기술통계분석을 통해 초등학생의 과도한 사회화 현황을 파악하고, 초등학생의 과도한 사회화에 대한 차이분석을 추가로 수행하였다. 3. 원인 분석: 이 부분은 주로 사회화의 영향요인을 기반으로 하며, 조사 및 인터뷰 자료를 결합하여 초등학생의 과도한 사회화에 영향을 미치는 요인을 심층 분석하여 초등학생의 과도한 사회화에 대한 대책 연구를 기반으로 한다. 4. 영향 분석: 이 부분은 주로 초등학생의 과도한 사회화가 초등학생 자신, 사회 및 국가의 미래 발전에 미치는 부정적인 영향을 분석하고 부모, 학교 및 사회가 초등학생의 과도한 사회화에 주의를 기울일 것을 촉구한다. 5. 대책 연구: 초등학생의 과도한 사회화를 초래하는 영향 요인을 대상으로 한다.

주요 용어: 초등학생, 과잉사회화, 교육사회학.

간호대학생의 학업지연행동에 미치는 영향 요인

이은미¹, 오윤정²

요 약

본 연구는 간호대학생의 학업소진, 스마트폰 중독 및 학업지연행동 간 관계를 파악하고 학업지연행동에 미치는 영향을 확인하기 위해 시도된 서술적 조사연구이다. 연구대상자는 G광역시 소재 대학의 간호학과에 재학 중인 학생 195명으로, 자료수집은 자가보고식 설문지를 이용하여 2024년 4월 10일에서 4월 25일까지 실시하였다. 수집된 자료는 SPSS 25.0 program을 이용하여 기술적 통계, t-test, ANOVA, Pearson correlation coefficient, 다중회귀분석을 시행하였다. 연구결과 전체 연구 대상자는 193명이며, 평균 연령은 20.76±1.67세였다. 대상자의 일반적 특성에 따른 학업지연행동을 분석한 결과 거주형태($F=5.700, p=.004$), 하루 스마트폰 사용시간($F=6.255, p=.002$), 스마트폰 사용기간($F=6.411, p=.002$), 스마트폰 사용목적($F=3.439, p=.018$)에 따라 통계적으로 유의한 차이가 있었다. 학업지연행동은 학업소진($r=.387, p<.001$), 스마트폰 중독($r=.414, p<.001$)과 유의한 상관관계를 보였고, 학업소진은 스마트폰 중독($r=.544, p<.001$)과 유의한 상관관계를 보였다. 학업지연행동에 영향을 미치는 요인은 스마트폰 중독($\beta=.295, p<.001$)과 학업소진($\beta=.186, p=.002$)인 것으로 나타났고, 총 설명력은 17.0%였다. 따라서 간호대학생의 스마트폰 중독과 학업소진을 줄일 수 있는 교육프로그램을 개발하여 간호교육과정 동안 지속적으로 제공함으로써 학업지연행동을 감소시킬 수 있는 전략이 필요하다.

주요 용어 : 간호대학생, 학업소진, 스마트폰 중독, 학업지연, 중독.

1. 서론

누구나 한 번쯤은 지금 해야 할 일들을 잠시 후로 미루고, 또 오늘 해야 할 일들을 나중에 미룬 경험이 있을 것이다. 이러한 지연행동은 일상생활에서 흔히 관찰할 수 있는 행동으로, 지연행동 중에서도 과제 제출 기간이 지나서야 겨우 과제를 내거나 시험준비를 미루다가 시험 전날에 벼락치기로 공부하는 등 학업 상황에서 발생하는 지연행동을 학업지연행동이라고 한다(Solomon & Rothblum, 1984). 이러한 학업지연행동은 초등학생에서 성인에 이르기까지 다양한 연령대에서 나타나는데(박승호, 서은희, 2005), 특히, 대학생은 중, 고등학교 때보다 비교적 자율적인 분위기에서 학업을 수행하기 때문에 학업지연행동이 나타날 가능성이 높다(유지원, 강명희, 김은희, 2013; Klingsieck, Fried, Horz, & Hofer, 2012). 이에 본 연구는 간호대학생을 대상으로 학업지연행동에 영향을 미치는 요인을 확인하고, 학업지연행동을 감소시키기 위한 전략개발의 기초자료로 활용하고자 한다.

¹광주광역시 광산구 첨단중앙로 23, 남부대학교 간호학과 조교수. E-mail: jjang9312@nambu.ac.kr

²(교신저자) 광주광역시 광산구 첨단중앙로 23, 남부대학교 간호학과 조교수. E-mail: agnes@nambu.ac.kr

2. 연구방법

본 연구는 간호대학생의 학업소진, 스마트폰 중독 이 학업지연행동에 미치는 영향을 확인하기 위한 서술적 조사연구이다. 대상자 수는 G*Power 3.1.9 프로그램을 이용하여 효과크기 .15, 유의수준 .05, 검정력 .95, 관련 변인의 수를 9개로 설정하여 회귀분석에 필요한 최소 표본 수를 산출한 결과 166명이었고, 탈락률을 고려하여 총 200부를 배부하였다. 회수된 설문지 중 응답이 불충분하거나 누락된 5부를 제외하고 총 195부를 자료 분석에 이용하였다. 자료수집은 2024년 4월 10일부터 4월 25일까지 자가보고식 설문지를 활용하여 이루어졌다.

3. 연구결과

3.1 일반적 특성에 따른 학업지연행동

본 연구에서 대상자의 일반적 특성과 일반적 특성에 따른 학업지연행동은 Table 1과 같다. 전체 연구 대상자는 193명이며, 평균 연령은 20.76±1.67세였다. 대상자의 일반적 특성에 따른 학업지연행동을 분석한 결과 거주형태($F=5.700, p=.004$), 하루 스마트폰 사용시간($F=6.255, p=.002$), 스마트폰 사용기간($F=6.411, p=.002$), 스마트폰 사용목적($F=3.439, p=.018$)에 따라 통계적으로 유의한 차이가 있었다.

3.2 대상자의 학업소진, 스마트폰 중독, 학업지연행동 간의 상관관계

대상자의 학업소진, 스마트폰 중독, 학업지연행동 간의 상관관계는 Table 2와 같다. 학업소진과 스마트폰 중독($r=.414, p<.001$), 학업지연행동($r=.387, p<.001$) 간에는 유의한 정적 상관관계가 있었고, 스마트폰 중독과 학업소진($r=.544, p<.001$) 간에도 유의한 정적 상관관계가 있었다.

Table 1. Differences of Academic procrastination behavior by general characteristics (N=195)

Characteristics	Categories	n(%) or M(SD)	M(SD)	t or F(p) Scheffe
Age(year)	range 19-28	20.76±1.67		
Gender	Male	33(16.9)	2.51(0.50)	-2.017
	Female	162(83.1)	2.70(0.48)	(.050)
Type of residence	With familya	90(46.2)	2.60(0.49)	5.700*
	Living aloneb	41(21.0)	2.89(0.52)	(.004)
	Dormitoryc	64(32.8)	2.62(0.43)	b>a
Smartphone usage time per day	1-3hra	17(8.7)	2.69(0.39)	6.255*
	3-5hrb	67(34.4)	2.50(0.53)	(.002)
	More than 5hrc	111(56.9)	2.77(0.46)	c>b
Period of use of smartphone	1-3yra	2(1.0)	2.79(0.00)	6.411*
	3-5yrb	2(1.0)	1.47(0.00)	(.002)
	More than 5yrc	191(98.0)	2.68(0.48)	a,c>b
Purpose of using smartphone	Search for informationa	10(5.1)	2.26(0.51)	
	Interpersonal relationshipb	106(54.4)	2.74(0.46)	
	Watch the videoc	65(33.3)	2.63(0.47)	3.439*
	Basic Features (Call, text message)d	14(7.2)	2.58(0.66)	(.018)

Table 2. Correlations among Academic burnout, Smartphone addiction and Academic procrastination behavior (N=195)

Variables	Academic burnout r(p)	Smartphone addiction r(p)	Academic procrastination behavior r(p)
Academic burnout	1		
Smartphone addiction	.544 ($<.001$)	1	
Academic procrastination behavior	.387 ($<.001$)	.414 ($<.001$)	1

Table 4. Influencing Factors on Academic procrastination behavior (N=195)

Variables	B	SE	β	t	p
	1.544	0.19		8.20	$<.001$
Smartphone addiction	0.28	0.08	.295	3.69	$<.001$
Academic burnout	0.16	0.68	.186	2.32	.002

Adj R2=.170, F=19.437, $p<.001$

3.4 대상자의 학업지연행동에 영향을 미치는 요인

대상자의 학업지연행동에 영향요인을 확인한 결과는 Table 3과 같다. 본 연구에서 대상자의 학업지연행동의 가장 중요한 가장 중요한 예측요인은 스마트폰 중독($\beta=.295$, $p<.001$)이었고, 그 다음으로는 학업소진($\beta=.186$, $p=.002$)인 것으로 나타났다.

4. 논의

본 연구는 간호대학생의 학업소진, 스마트폰 중독 및 학업지연행동 간 관계를 파악하고 학업지연행동에 영향을 미치는 요인을 확인하기 위해 시도되었다. 본 연구결과를 토대로 간호대학생들의 학업소진, 스마트폰 중독, 학업지연행위를 고려한 간호교육과정 개발과 학업지연행위를 줄이기 위한 교육프로그램의 개발 및 효과검증이 이루어져야 할 것을 제언한다.

References

- Choi, E. Y., & Shin, N. M. (2020). The Relationship Between SNS fatigue and university students' behaviors of academic procrastination and learning flow. *Journal of Knowledge Information Technology and Systems*, 15(3), 373-382.
- Klingsieck, K. B., Fries, S., Horz, C., & Hofer, M. (2012). Procrastination in a distance university setting. *Distance Education*, 33(3), 295-310.
- Park, S. H. & Seo, E. H. (2005). Women college students' academic procrastination. *Korean Journal of Educational Research*, 43(2), 115-134.
- Solomon, L. J., & Rothblum, E. D. (1984). Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of Counseling Psychology*, 31, 503-509.
- You, J. W., Kang, M. H., & Kim, E. H. (2013). A relationship among task value, academic self-efficacy, motivation, self-regulated learning and academic procrastination in a college e-learning course. *The Journal of Korean Association of Computer Education*, 16(1), 81-95.

중국대학생의 전공선택과 직업선택의 충돌 문제

오적¹, 양정호²

요약

중국 경제의 지속적인 발전, 산업 구조의 전환 및 고등 교육의 지속적인 발전으로 인재 수요 구조는 큰 변화를 겪었고 대학생의 고용 상황은 점점 더 심각해졌으며 고용 문제는 사회의 모든 측면에서 광범위한 관심을 끌었다. 대학생의 전공 선택과 직업 선택의 충돌 문제에 대한 연구는 졸업생들이 위치한 직책의 대학 교육에 대한 진정한 요구를 깊이 이해하는 데 도움이 될 것이며 대학 교육 조정에 큰 의미가 있다. 중국 대학생의 고용 질에 대한 관련 연구에 따르면 대학생의 직업과 전공의 상관관계는 66%에 불과하며 전공 선택과 직업 선택이 충돌하는 현상이 만연한다. 대학생의 경력개발에 대한 평가연구에 따르면 전공선택과 직업선택의 충돌은 대학생의 취업의 질, 취업만족도, 취업안정성에 중요한 영향을 미치며, 심지어 일부 대학생의 잠재성 또는 현재성 실업으로 이어져 대학생 자신, 가족 및 경제 사회 발전에 많은 부정적인 영향을 미친다. 많은 문헌자료의 열람을 바탕으로 구체적인 자료를 바탕으로 문제발견, 문제분석, 문제해결의 사고방식을 바탕으로 대학생의 전공선택과 직업선택의 충돌문제를 체계적으로 연구하고 발견하며, 전공선택과 직업선택의 충돌로 이어지는 심층적인 원인을 객관적으로 분석하여 취업압박 완화, 현재의 심각한 취업상황 개선, 대학생 취업의 질 향상에 참고가 된다. 첫째, 대학생의 전공선택과 직업선택의 충돌현상을 분석하고, 학생의 전공선택과 직업선택의 분석을 통해 전공선택과 직업선택의 상호영향, 상호제한 및 촉진의 관계를 설명하고, 둘째, 2018년부터 올해까지 전국 대학생 취업의 업종별 변화 추이와 취업지역의 데이터를 바탕으로 우리나라 대학생들의 전공선택과 직업선택의 충돌현황을 분석하였다. 대학생의 전공선택과 직업선택의 충돌 원인에 대한 상관분석을 수행한다.

주요 용어 : 대학생, 전공선택, 직업선택, 갈등.

1. 서론

전공 선택은 대학 캠퍼스에 입학하는 학생들이 대학 단계에서 고등 교육을 받는 핵심 단계 중 하나이며 향후 4년 동안 공부할 전공을 선택하는 것이다. 전공 선택은 능력 배양과 성격 특성 형성에 영향을 미칠 뿐만 아니라 미래에 어떤 직업에 종사할지를 선택하는 중요한 근거이기도 한다. 수능 후 지원에서 전공을 선택한 뒤 전공을 바꿀 확률은 매우 낮다. 대학에 입학한 뒤 당초 선택한 전공이 정작 배우고 싶은 내용이 아니라는 사실을 알게 되면 많은 학생들이 학습 의욕을 떨어뜨린다. 원하는 학과에 진학하지 못하면 대학생들은 공부에 흥미를 잃어 장래 취업에 지장을 줄 수 있다.

¹03063 서울특별시 종로구 성균관로 25-2, 성균관대학교 교육학과 석사과정. E-mail: giggle0705@gmail.com

²(교신저자) 03063 서울특별시 종로구 성균관로 25-2, 성균관대학교 교육학과 교수. E-mail: jyang@skku.edu

21세기 초 중국의 고등교육이 대중화 단계에 진입한 이후 중국의 고등교육 발전규모가 급속히 확대되어 2022년 중국의 대학졸업생 규모가 처음으로 1000만 명을 넘어섰고, 이렇게 많은 대학졸업생들은 자기선택을 통해 취업해야 하며, 취업난은 대학생들이 졸업 후 직면해야 하는 중요한 현실이 되었다. 대학생들의 취업난은 한편으로는 졸업생 공급이 대학생에 대한 사회적 수요를 초과하기 때문, 즉 대학생의 수가 일자리 수를 초과하고, 제공할 수 있는 일자리가 충족된 후에도 대졸자는 여전히 일부 남아 있다. 한편, 대학생의 전공은 직장에서 제공하는 직업과 일치하지 않으며, 이는 대학생의 전공 선택과 직업 선택의 충돌로 나타난다. 전공선택과 직업선택의 충돌은 다음과 같은 두 가지 측면에서 반영되는데, 하나는 대학생이 전공과 일치하는 직업을 찾지 못하고, 다른 하나는 회사가 원하는 직책에 적합한 대학생을 구하지 못하는 것이다. 따라서 대졸자의 전공선택과 직업선택의 충돌은 대학생들의 취업문제로 이어질 수 있다. 극심한 취업 압박으로 인해 부정적인 생각을 갖고 전공과 직업을 무시하는 대학생들이 점점 더 많아지고 있으며, 먼저 취업한 다음 직업을 바꾸고, 먼저 일자리 문제를 해결한 다음 향후 발전을 고려하고 있다. 이러한 태도로 인한 높은 취업률은 대학생의 취업난과 실업 문제를 은폐하고 있으며, 이러한 행동은 졸업생들이 직장에서 안정적이지 못하고 대학생의 취업의 질과 장기적인 발전을 보장할 수 없다.

2. 대학생의 전공선택과 직업선택의 충돌 표현

동부 지역은 인재 흡인력이 강하고 중부와 동북 지역은 졸업생 유출이 많다. 학부 졸업생 취업 지역 분포는 2022년 동부지역(51.9%)이 가장 높고 서부지역(24.6%)이 뒤를 잇고 있으며, 지역별 졸업생 비율과 졸업 후 리얼 취업을 종합하면 동부지역(졸업생 비율 38.5%, 졸업 후 리얼 취업 비율 88.0%)이 가장 강하고 졸업생 유입이 많으며 중부지역(졸업생 비율 26.7%, 졸업 후 리얼 취업 비율 85.2%)과 동북지역(졸업생 비율 9.4%, 졸업 후 리얼 취업 비율 84.4%)의 졸업생 유출이 많다 (Table 1참조).

마이클스의 중국 산업 분류 시스템에 따르면 이번 평가는 졸업생이 고용된 327개 산업을 대상으로 한다. 교육 및 건설 산업에서 졸업생의 고용 비율이 크게 감소하고 정보 기술 산업에 대한 수요가 감소하고 장비 제조 및 에너지 공급 분야에서 수요가 증가했다. 졸업생 취업 업종별 비율을 보면 2022년 졸업생이 반년 만에 가장 많이 취업한 업종은 '교육업'(13.0%)이었고, '정보전송·소프트웨어·정보기술 서비스업'(8.7%), '건설업'(7.7%), '금융업'(7.3%) 등이 뒤를 이었다. 2022학년도 학부 졸업생 중 교육훈련기관 취업률(5.1%)은 2021학년도(6.0%)에 비해 계속 낮아진 것이다. 이와 함께 '건설업'에서 졸업생 취업률도 크게 감소했는데, 특히 주택건설 분야는 2022년 3.3%로 5년 간 1.2%포인트 하락했고, 인프라 건설 분야는 안정세를 유지해 2022년 3.0%로 나타났다(Table 2참조).

Table 1. 2022년 졸업생 취업 지역 분포

각 지역	2022년 졸업생 취업 비율	2022년 졸업생 실제 인원수 비율	졸업 후 리얼 취업 비율
중국 동부 지역	51.9%	38.5%	88.0%
중국 서부 지역	24.6%	25.4%	85.9%
중국 중부 지역	19.7%	26.7%	85.2%
중국 동북 지역	3.8%	9.4%	84.4%

Table 2. 2018-2022년 졸업생 취업의 주요 업종별 변화 추세

업종별	2022년	2021년	2020년	2019년	2018년	5년간의 변화
교육업	13.0%	14.0%	17.0%	15.9%	14.9%	-1.9%
정보기술 서비스업	8.7%	9.2%	9.0%	8.9%	8.8%	-0.1%
건축업	7.7%	8.6%	9.0%	8.9%	9.1%	-1.4%
금융업	7.3%	7.2%	7.5%	7.8%	8.1%	-0.8%
정부 및 공공관리	6.8%	6.4%	6.2%	6.0%	6.0%	0.8%
전자전기제조업	6.8%	6.2%	5.7%	5.7%	5.6%	1.2%
의료 및 간호 서비스업	6.0%	6.0%	5.9%	6.0%	6.2%	-0.2%
디자인컨설팅서 비스업	5.0%	5.3%	5.1%	5.8%	5.5%	-0.5%
문화,체육,유흥업	4.5%	4.6%	4.2%	4.6%	4.2%	0.3%
소매업	3.6%	3.8%	3.5%	3.4%	4.0%	-0.4%
전기 및 물 생산과 공급업	3.0%	2.6%	2.4%	2.2%	1.6%	1.4%
기계설비제조업	2.9%	2.6%	2.5%	2.4%	2.4%	0.5%
병원설비제조업	2.7%	2.4%	2.1%	2.0%	1.8%	0.9%
운수업	2.2%	2.3%	2.2%	2.4%	2.2%	0%
화학공업	2.1%	1.8%	1.8%	1.8%	1.9%	0.2%
환경보호보조업	2.0%	1.7%	1.7%	1.9%	2.1%	-0.1%

3. 대학생의 전공선택과 직업선택의 충돌원인 분석

(1) 대학과 학생 간의 정보 비대칭

대학생은 전공을 선택할 때 대학이 대학생을 선택하는 과정이기도 하며 대학생과 대학은 시장 거래의 쌍방으로 간주될 수 있다. 대학입시가 끝난 후 졸업생은 자신이 취득한 성적과 각 학교의 합격점수와 흥미특기에 따라 자신의 학교와 전공을 선택하는데, 대학을 갓 졸업한 고등학교 졸업생은 학교와 전공에 대한 인식이 매우 제한적이며, 학교의 상황과 전공에 대한 인식이 정확하지 않고, 가족과 친구로부터 얻은 정보도 정확하지 않으며, 학교가 학생 모집의 관점에서 발표하는 정보는 기본적으로 학교에 유리하며, 기본적으로는 학생에게 불리한 요소를 밝히지 않고 홍보하기 때문에 학생과 학교 사이에 정보의 비대칭이 존재하여 학생들이 올바른 전공을 선택하는 데 도움이 되지 않는다. 동시에 현재 학교는 학생에 대한 평가 및 입학 기준은 대학 입시 점수이며 특기, 흥미 및 능력을 포함한 학생의 종합적인 자질에 대해서는 무지하다. 학생의 점수와 각 전공의 점수에 따라 높은 것에서 낮은 것으로 입학하여 학교와 학생 간의 정보 비대칭이 발생하고, 학교 전공에 의한 학생 입학은 향후 발전에 부정적인 영향을 미칠 수 있다.

(2) 사회와 대학 간의 정보 비대칭

대학교육의 목적은 사회에 적합한 인재를 양성하고 사회 발전의 요구를 지속적으로 충족시키는

것이므로 학교는 사회의 요구를 충분히 이해해야 한다. 그러나 학교로서 현재로서는 사회와 심각하게 동떨어져 있으며 학교는 학술연구와 이론학습을 중시하고 있으며, 사회발전을 위해서는 이론 지식 외에 응용형 인재가 더욱 필요하며, 대학 교사는 기본적으로 기업체에 근무하지 않고, 배운 이론지식도 실제와 동떨어져 학교가 사회적 필요에 따라 회사의 요구에 부합하는 인재를 양성할 수 없어 학교와 회사 간의 정보 비대칭이 발생한다. 마찬가지로 회사와 학교 간의 정보 소통이 부족하여 학교도 회사의 필요에 따라 전문 인재를 양성할 수 없다. 이 둘 사이의 정보 비대칭은 대학생의 취업에 부정적인 영향을 미치며 대학생의 전공 선택과 직업 선택의 갈등을 일으키는 중요한 요소이다.

(3) 대학과 학생 간의 정보 비대칭

한편, 교육의 발전은 경제 발전 수준에 의해 제한되고 과학 기술의 발전과 교육 발전에 뒤처져 있으며, 인재에 대한 사회의 요구를 충족시키는 것을 목표로 교육의 지연이 불가피하다. 교육의 중요한 단계로서 고등 교육은 자체의 지연을 피할 수 없다. 반면에 교육은 미래 지향적인 사업이며 교육의 일부로서 고등 교육은 더욱 그렇다. 학교에서 학생을 양성하는 데 최소 4년의 시간이 소요된다는 것은 교육 및 인재 양성 주기가 길고 결과가 느리다는 것을 충분히 보여주므로 교육의 발전 규모, 교육의 발전 수준, 교육 자금 투입 등을 잘 예측해야 하며, 현재의 사회 경제 발전 수준을 어느 정도 초과해야 하며, 이는 또한 교육의 선진성에 대한 요구 사항이기도 하다. 최근 몇 년 동안 많은 대학의 전공 설정이 시장의 요구와 학교 자체의 특성을 고려하지 않고 다양성과 종합성을 갖추는 경향이 있어 많은 대학 졸업생이 다양한 자격증을 가지고 있지만 취업 시장은 여전히 막연하여 대학생의 전공 선택과 직업 선택의 충돌이 발생한다. 전공 설정 상황과 직업 수요의 충돌은 두 가지 상태를 보여주는데, 하나는 사회적 요구에 뒤처진 전공 설정으로 인한 대학생의 공급 부족이고, 다른 하나는 전공 설정의 미래 지향성 부족으로 인한 대학생의 공급 과잉이다.

References

- Mycos Data Institute of Technology. (2023). China Undergraduate Employment Report 2023, *Society Social Sciences Literature Press*.
- John Robst, Jennifer Van Gilder. (2016). Salary and job satisfaction among economics and business graduates: The effect of match between degree field and job, *International Review of Economics Education*, 16(1), 30-40.
- Mingcheng Fan, Xiaowei Chen. (2009). Basic analysis of Chinese university students' major choice, *University Education Science*, (03), 91-96. (in China).

한국형 심방세동 위험 점수 모형 KAF 개발

윤지원¹, 남진현²

요약

요약심방세동(Atrial Fibrillation, AF)은 불규칙한 심장 박동을 특징으로 하는 부정맥 질환으로, 특히 고령 인구에서 발생률이 꾸준히 증가하고 있습니다. 한국에서 심방세동의 유병률은 최근 몇 년간 급증했으나, 서양 인구를 대상으로 개발된 기존의 위험 점수 모형은 한국인에게 정확한 예측을 제공하지 못하는 한계가 있습니다. 아시아 인구를 대상으로 한 모형이 존재하지만, 한국인을 대상으로 했을 때 그 성능은 제한적이었습니다. 본 연구는 국민건강보험공단의 표본 코호트 자료를 활용하여 한국인 맞춤형 심방세동 위험 점수 모형(KAF)을 개발하는 것을 목적으로 합니다. 연구 기간은 2007년부터 2019년이며, 건강검진을 받은 환자 중 건강검진 날짜 기준 이전 2년간의 청구 데이터를 통해 합병증과 동반질환을 확인했습니다. 기술통계, 유병률, 단변량 Cox 회귀분석을 통해 심방세동 발병에 주요한 위험 요인을 선택했고, 이후 다변량 Cox 회귀분석 수행 후 유의한 위험 요인을 통해 최종 위험비(HR)를 도출했습니다. 이를 바탕으로 상대적인 위험 점수를 계산하여 KAF 모형을 만들었습니다. 이 모형은 성별, 연령, BMI, 고혈압, 고지혈증 등 다양한 인자를 포함하고 있으며, 심방세동 발생 위험을 점수화하여 환자를 위험도별로 그룹화할 수 있음을 확인했습니다. KAF 모형은 기존에 제시된 모형보다 높은 분류 성능을 보여, 한국형 심방세동 위험 점수 모형으로 적합함을 보였습니다. KAF 모형은 환자 스스로 심방세동 위험 정도를 판단 가능하게 하고, 이에 따라 조기 진단 및 예방적 치료에 활용 가능하여 심방세동으로 인한 사회적 비용을 경감 할 수 있을 것입니다.

주요 용어: 심방세동, KAF 모형, 위험 점수, 한국인 맞춤형 예측 모형.

¹30019 세종특별자치시 세종로 2511, 고려대학교 일반대학원 응용통계학과 석사과정.

E-mail: yjya0201@korea.ac.kr

²(교신 저자) 세종특별자치시 세종로 2511, 고려대학교 빅데이터사이언스학부 부교수.

E-mail: jinhnam@korea.ac.kr

층화 직접질문 조건부 무관질문모형*

이기성¹, 홍기학², 손창균³, 박근화⁴, 홍성준⁵

요약

본 논문에서는 매우 민감한 조사에서 모집단이 층으로 구성되어 있을 때 덜 민감한 속성에 대하여 응답자들에게 직접질문을 하여 “예”라고 응답한 사람들에게만 무관질문모형을 사용하여 민감한 속성을 층화 추정할 수 있는 층화 직접질문 조건부 무관질문모형을 제안하였다.

주요용어 : 덜 민감한 속성, 직접질문, 무관질문모형, 층화추출, 조건부 모형.

1. 서론

Warner(1965)는 응답자의 신분이나 비밀을 노출시키지 않고서 민감한 질문에 대한 정보를 이끌어 낼 수 있는 확률화응답모형을 처음으로 제시하였으며, Greenberg et al.(1969)은 민감한 질문과 배반되는 질문 대신에 민감한 질문과 전혀 무관한 질문을 사용하는 무관질문모형을 제안하였다. 그리고 Loynes(1976)는 민감한 질문과 강요된 응답 “예”로 구성된 확률장치를 이용하는 강요질문모형을 제안하였으며, Carr, Marascuilo.(1982)은 덜 민감한 속성과 강요질문으로 구성된 확률장치를 통해 “예”라고 응답한 사람들에게만 Loynes의 강요질문모형을 사용하도록 하는 조건부 확률화응답모형을 제안하였다. 또한 Lee, Hong.(2000)은 Carr, Marascuilo.의 모형에서 사용한 강요질문모형 대신에 무관질문모형을 적용한 조건부 무관질문모형을 제안하였으며, Ahn, Lee(2003)는 덜 민감한 속성에 대하여 직접질문을 하는 직접질문에 의한 양적속성 조건부 무관질문모형을 제안하였다.

본 논문에서는 매우 민감한 조사에서 모집단이 층으로 구성되어 있을 때 덜 민감한 속성에 대하여 응답자들에게 직접질문을 하여 “예”라고 응답한 사람들에게만 무관질문모형을 사용하여 민감한 속성을 층화 추정할 수 있는 층화 직접질문 조건부 무관질문모형을 제안하고자 한다. 그리고 층화 직접질문 조건부 무관질문모형이 층화 무관질문모형과 층화 Carr, Marascuilo.의 모형보다 효율적이 되는 조건을 제시하고자 한다.

2. 층화 직접질문 조건부 무관질문모형

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

¹55338 전북 완주군 삼례읍 삼례로 443 우석대학교 아동사회복지학부 교수. E-mail : gisung@woosuk.ac.kr

²58245 전남 나주시 건재로 185 동신대학교 컴퓨터학과 교수. E-mail : khkhong@dsh.ac.kr

³38066 경북 경주시 동대로 123 동국대학교 빅데이터·응용통계학과 교수. E-mail : ckson85@dongguk.ac.kr

⁴07511 서울특별시 강서구 금남화로 154 한국문화관광연구원 선임전문위원. E-mail : pkhkwen@kcti.re.kr

⁵13822 경기도 과천시 경마공원대로 107 마사회 말산업연구소 연구원. E-mail : hsj8129@naver.com

이 절에서는 모집단이 층으로 구성되어 있고 얻고자 하는 속성이 민감할 때, 덜 민감한 속성에 대하여 특별히 확률장치를 사용하지 않고 응답자들에게 직접질문을 하여 “예”라고 응답한 사람들만을 대상으로 확률장치를 이용하는 무관질문모형을 사용하도록 하여 민감한 속성에 대한 층화 추정을 할 수 있는 층화 직접질문 조건부 무관질문모형을 제안하고자 한다.

크기 N 인 모집단이 각 층의 크기가 N_h ($h = 1, 2, \dots, L$)인 상호 배반인 L 개의 층으로 구성되어 있다고 하고, 모집단의 각 층의 크기를 알고 있다고 가정하자. 이때, h 층에서 덜 민감한 속성 B_h 를 가지고 있는 응답자들은 직접질문에 완전히 진실된 응답을 한다고 가정한다.

h 층의 첫 번째 단계에서 n_h 명의 응답자들은 “당신은 덜 민감한 속성 B_h 를 가지고 있습니까?”라는 직접질문에 “예”와 “아니오”로 응답하게 된다. h 층의 첫 번째 단계에서 응답자가 완전히 진실된 응답을 하는 경우 “예”라고 응답할 확률을 λ_{h1} 이라 하면, 이는 덜 민감한 속성 B_h 의 모비율 π_{h1} 과 일치하게 된다. 이 때, “예”라고 응답한 사람들의 수를 n_{h1} 명이라 하면, $\hat{\lambda}_{h1} = n_{h1}/n_h$ 이 되므로 덜 민감한 속성 B_h 에 대한 모비율 π_{h1} 의 추정량은 $\hat{\pi}_{h1} = n_{h1}/n_h$ 이 된다.

두 번째 단계에서는 h 층의 첫 번째 단계에서 “예”라고 응답한 n_{h1} 명의 응답자에게만 다음과 같은 Greenberg et al.의 무관질문모형의 확률장치를 통해 선택된 질문에 대하여 응답하도록 한다.

	질문내용	선택확률
질문 1	당신은 민감한 속성 A 를 가지고 있습니까?	p_h
질문 2	당신은 무관한 속성 Y 를 가지고 있습니까?	$1-p_h$

이 때, 응답자들은 확률장치에 의해서 선택된 질문에 대하여 “예” 또는 “아니오”라고 응답하게 된다. 따라서 h 층의 응답자가 첫 번째 단계에서 “예”라고 응답했다는 조건하에서 두 번째 단계에서 “예”라고 응답할 확률을 λ_{h2} 라 하면, 그 확률은 다음과 같다.

$$\lambda_{h2} = p_h \frac{\pi_{h2}}{\pi_{h1}} + (1-p_h)\pi_{hy}. \quad (2.1)$$

여기서, π_{h2} 는 h 층의 민감한 속성에 대한 모비율이며, π_{hy} 는 h 층의 무관한 속성에 대한 모비율이며 알고 있다고 가정한다.

h 층의 n_{h1} 명의 응답자들 중에서 “예”라고 응답한 사람의 수를 n_{h2} 라 하면 $\hat{\lambda}_{h2} = n_{h2}/n_{h1}$ 가 되므로 h 층의 민감한 속성에 대한 모비율 π_{h2} 의 추정량 $\hat{\pi}_{h2}$ 는 다음과 같다.

$$\hat{\pi}_{h2} = \frac{1}{p_h} \hat{\pi}_{h1} [\hat{\lambda}_{h2} - (1-p_h)\pi_{hy}] = \frac{1}{n_h p_h} [n_{h2} - (1-p_h)\pi_{hy} n_{h1}]. \quad (2.2)$$

따라서 민감한 속성 A 에 대한 모비율 π_2 의 추정량 $\hat{\pi}_2$ 는 다음과 같다.

$$\hat{\pi}_2 = \sum_{h=1}^L W_h \hat{\pi}_{h2} = \sum_{h=1}^L W_h \frac{1}{n_h p_h} [n_{h2} - (1-p_h)\pi_{hy} n_{h1}], \quad W_h = \frac{N_h}{N}. \quad (2.3)$$

<정리 1> 추정량 $\hat{\pi}_2$ 는 모비율 π_2 의 비편향추정량이다.

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

$$V(\hat{\pi}_2) = \sum_{h=1}^L W_h^2 \frac{1}{n_h p_h^2} [\pi_{h1}(1-p_h)\pi_{hy} \{1-(1-p_h)\pi_{hy}\} - p_h \pi_{h2} \{2(1-p_h)\pi_{hy} + p_h \pi_{h2} - 1\}] . \quad (2.4)$$

3. 효율성 비교

이 절에서는 제안한 층화 직접질문 조건부 무관질문모형이 Ahn, Lee(2000)이 제안한 층화 무관질문모형과의 Carr, Marascuilo의 모형보다 효율적이 되는 조건을 제시하고자 한다.

층화 무관질문모형에서 민감한 속성에 대한 모비율 π 의 추정량 $\hat{\pi}_g$ 의 분산은 다음과 같다.

$$V(\hat{\pi}_g) = \sum_{h=1}^L \frac{W_h^2}{n_h} \left[\pi_{h2}(1-\pi_{h2}) + \frac{(1-p_h)[p_h \pi_{h2}(1-2\pi_{hy}) + \pi_{hy} \{1-(1-p_h)\pi_{hy}\}]}{p_h^2} \right] . \quad (3.1)$$

제안한 층화 직접질문 조건부 무관질문모형의 분산 식(2.4)와 층화 무관질문모형의 분산 식(3.1)을 이용하여 $V(\hat{\pi}_g) - V(\hat{\pi}_2)$ 를 구해보도록 하자.

$$V(\hat{\pi}_g) - V(\hat{\pi}_2) = \sum_{h=1}^L \frac{W_h^2}{n_h p_h^2} [(1-\pi_{h1})(1-p_h)\pi_{hy} \{1-(1-p_h)\pi_{hy}\}] . \quad (3.2)$$

식(3.2)로부터 $p_h \neq 1, \pi_{h1} \neq 1$ 이면 $V(\hat{\pi}_g) - V(\hat{\pi}_2) > 0$ 이므로 이 조건은 일반적으로 성립되므로 $V(\hat{\pi}_2) < V(\hat{\pi}_g)$ 가 됨을 알 수 있다. 다음으로, Carr, Marascuilo의 모형에 층화추출법을 적용하여 얻은 민감한 속성 A 에 대한 모비율 π_2 의 추정량 $\hat{\pi}_c$ 의 분산은 다음과 같다.

$$V(\hat{\pi}_c) = \sum_{h=1}^L \frac{W_h^2}{n_h p_h} [\{p_h \pi_{h1} + (1-p_h)\}(1-p_h) - \pi_{h2}(1-2p_h + p_h \pi_{h2})] . \quad (3.3)$$

제안한 층화 직접질문 조건부 무관질문모형의 분산 식(2.4)와 층화 Carr, Marascuilo의 모형의 분산 식(3.3)으로부터 $V(\hat{\pi}_c) - V(\hat{\pi}_2)$ 을 구해보면 다음과 같다.

$$V(\hat{\pi}_c) - V(\hat{\pi}_2) = \sum_{h=1}^L \frac{W_h^2}{n_h p_h^2} (1-p_h) [\pi_{h1} \{ p_h^2 - \pi_{hy} (1 - (1-p_h)\pi_{hy}) \} + p_h (1-p_h) - 2p_h \pi_{h2} (1-\pi_{hy})] . \quad (3.4)$$

식(3.4)로부터 $V(\hat{\pi}_2) < V(\hat{\pi}_c)$ 를 만족하는 조건을 구해보면 식(3.5)을 얻을 수 있다.

$$\pi_{h2} < \frac{\pi_{h1} [p_h^2 - \pi_{hy} \{ 1 - (1-p_h)\pi_{hy} \}] + p_h (1-p_h)}{2p_h (1-\pi_{hy})} , \quad \pi_{hy} \neq 1 . \quad (3.5)$$

따라서 제안한 층화 직접질문 조건부 무관질문모형은 층화 Carr, Marascuilo의 모형에 비해 식(3.5)와 같은 조건에서 효율적임을 알 수 있다.

이 때, 식(3.5)에 의한 해석학적인 비교가 어려우므로 수치적인 비교를 통하여 효율성을 비교해

보고자 한다. 모집단이 2개의 층으로 되어 있고, $W_1 = 0.7$, $W_2 = 0.3$ 이라 가정하자. 그리고 $n = 1,000$ ($n_1 = 700$, $n_2 = 300$) 이고 h ($h = 1, 2$) 에서 $\pi_{h1} = 0.5$ 일 때, π_{h2} 와 p_h 및 π_{hy} 를 변화시켜 가면서 $\hat{\pi}_c$ 에 대한 $\hat{\pi}_2$ 의 상대효율 $RE = V(\hat{\pi}_c)/V(\hat{\pi}_2)$ 를 계산한 결과 다음 Table 1을 얻었다.

Table 1에서 1보다 큰 값은 제안한 층화 직접질문 조건부 무관질문모형이 층화 Carr, Marascuilo. 의 모형보다 더 효율적임을 나타낸다. Table 1에서 제안한 층화 직접질문 조건부 무관질문모형은 π_{h2} 값이 작고 π_{hy} 값이 작을수록 층화 Carr, Marascuilo. 의 모형보다 효율이 좋게 나타나고 있었다.

Table 1. Efficiency comparison

π_{h2}	π_{hy}				
	p_h	0.1	0.2	0.3	0.4
0.1	0.1	1.5779	0.9662	0.7510	0.6564
	0.3	2.8825	2.0469	1.6651	1.4641
	0.5	2.7924	2.2424	1.9220	1.7209
	0.7	2.1153	1.8798	1.7078	1.5786
	0.9	1.3333	1.2871	1.2454	1.2076
0.2	0.1	1.2133	0.8016	0.6436	0.5739
	0.3	1.8733	1.5120	1.3219	1.2208
	0.5	1.7108	1.5434	1.4343	1.3653
	0.7	1.3662	1.3118	1.2699	1.2385
	0.9	1.0815	1.0725	1.0644	1.0570

4. 결론

본 논문에서는 매우 민감한 조사에서 모집단이 층으로 구성되어 있을 때 덜 민감한 속성에 대하여 응답자들에게 직접질문을 하여 “예”라고 응답한 사람들에게만 무관질문모형을 사용하여 민감한 속성을 층화 추정할 수 있는 층화 직접질문 조건부 무관질문모형을 제안하였다. 그리고 층화 직접질문 조건부 무관질문모형이 층화 무관질문모형과 층화 Carr, Marascuilo. 의 모형보다 효율적이 되는 조건을 제시하였으며, 제안한 층화 직접질문 조건부 무관질문모형이 π_{h2} 값이 작고 π_{hy} 값이 작을수록 층화 Carr, Marascuilo. 의 모형보다 효율이 좋게 나타남을 수치적으로 보였다.

Reference

- Ahn, S. C., Lee, G. S. (2003). A conditional unrelated question model with quantitative attribute by direct question, *Journal of the Korean Data Analysis Society*, 5(3B), 581-589. (in Korean).
- Carr, J. W., Marascuilo, L. A. (1982). Optimal randomized response models and methods for hypothesis testing, *Journal of Educational Statistics*, 7, 295-310.
- Greenberg, B. G., Abul-Ela, A. A., Simmons, W. R., Horvitz, D. G. (1969). The unrelated question randomized response model : Theoretical framework, *Journal of the American Statistical Association*, 64, 520-539.
- Lee, G. S., Hong, K. H. (2000). A conditional unrelated question randomized response model, *Korean Journal of Computational and Applied Mathematics*, 8(1), 253-260.
- Loynes, R. M. (1976). Asymptotically optimal randomized response procedures, *Journal of the American Statistical Association*, 71, 924-928.
- Warner, S. L. (1965). Randomized response ; A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, 60, 63-69.

Comparison of algorithms for estimating regression splines using lasso penalty^{*}

Eun-Ji Lee¹, Jong-Beom Park², Dong-Young Lee³, Jae-Hwan Jhong⁴

Abstract

The Lasso regression model is a penalization technique that facilitates variable selection from high-dimensional data. When combined with a truncated power spline, Lasso regression addresses the knot selection problem from a nonparametric regression perspective. We compare the performance of three methods—coordinate descent, quadratic programming, and the alternating direction method of multipliers—in fitting Lasso regression spline models. Through simulations and analyses of two real datasets, we conduct numerical studies to evaluate the performance of these methods. By comparing their performance under various conditions, this study ultimately offers recommendations for the most suitable algorithm for different scenarios.

Keywords : Alternating direction method of multipliers; Coordinate descent algorithm; Knot selection; Lasso penalty.

^{*}This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00342014, RS-2024-00440787, RS-2024-00406939).

¹MS-Ph.D. integrated course student, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea. E-mail: ejkrsu97@gmail.com

²Bachelor's student, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea, whd3250@naver.com

³Bachelor's student, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea, dtsy5891@gmail.com

⁴(Corresponding Author) Associate Professor, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea. E-mail: jjh25@chungbuk.ac.kr

Dynamic Time Warping을 활용한 포트폴리오 최적화 방법

이석준¹, 정재홍²

요약

본 연구는 시계열 자료 간의 유사성을 비교하는 알고리즘인 Dynamic Time warping (DTW) 알고리즘을 사용하여 금융 자산 간의 공분산 구조를 새롭게 정의한 포트폴리오 최적화 모델을 제안한다. DTW는 한 시계열 자료의 첫 시점에서부터 마지막 시점까지 차례대로 다른 시계열 자료 내의 데이터 중 가장 가까운 시점을 찾고, 그 거리를 누적하여 유사도를 측정하는 방법이다. 이 과정을 통해 얻어진 DTW 값을 변환하거나 공분산 함수에 적용하여 새롭게 공분산 구조를 정의한 후 이를 포트폴리오 최적화 문제에 적용하기로 한다. 본 연구에서 제안하는 Scaled Dynamic Time Warping (SDTW)와 Covariance-Dynamic Time Warping (CDTW) 두 가지 모형과 공분산에 기반을 둔 전통적인 포트폴리오 모형들을 비교하는 실증분석을 진행하였다. 실증분석에는 한국 주식 시장의 자산 30종목을 활용하였다. 자산은 10개의 섹터에서 각각 시가총액 상위 3가지 종목으로 선택하였다. 포트폴리오 구성 전략으로는 Maximum Diversification Portfolio (MDP), Equally Risk Contributed Portfolio (ERCP), Hierarchical Risk Parity (HRP)를 사용하였다. 포트폴리오 성과는 Sharpe Ratio와 누적 수익률을 비롯한 7가지 평가지표를 사용하여 비교하였다. 연구 결과, Maximum Drawdown를 제외한 평가지표들에서 전통적인 포트폴리오 모형들보다 제안한 모형들이 향상된 결과를 확인할 수 있었다. 세부적으로는 포트폴리오 구성 전략에 따라 높은 Sharpe ratio를 보이는 모형이 달랐다. MDP에서는 SDTW가 높은 Sharpe ratio를 보인 반면, ERCP와 HRP에서는 CDTW가 높은 모습을 보였다.

주요 용어 : 포트폴리오 최적화, Dynamic Time Warping, 공분산 함수, SDTW, CDTW.

¹04763 서울시 성동구 왕십리로 222, 한양대학교 응용통계학과 석사과정. E-mail: stonejune@hanyang.ac.kr

²(교신저자)04763, 서울시 성동구 왕십리로 222, 한양대학교 수학과 부교수. E-mail: jaehongjeong@hanyang.ac.kr

소셜 빅데이터에 나타난 고독에 대한 인식 분석*

전란영¹

요약

본 연구의 목적은 최근 현대인의 삶과 심리적 안녕의 중요한 주제로 주목받는 고독에 대한 사회적 인식을 알아보는 것이다. 이를 위해 소셜 빅데이터에 언급된 고독에 대한 게시 글을 분석하였다. 주요 소셜 네트워크 서비스인 인스타그램과 블로그에 2023년 11월에서 2024년 10월까지 게시된 글을 수집하여 분석하였다. 자료수집 및 분석에는 썬트렌드(Sometrend) 서비스를 사용하였다. 주요 분석 방법은 빈도 분석, 연관어 분석, 감성 분석이었다. 분석 결과, 첫째, 고독에 관한 기간 내 총 언급량은 73,595건(인스타그램 19,819건, 블로그 53,776건)이었다. 둘째, 연관어 분석 결과, '삶', '마음', '시간', '생각', '책', '인생', '인간' 등이 고독과 관련된 주요 키워드로 나타났다. 셋째, 감성 분석 결과, 고독과 관련된 주요 감성어는 '외로움', '사랑', '고통', '즐거움', '행복', '슬픔' 등이었다. 게시 글에 나타난 고독에 대한 긍정과 부정 평가 비율은 각각 50.7%와 46.7%였다. 이러한 결과는 고독이 주로 삶이나 시간과 함께 인식되는 경향이 있음을 나타낸다. 또한 고독이 긍정적으로 인식되기도 하지만 부정적으로 인식되기도 하여 고독에 대한 양가적 감정이 존재할 수 있음을 시사한다. 본 연구는 고독에 대한 사회적 인식을 조명하였기에 향후 고독을 더 깊이 탐구하고 관련 논의를 발전시키는 데 필요한 기초 자료로 활용될 수 있다.

주요 용어: 고독, 외로움, 사회적 인식, 소셜 빅데이터 분석, 감성 분석.

*이 논문은 2023년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2023S1A5B5A16078886).

¹03186 서울특별시 종로구 새문안로 92, 사)한국인문사회총연합회 학술연구교수. E-mail: ryjeon5555@naver.com

로버스트 왜도를 기반으로 한 이상치 탐지에 관한 연구: 영향함수 산출 및 산업자료로의 응용

정윤채¹, 정수빈², 박민수³

요 약

본 연구는 비대칭도를 측정하는 주요 통계량인 사분위수 왜도(quantile skewness), 팔분위수 왜도(octile skewness), medcouple의 특징을 비교하기 위해 영향함수를 활용한다. 사분위수 왜도는 분포의 양 극단값에 강건한 특성을 보이며, 팔분위수 왜도는 전체 분포를 반영하는 특성이 두드러진다. Medcouple은 두 통계량의 장점을 결합한 강건한 통계량으로 평가받고 있다. 본 연구에서는 각 통계량의 영향함수를 산출하고, 이를 대칭 분포 또는 비대칭 분포에서 시각화하여 분석한다. 사분위수 왜도와 팔분위수 왜도는 계단 함수 형태의 영향함수를 보이나 medcouple은 중위수를 제외한 부분에서 연속적인 함수 형태를 보인다. 이러한 계단 함수 형태는 작은 데이터 변화에 민감하게 반응하여 일관된 분석을 어렵게 만든다. 대칭인 분포에서 팔분위수 왜도는 데이터 분포 전반에서 높은 영향력을 가지며, 사분위수 왜도는 극단값에서 강건했으나 사분위수 범위 내에서는 민감도가 증가한다. Medcouple은 전반적으로 두 통계량 사이의 민감도를 가지나, 중위수 부근에서 민감도가 크게 증가하는 함수로 표현된다. 치우친 분포에서는 치우친 방향에 위치한 데이터에 대해 팔분위수 왜도와 사분위수 왜도에 대한 민감도는 증가하지만, medcouple은 대칭분포와 유사한 패턴을 유지한다. 이러한 분석을 통해, medcouple이 다양한 형태를 가정한 시뮬레이션 자료에서 이상치 탐지에 가장 안정적인 성능을 보였으나, 영향함수에서 보이는 단점들을 보완할 보정된 통계량 개발을 필요로 한다.

주요 용어 : 비대칭 분포, 사분위수 왜도, 팔분위수 왜도, medcouple, 영향함수.

¹34134 대한민국 대전광역시 유성구 대학로 99, 충남대학교 정보통계학과 학부재학생.

E-mail: bella55560@gmail.com

²34134 대한민국 대전광역시 유성구 대학로 99, 충남대학교 통계데이터사이언스학과 석사과정.

E-mail: tnqls2053@gmail.com

³(교신저자) 34134 대한민국 대전광역시 유성구 대학로 99, 충남대학교 통계데이터사이언스학과 조교수.

E-mail: minsu.park51@gmail.com

베이지안 기법을 적용한 당뇨병 및 당뇨합병증 예측 딥러닝 모델 설계

조영찬¹, 박성희², 이경준³

요약

당뇨병은 고혈당을 특징으로 하는 대사성 질환이며, 전 세계적으로 그 유병률이 계속해서 증가하고 있다. 당뇨병은 뇌, 심장, 신장을 비롯한 여러 신체 기관에 다양한 합병증을 유발할 수 있어, 조기 발견을 예측하는 것이 매우 중요하다. 본 연구에서는 한국인 유전체역학조사사업 (KoGES) 및 안성·안산 지역사회의 기본 코호트에서 수집된 10,030명의 데이터를 활용하여 2년간격으로 10년간 추적한 자료를 바탕으로 분석을 진행했다. 10년 이내에 당뇨병이 발생할 확률을 예측하기 위해 RNN, LSTM, GRU, 1차원 합성곱 신경망(CNN) 모델을 결합한 앙상블 모델을 구축했다. 각 모델의 노드 수와 배치 크기는 베이지안 최적화 기법을 통해 최적화하였고, 그리드 서치로 도출된 모델 성능과 비교했다. 이후, 당뇨병의 주요 합병증인 심혈관질환과 신장질환의 발병 가능성을 예측함으로써, 당뇨병의 발병 메커니즘을 보다 깊이 있게 이해하고, 이를 바탕으로 맞춤형 환자 관리 방안을 제시하여 정밀 의료의 발전 및 의료 기술 향상을 도모하고자 한다.

주요 용어 : 당뇨병, 딥러닝, 베이지안 최적화, 코호트.

¹39177 경상북도 구미시 대학로 61, 국립금오공과대학교 일반대학원 수학과, 석사과정.

E-mail : dudcks6436@naver.com

²39177 경상북도 구미시 대학로 61, 국립금오공과대학교 일반대학원 수학과, 석사과정.

³(교신저자) 39177 경상북도 구미시 대학로 61, 국립금오공과대학교 수리빅데이터학과, 조교수.

Llama-3 모델을 활용한 법정감염병 정보 챗봇 구현*

주진범¹, 형도훈², 문형빈², 김창훈⁴, 노맹석⁵

요약

최근 인공지능의 발전의 가속화에 따라 다양한 분야에서 인공지능을 접목한 기술 및 서비스 개발이 활발하게 진행되고 있는 상황이다. 특히, 현재 자연어 처리 분야에서 연구가 활발하게 진행되었으며 이에 따라 다양한 대규모 언어모델(Large Language Model)들이 공개되었다. 이는 대규모 언어모델 활용에 대한 접근성을 향상시켰으며 그 결과 다양한 분야에서 자신들의 데이터를 공개된 대규모 언어모델에 미세 조정하여 독자적인 서비스를 구현 중이다. 대표적으로 감정분석, 챗봇, 정보요약 등의 형태로 구현되고 있다. 챗봇 서비스는 이용자에게 제공될 정보 범위 내의 질문을 실시간으로 받아서 응답하는 서비스를 제공하며 이는 이용자에게 신속한 정보 제공, 실시간 상호작용 제공자에게는 단순 질의응답의 자동화를 통하여 등의 편리함을 제공한다. 이에 따라 해당 연구에서는 최근 공개된 Llama-3-8B의 모델을 활용하여 몇 개의 법정 감염병에 대해 사전에 제작된 질문과 대답을 미세 조정하여 챗봇 구현을 시도했다. 구체적으로 법정감염병 중 1급과 2급에 해당되는 질병 10개를 추출하여 총 1732개의 질의응답을 제작한 후 Google의 colab 서비스의 A100 환경에서 Meta-Llama-3-8B-Instruct model을 활용하여 Batch Size = 4, Optimizer = AdamW, Epoch = 10으로 hyper parameter 설정을 한 후 미세 조정을 진행하였다. 그 결과 Epoch = 10의 환경에서 최종 Loss-rate는 0.2096이다.

주요 용어 : Llama-3, 챗봇, 감염병, LLM, FineTuning.

*이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (RS-2023-00242528).

¹48513 부산광역시 남구 용소로 45(대연동, 국립부경대학교대연캠퍼스) 부경대학교 일반대학원 데이터공학과, 석사 석사과정. E-mail: apollon3110@pukyong.ac.kr

²48513 부산광역시 남구 용소로 45(대연동, 국립부경대학교대연캠퍼스) 부경대학교 일반대학원 데이터공학과, 석사 석사과정. E-mail: hdhd4422@naver.com

³(교신저자) 48513 부산광역시 남구 용소로 45(대연동, 국립부경대학교대연캠퍼스) 부경대학교 정보융합대학 빅데이터융합과 조교수. E-mail: hbmoon@pknu.ac.kr

⁴50612 경상남도 양산시, 물금읍, 부산대하로 49, 부산대학교 의과대학 예방의학 및 직업환경의학교실 교수. E-mail: kchprev@gmail.com

⁵48513 부산광역시 남구 용소로 45(대연동, 국립부경대학교대연캠퍼스) 부경대학교 정보융합대학 빅데이터융합과 교수. E-mail: msnoh@pknu.ac.kr

Robust Slope Estimation in Progress Data from AI-Driven ALEKS: Application of Theil-Sen Estimator*

왕흔이¹, 한현석²

요 약

COVID-19 팬데믹은 온라인 교육과 인공지능(AI) 기반 학습 분석 플랫폼의 급속한 발전을 촉진시켜 대규모 학습 데이터 수집을 가능하게 했으며, 이를 통해 전통적인 데이터 수집 방법의 한계를 극복할 수 있었다. 본 논문은 AI 기반 플랫폼인 ALEKS의 데이터를 활용하여, 일반 최소 제곱법(OLS)과 비모수적 테일-센(Theil-Sen) 방법이 학생의 학습 진전도를 평가하는 데 얼마나 효과적인지를 비교하고, 다양한 조건에서 데이터 분석의 신뢰성과 정확성을 보장할 수 있는 방법을 논의하였다. 연구 결과, OLS와 테일-센 방법은 전반적으로 유사한 기울기 추정치를 제공했지만, 테일-센 방법은 극단적인 데이터에 대해 더 높은 신뢰성을 보인다. 이는 특히 저성취 학생의 데이터 분석에서 두드러지며, 비모수적 방법의 필요성을 시사한다. 본 연구는 학습 진전도 점검 시 보다 정확하고 효과적인 결정을 위해 비모수적 방법을 포함한 다양한 분석 접근법의 필요성을 강조하고 있다. 또한, AI 기반 데이터 분석의 정밀성과 결과의 안정성을 높이기 위한 추가 연구의 필요성을 제기한다.

주요 용어 : Theil-Sen, OLS, Estimation, ALEKS.

*본 논문은 전북대학교의 지원을 받아서 연구된 것임.

¹대한민국 전주시 백제대로 567, 전북대학교 심리학과 석사 과정.

E-mail: mollymolly12345678@gmail.com

²(교신저자)대한민국 전주시 덕진구 백제대로 567, 전북대학교 심리학과 교수. E-mail: hysh@jbnu.ac.kr

효과적인 국문 판결문 요약을 위한 거대언어모형 비교연구*

홍인석¹, 최호식²

요약

소송을 통해 분쟁을 해결하려는 사람이 늘면서 개인이 판결문에서 재판부의 판단을 명확히 이해하고 효과적으로 방어권을 행사할 필요성이 커지고 있다. 본 연구는 최근 주목받는 자연어 처리 기술을 활용해 국문 판결문 중 재판부 판단 부분을 효과적으로 요약하는 모델이 무엇인지 비교하고 그 결과를 분석하는 것을 목적으로 한다. 최신 자연어 처리(NLP) 기술을 기반으로 자동 요약 모델을 선택해 판결문 중 판단 또는 재판부 판단 요약을 효과적으로 할 수 있는 것은 무엇인지 살펴본다. 연구에서는 범용적인 언어모델인 KoGPT Mini, Flan-T5 Small/Medium, KoAlpaca 7B 등을 활용하여 민사 판결문 중 재판부 판단의 원문에서 중요한 문장을 요약할 수 있는 추출요약(extractive summarization) 방법을 고찰해본다. 요약된 텍스트는 평가 지표인 ROUGE (Recall-Oriented Understudy for Gisting Evaluation)를 고려하여 정확성을 정량적으로 평가한다. 아울러 전문법조인 평가를 통해, 재판부 판결문의 법적 의미가 잘 담겨있는지를 정성적으로 평가한다.

주요 용어 : 판결문, 판결문 요약, 법률 LLM, 법률 문서

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

¹02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 도시빅데이터융합학과 석사과정.

E-mail: mystic@chosunbiz

²02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 일반대학원 도시빅데이터융합학과, 교수.

E-mail: choi.hosik@gmail.com

중국 고등교육 입학기회의 공정성 연구

귀쯔페이¹, 양정호²

요약

이 논문의 서론 부분은 주로 고등 교육 입학 기회의 공정성에 대한 관련 문헌을 분류하고 분석하며, 논문의 본문은 주로 세 부분으로 구성된다. 첫 번째 부분은 교육의 공정성에 대한 관련 개념을 정의하고 관련 이론적 근거를 정리 및 연구하며 기초 교육의 공정성과 고등 교육의 공정성의 차이점에 대해서도 논의하고 기초 교육의 공정성은 '평등'을 추구하는 것이고 고등 교육의 공정성은 '공정'과 '효율'을 모두 고려하여 후속 연구를 위한 개념 기반을 마련한다. 두 번째 부분은 주로 고등 교육 입학 기회의 형평성에 대한 추가 분석을 수행하고 중국의 고등 교육 입학 기회의 공정성에 영향을 미치는 세 가지 요인, 즉 시험 제도의 공정성, 대학 모집 계획의 공정성 및 대학 입학 제도의 공정성을 도출한다. 세 번째 부분은 입시제도의 공정성, 대학 모집 계획의 공정성, 대입제도의 공정성에 대해 논의한다. 중국의 고등 교육 입학 기회에 영향을 미치는 이 세 가지 요소는 현황 분석과 외국 경험의 참고 자료를 기반으로 해당 정책 제안을 제시하여 중국 시험 입학 시스템의 공정성, 체계성 및 과학성을 보장한다. 본 논문은 고등 교육 입학 기회의 공정성을 국가 정책의 관점에서 살펴보고, 현대 교육의 공정성에 관한 이론을 적용한다. 또한 문헌 분석, 비교 분석, 실증 연구 등의 연구 방법을 통해 시험의 공정성, 입학 계획의 공정성, 입학 절차의 공정성이라는 3차원 제도 수준과 그 발전 역사를 분석하며, 이와 관련된 문제와 영향 요인에 대해 깊이 연구한다. 선진국의 경험을 참고하고, 차세대 정보 기술 발전의 최신 연구 성과를 결합하여 상기 문제를 해결하기 위한 새로운 아이디어, 방법, 대책을 모색하며, 관련 연구와 실천 지도를 위한 새로운 참고 자료를 제공한다.

주요 용어 : 고등교육, 입학기회 공정, 시험제도의 공정성, 모집계획의 공정성, 대입제도의 공정성.

1. 서론

중국의 고등 교육이 대중화 단계에 진입함에 따라 고등 교육에 대한 사람들의 요구가 점점 더 시급해지고 있으며, 고등 교육의 질과 양에 대한 기대가 높아지고 있다. 교육 공정성의 출발점으로 서 고등 교육 입학 기회는 시민이 고등 교육을 동등하게 받을 권리가 있는지를 측정하기 위한 전체 조건이자 기준이다. 고등 교육 입학 기회의 공정성에 대한 사회적 관심이 지속되는 이유는 중국 사회의 주요 모순의 변화, 시민의 평등한 교육 권리 등과 불가분의 관계에 있기 때문이다. 고등

¹03063 서울특별시 종로구 성균관로 25-2 성균관대학교 호암관 50519호, 성균관대학교 교육사회학과 석사과정. E-mail: guozifei1127@163.com

²(교신저자)03063 서울특별시 종로구 성균관로 25-2 성균관대학교 호암관 50519호, 성균관대학교 교육사회학과 교수.E-mail: jyang@skku.edu

교육은 개인의 능력과 수준 향상에 핵심적인 역할을 하며, 국가와 사회의 진보에 강력한 추진력을 제공한다. 최근 몇 년 동안 중국의 고등 교육 수준도 지속적으로 향상되어 대중화되는 추세를 보이고 있으며, 사람들은 더 높은 수준의 교육을 받기를 원하지만, 중국의 고등 교육 자원의 전반적인 배치에는 불균형이 존재한다. 특히 다른 성(省)과 도시, 농촌 간의 고등 교육 기회에는 여전히 큰 차이가 있으며, 고등 교육 입학 기회의 공정성은 사회 각계각층의 광범위한 관심의 초점이 되고 있다.

2. 본론

2.1. 고등교육 입학기회의 공정성 문제 분석

고등 교육 입학 기회의 형평성에 영향을 미치는 요인

(1) 질 높은 입학 기회 형평성 부각

2015년부터는 40%에서 2023년 60.2%로 점차 확대되었으며, 고등교육 시스템은 매우 광범위하게 발전하고 있다. 또한, 양적으로 입학 기회의 형평성을 보장하는 동시에, 중국 정부는 학생의 능력 평가와 선발에 점점 더 주의를 기울이기 시작했다. 그 결과, 고품질 고등교육 입학 기회의 형평성이 부각되기 시작했으며, 이는 대학 입학시험 내용의 개혁에도 반영되었다.

이 논문에서는 중국의 고등교육 현황을 분석하기 위해, '중국 고등교육 입학 기회의 형평성 연구'에서 리원청이 정의한 상대 풍부도(RA; Relative Abundance) 개념을 참고하다. 여기서 RA는 인구 100만 명당 일반 대학의 수를 기준으로 각 성(省)의 대학 분포의 상대적 풍부함을 나타낸다. 2024년 중국 교육부의 최신 통계에 따라 각 성(省)의 RA 값을 정렬하고, 그 결과를 표로 제시했습니다. 2024년 기준 전국 각 성의 인구 100만 명당 대학 수(RA)에는 여전히 큰 격차가 있음을 알 수 있다. 분석 결과, 전국 성급 행정구역(홍콩, 마카오, 대만 제외)에서 RA 값이 가장 높은 지역은 여전히 베이징으로 4.21이었으며, 텐진과 닝샤가 각각 4.11과 3.02로 그 뒤를 이었다. 반면 저장, 산둥, 광둥은 상대적으로 풍부도가 가장 낮은 3개성으로 각각 1.64, 1.59, 1.3을 기록했다. 일부 도시나 지역은 고등교육기관이 많아 해당 지역의 학생 수가 많은 편이다. 이는 중국 대학의 전국적 분포가 여전히 불균형하다는 것을 보여준다.



2.2. 입시 제도의 공정

제1절 시험제도의 현황 분석

(1). 시험 횟수

시민이 고등 교육 시스템에 진입하는 가장 중요한 방법으로서 수능의 중요성은 자명하다. 그러나 중국에서는 수험생이 시험을 치를 때 많은 우연적 요소와 돌발 상황이 발생할 수 있다. 공정한 개인이라 할지라도 단 한 번의 시험으로 합격할 학교가 결정되는 것은 적절하지 않다.

(2). 시험 형식

중국은 2003년부터 대학의 자율 입학과 수시모집 제도를 공식적으로 도입했다. 그러나 이에 상응하는 감독 체계가 부재하여, 고등 교육 분야에서 일부 불공평한 현상이 발생했다.

제2절 해외 시험 제도의 경험을 거울삼다

(1) 미국의 시험제도 (2) 한국의 시험제도

제3절 입시 제도의 공정 실현을 위한 건의

(1) 시험 기회를 더 많이 제공한다 (2) 대입 시험 진입 조건의 제한을 완화하다. (3) 기초 교육을 대대적으로 발전시키다 (4) 선진적인 지능 기술을 보급하다.

2.3. 대학 입학 계획의 공정성

제1절 대학 모집 계획의 현황 분석

(1) 할당제도와 그로 인한 불공정. 현재 중국의 대학 입학 계획은 주로 두 가지 방식으로 운영된다. 첫 번째는 현재 대학 입시 모집에 적용되는 방식으로, 이를 '성 할당제'라고 한다. 이 방식에서는 대학이 미리 각 성, 자치구, 직할시에 모집 정원을 할당한 후, 해당 지역 내 학생들의 성적에 따라 점수를 배열하고, 할당된 정원수에 따라 합격선을 설정하여 입학을 결정한다. 두 번째 방식은 현재 대학원생 모집에서 채택하는 정원 할당 방식이다.

(2) 시험지의 차이와 그로 인한 불공정

중국의 학생들은 대부분 입시를 통해 대학에 진학해야 하지만, 시험지의 내용과 합격선은 성마다 다르다. 본 논문에서는 2024년 각 성에서 채택한 수능 시험지를 예로 들어 분석을 수행했다. 2024년 각 성의 시험지는 주로 전국공통시험과 자율출제시험 두 가지 유형으로 나뉘며, 전국공통 시험은 1권, 2권, 전국 갑권으로 세분화된다. 저자는 분석을 위해 동일한 시험지를 사용하는 일부 성의 학부 커트라인을 비교하여 분석하기로 결정했다. 분석 결과는 아래 그림과 같다.

(3) 대입 계획의 속지화와 그로 인한 불공정

중국의 대학은 주로 중앙정부가 지원하는 중앙부처 핵심대학, 성급 재정지원대학, 시 재정지원 지방대학의 세 가지 유형으로 구성되어 있다. 산하 대학은 주로 전국을 대상으로 모집하고, 성급 대학은 주로 지방에서 모집하며, 지방 대학은 주로 지역 학생을 모집한다. 현실적으로 부처와 지방 모두 다른 지역보다 훨씬 높은 지역 입학 계획을 가지고 있다. 이는 같은 지역 내 수험생 간의 경쟁은 상대적으로 공정하지만, 지역 간 형평성은 크게 떨어짐을 의미한다.

이 논문에서는 베이징 대학의 입학 계획을 예로 들어 중국 입학 계획의 현지화를 분석한다. 2024년 전국 대입 시험에 응시하는 수험생은 약 1340만 명으로 예상됩니다. 참가자 수가 가장 많은 성은 허난성(136만 명)이고, 그 뒤를 산둥성(100만 명)이 이었다. 닝샤, 톈진, 하이난, 베이징, 칭하이, 상하이, 티베트 등 인구가 10만 명을 초과하지 않는 7개 성이 있으며, 베이징대의 속주 모집 인원은 중국의 다른 어느 지역을 훨씬 능가하며, 2024년 베이징의 대학 입시 응시자 수는 68,000명에 불과하다. 반면, 산둥성, 허난성 등 인구가 많은 성과 개발 수준이 낮은 성의 응시자 수는 베이징 응시자 수의 10배에 불과하지만, 모집 인원은 베이징 모집 인원의 몇 분의 1에 불과하다. 모집 계획 수의 차이로 인해 다른 성, 특히 인구가 많은 성과 상대적으로 낙후된 지역의 응시자들은 베이징 대학과 같은 지역화 경향이 더 심각한 대학에 입학하는 것이 어려워지며, 양질의 고등교육을 받을 기회가 더욱 제한된다.

제2절 해외 대학 신입생 모집 계획의 경험을 거울로 삼다
미국의 긍정적 계획(affirmative action)

제3절 대학 입학 계획의 공평함을 실현하기 위한 건의

(1).부속 대학과 속지의 학생 모집 계획을 합리적으로 조절한다. (2) 점차 전국적인 시험 형식을 보급하다 (3) 대학 수시모집의 형평성 촉진 (4) 단계적 개혁 할당제 및 그 할당량 분배 원칙 (5) 원격 교육의 새로운 모델과 선진 정보 기술 수단을 대대적으로 보급하다

2024년 중국 전국 대학수학능력시험 지역별 시험지 종합표		2024년 전국 일부 지역 학부 최저 합격선		
시험지 종류	2024년 대학수학능력시험 성	지역	문과	이과
신수능 I권	광둥성, 푸젠성, 후베이성, 후난성, 장쑤성, 허베이성, 산둥성, 저장성, 장시성, 안후이성, 허난성	안후이(전국 I권)	512	514
		저장(전국 I권)	595	595
신수능 II 권	랴오닝성, 충칭, 하이난성, 산시성, 신장, 광시, 구이저우성, 헤이룽장성, 간쑤성, 지린성, 윈난성	호북(전국 I권)	530	525
		후난(전국 I권)	497	481
		광시(전국 II 권)	519	501
		흑룡강(전국 II 권)	486	480
전국 갑권	쓰촨성, 티베트, 네이멍구, 닝샤, 산시성, 칭하이성	하이난(전국 II 권)	568	568
		구이저우(전국 II 권)	505	482
		쓰촨(전국 갑권)	529	539
베이징권	베이징	칭하이(전국 갑권)	411	343
톈진권	톈진	내몽골(전국 갑권)	478	471
상하이권	상하이	북경(자주명제)	523	523
		상하이(자주명제)	503	503
		톈진(자주명제)	563	563

순번	성.	2019학년도 대학수학능력시험 문항의 수(인원)	특정대학의 계획 학생 모집 인원				실명
			이과	문과생	합계	비율(%)	
1	취안성	136	126	32	58	0.43	
2	산둥	100	44	22	66	0.66	
3	허베이	88	18	22	40	0.45	
4	쓰촨성	88	21	26	47	0.57	
5	윈둥	76	24	24	48	0.63	
6	후난성	72	31	31	62	0.86	
7	허난성	67	21	18	39	0.58	
8	광시성	65	22	20	42	0.65	
9	후베이	52	40	29	69	1.33	
10	허베이	50.6	21	23	44	0.87	
11	광둥성	48	20	28	48	1.00	
12	광시성	48	18	20	38	0.79	
13	윈난성	40.6	9	10	19	0.47	
14	저장	40.5	24	7	31	0.77	
15	산시성	35.4	21	30	51	1.44	
16	충칭	35	13	16	29	0.83	
17	산시성	35	25	20	51	1.46	
18	간쑤성	26	14	15	29	1.12	
19	후베이	24.7	25	20	45	1.82	
20	대륙성	23	10	10	20	0.87	
21	신장	23	12	17	29	1.26	
22	베이징	21.1	16	16	32	1.52	
23	허난성	20.3	22	22	44	2.17	
24	광둥성	19.2	14	24	38	2.88	
25	닝샤	8.2	8	9	17	2.07	
26	텐진	7.4	19	17	36	4.86	
27	허베이	7.3	6	8	14	1.92	
28	베이징	6.8	80	108	188	27.65	
29	칭하이성	5.9	7	8	15	2.54	
30	장쑤성	5.0	5	5	10	1.00	
31	티베트	3.9	4	4	8	2.05	

2.4. 대학 입학 제도의 공정성

중국 대학 입학 전형 규칙은 대학 입학 규정의 중요한 내용이며, 주로 등록 학생의 전공 배치, 입학 운영 및 근거와 같은 제도적 규범을 포함하여 교육부의 관련 규정에 따라 대학에서 공식화된다. 위의 규칙의 공정성은 사회적 관심의 초점이자 교육 관리부서의 관리 및 통제의 핵심이다. 이 장에서는 대학 입학 규칙의 현황과 공정성을 분석하고, 미국 영국의 관련 경험을 바탕으로 개선 대책과 제안을 제시한다.

대학 입학 규칙의 공정한 실현을 위한 제언

1. 학생 모집에 대한 감독 및 책임제를 완비한다.
2. 업무 효율성을 높이고 정보를 적시에 공개한다.
3. 새로운 대입 시험 지원 방식을 적극적으로 탐색한다.
4. 대학이 더 많은 학생 모집 자율권을 확보하도록 하고, 특기자 전형 규칙의 공정성을 보장한다.
5. 과학적이고 합리적인 종합 평가 시스템을 구축한다.

3. 결론

고등교육 입학 기회의 불평등은 중국의 사회 발전 과정에서 다양한 요인으로 인해 발생하는 복합적인 사회 문제이다. 중국에서 고등교육을 받는 것은 여전히 국민의 기본권으로 완전히 자리 잡지 못했으며, 교육비 부담은 주로 학생 본인과 가정에 의존하고 있다. 대입 전형 과정에서 어떤 측면에서는 공정하다고 여겨지는 현상이 다른 측면에서는 불공정하게 보일 수 있다. 이에 따라 이 글은 교육의 공정성을 정의하고, 기초교육의 공정성과 고등교육의 공정성의 차이를 논의하고자 한다. 기초교육(특히 의무교육)의 공정성은 '교육 자원의 평등'을 달성하기 위해 최선을 다해야 하며, 모든 사람이 현대 사회에 필요한 시민으로서의 기본 소양을 갖추도록 하는 데 중점을 두어야 한다. 반면, 고등교육의 공정성은 '기회의 평등'을 바탕으로 '공정'과 '효율성'을 균형 있게 고려해야 한다. 고등교육의 공정성을 평가하는 세 가지 기준을 요약하면 다음과 같다. 첫째, 모든 사람에게 입학 기회가 보장되어야 한다. 둘째, 모든 사람은 동일한 평가 기준에 따라 심사를 받아야 한다. 셋째, 모든 사람은 동일한 입학 기준을 적용받아야 한다. 전문적인 분석을 통해 현재 중국의 고등교육 입학 기회에 영향을 미치는 세 가지 주요 요인을 도출할 수 있다. 그것은 시험의 공정성, 계획의 공정성, 대입 제도의 공정성이다. 이러한 요인들에 대한 상세한 분석과 더불어, 국내외의 관

런 선진 사례를 참고하여 다음과 같은 제안을 제시한다. 시험 횟수를 늘리고 종합 평가를 적절히 시행하며, 대학 입시 접근 조건의 제한을 완화하고, 산하 대학의 입학 계획을 확대하며, 국가 시험 지 형식을 채택하고, 할당제를 점진적으로 개혁하며, 국가의 재정 지원을 확대하고, 현지화를 줄이며, 관련 법률을 제정하고, 입학 감독 및 책임 체계를 개선하며, 효율성을 높이고 정보 공개를 실현하며, 새로운 대학 입시 지원 방식을 적극적으로 탐색해야 한다. 또한, 신경교육 이론의 지도하에 새로운 원격 교육 모델, 빅 데이터 분석 기술, 인공지능 기술의 적용을 적극적으로 추진해야 한다. 각 국가의 특수한 여건을 충분히 고려하며, 국내외의 우수한 경험을 선택적으로 참고해야 한다. 이를 통해 중국 특색의 고등교육 발전 모델을 모색하고, 고등교육 입학 기회의 실질적이고 공평한 실현을 지속적으로 추진해야 한다.

References

- James S. Coleman.(1968). The Concept of Equality of Educational Opportunity, *Harvard Educational Review*, Vol.38.
- John Rawls.(1989). The Domain of the Political and Overlapping Consensus, *NewYork University Law Review*, Vol.64, No. 2, pp. 242.
- Liu Jingming.(2014). Capacity and Origin: Mechanism Analysis of the Allocation of Admission Opportunities to Higher Education, *Chinese Social Sciences*, No. 8.
- Tang Ying.(2003). "Inquiry into the Equal Opportunities of Higher Education in China's Minorities", *Education and Economy*, No. 2.
- China Education Online: 2024 National College Entrance Examination Applicants, <http://gaokao.eol.cn/gkbn>
- Ministry of Education: Education Law of the People's Republic of China, http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_619/200407/1316.html

Statistical Inference Methods for Two-Parameter Exponential Distribution in Generalized Progressive Hybrid Censoring Schemes

*Jihyeon Kim*¹, *Yeongeun Hwang*², *Kyeongjun Lee*³

Abstract

Hybrid censoring schemes are extensively utilized in life-testing experiments to reduce experimental time and costs. The Type-II progressive hybrid censoring scheme (PHCS) combines the features of both progressive Type-II and Type-I censoring. This study investigates the statistical inference for a two-parameter exponential distribution under the generalized PHCS. Maximum likelihood estimates (MLEs) for the model parameters are derived, along with their joint and marginal moment-generating functions. Using these exact moments, bias-reduced estimators are proposed, and their distributions are carefully examined. Confidence intervals for the model parameters are constructed based on exact and asymptotic distributions of both the MLEs and bias-reduced estimators. The effectiveness of point and interval estimation methods is compared and evaluated through exact calculations and Monte Carlo simulations. Recommendations are made based on these analyses, and an illustrative example is provided.

Keywords : Generalized Progressive Hybrid Censoring, Maximum Likelihood Estimation, Two-Parameter Exponential Distribution, Bias-Reduced Estimators, Monte Carlo Simulations.

¹Graduate student, Department of Mathematics, Kumoh National Institute of Technology, Gumi 39177, Gyeongbuk, Korea. E-mail: rlawlgus4417@naver.com

²Ph.D. student, Department of Mathematics, Kumoh National Institute of Technology, Gumi 39177, Gyeongbuk, Korea.

³(Corresponding author) Associate professor, Department of Mathematics and Big Data Science, Kumoh National Institute of Technology, Gumi 39177, Gyeongbuk, Korea.

Exponentiality Testing for Progressive Type-II Censored Data with a Local Linear Regression Estimator of Kullback-Leibler Information

*SangHeon Jeong*¹, *Kyeongjun Lee*²

Abstract

This article presents a new method for goodness-of-fit testing for Progressive Type-II censorship data. The method leverages a new Kullback-Leibler information estimator based on local linear regression. We explore the properties of the proposed test statistic in detail, and a newly developed test statistic is applied to evaluate the exponential distribution using Progressive Type-II censorship data. In addition, a comprehensive simulation study determines the threshold value of the test statistic, and its power is compared with several existing test methods. Finally, two real-world data analysis cases are included to show the practical application of the method.

Keywords : Critical points, Exponentiality test, Kullback-Leibler information, Local linear regression, Monte Carlo simulation, Power study, Progressive Type-II censoring.

¹Graduate student, Department of Mathematics, Kumoh National Institute of Technology, Gumi 39177, Gyeongbuk, Korea. E-mail: jsh455437@naver.com

²(Corresponding author) Associate professor, Department of Mathematics and Big Data Science, Kumoh National Institute of Technology, Gumi 39177, Gyeongbuk, Korea.

현대화 교육에서의 AI 활용이 교사 교육 방식에 미치는 영향 탐구

양아청¹, 양정호²

요약

인공지능 기술의 급속한 발전으로 고등 교육계는 전례 없는 교수법 혁명을 겪고 있다. 이러한 기술 발전은 정보 전달 방식을 변화시켰을 뿐만 아니라 교육과 학습의 상호 작용 모델을 재정의 했다. 이 전환 기간 동안 대학 교사는 새로운 기술을 습득해야 할 뿐만 아니라 현대 학생들의 요구를 충족시키기 위해 교수 방법을 혁신해야 하는 중요한 역할을 한다. 본 연구는 주로 인공지능이 교사의 교수법에 미치는 종합적인 영향을 연구하며 교사의 정보화 교수능력의 현황 및 수요를 분석하여 현대 교육에서 교사의 교수능력 향상 방법을 탐구한다.

주요용어 : 인공지능, 교수법, 교육사회학.

1. 서론

클라우드 컴퓨팅, 핵심 알고리즘, 인터넷 빅데이터 등 핵심 기술이 돌파하면서 인공지능 발전이 신기원에 접어들었다. 교육 분야는 시대의 발전에 발맞춰 인공지능 산업과의 협력을 모색하기 위해 노력하고 있다. 인공지능 교육은 교사가 인공지능 기술을 사용하여 교육 설계를 수행하고 교육 자원을 제공하며 정확한 교육을 수행하여 학생들의 학습 효과를 향상시키는 활동을 말한다.

인공지능 교육이 수업에 가져오는 변화는 네 가지 측면으로 요약할 수 있다. 첫째, 학습 환경의 지능화이다. AI 기술은 로봇 교육 보조원과 같은 교육 실습에 널리 사용된다. 둘째, 교육 요소의 변화는 교육 조직, 교육 모델 및 학습 방법에 영향을 미친다. 예를 들어, 교수 조직은 전통적인 학급 단위 조직 형태에서 보다 자유롭고 유연한 동적 학습 형태로 바뀔 것이다. 셋째, 교수평가의 변화다. 전통적인 지식 기반 평가에서 지식, 소양 및 능력을 결합한 평가 방법을 강조한다. 넷째, 학교는 전통적인 고정된 장소와 상대적으로 폐쇄된 상태에서 지식을 전파하고 기술을 배양하는 일정 수준의 경계 없는 상태로 전환된다.

제4회 세계 스마트 컴퍼런스 '신 인프라 신교육 신경제 정상포럼'에서 전문가들은 '인간-기계 결합이 새로운 교사 교육 방식이 되고 교사와 인공지능의 결합은 진정으로 상호 보완적인 이점을 실현하고 교사의 업무 효율성을 크게 향상시킬 수 있다'고 제안했다. 한편, 'AI수업'은 인공지능을 통해 학생들의 창의적 사고와 창의력을 발굴하고 학생별 특성에 맞는 학습 방안을 마련하는 맞춤형 교육에 도움이 될 수 있다. 학생 교육 계획 수립 외에도 교육 감독 기능도 나타나고 있다. 한편,

¹03063 서울특별시 종로구 성균관로 25-2 성균관대학교 호암관 50519호, 성균관대학교 교육사회학과 석사과정.
E-mail: thea117@163.com

²(교신저자) 03063 서울특별시 종로구 성균관로 25-2 성균관대학교 호암관 50519호, 성균관대학교 교육사회학과 교수. E-mail: jyang@skku.edu

AI 수업은 컴퓨터 시스템, AR 장면 수업, 데이터베이스 검색 기능 및 전문가 시스템 등 현대적이고 첨단적인 스마트 기술을 활용하여 현 단계에서 각 학교의 전통적인 수업 방식에서 벗어나 학생들이 독특하고 지능적이며 효율적인 장소에서 학습할 수 있도록 하여 새로운 사물을 보다 많이 접할 수 있도록 한다.

2. 현대화 교육 모델의 현황 및 한계

- (1) 교사가 전 과정을 통제하고 학생의 적극성 높지 않다. 교과교사의 전공 배경, 교과목, 개인별 수업 스타일, 학급 규모 등 여러 요인으로 인해 일부 교사들이 소규모 교과과정을 개편하는 것을 제외하고는 대부분의 교사들은 여전히 전통적인 교과과정 위주의 수업방식을 채택하고 있다. 학생의 학습은 교사의 목적적이고 계획적이며 조직적인 지도 하에 이루어진다. 학습 주도자인 학생으로서 수동적으로 주입식 교육을 받는 것은 학생들의 열정을 충분히 동원하기 어렵고 학생들의 자기 인식 능력 향상을 저하한다.
- (2) 커뮤니케이션 방법은 많지만 구현 효과가 좋지 않다. 많은 교사들이 컴퓨터 및 정보 네트워크 기술을 사용하여 교실 콘텐츠를 풍부하게 하고 학생과의 상호 작용을 강화하려고 시도했지만 실행 효과를 더욱 개선해야 한다.
- (3) 현대화를 가르치는 정보 기술 수단을 직시하지 못했다. 컴퓨터와 정보 기술의 발전과 함께 페이퍼리스화는 일종의 학습 방식이 되었다. 하지만 일부 교사들은 이런 행위가 수업의 효과를 떨어뜨리고 정상적인 수업 질서를 저해한다고 낯은 사고와 편견을 갖고 있다.

3. AI 교육 주요 교수 패턴 탐구

(1) 스마트 환경 교육 모델

지혜롭게 공부하는 교수 환경은 가르치는 장소에 대해 비교적 높은 요구를 가지고 있다. 우선 교수환경의 '스마트화'는 스마트 학습장 구축에 필요한 요구다. 네트워크, 컴퓨터, 프로젝터 등은 교육 기반 장비이고 일부 교실에서는 지능인식 시스템, 3D 모델링 도구, AR 장비 등이 AI 교육 고급 장비다. 둘째, 스마트 환경에서의 '가용성'과 '적응성'은 교육의 지속 가능한 발전을 위한 기본 요건이다. 교사들의 교육전개의 지속 가능성을 보장하기 위해 스마트 환경에서의 스마트 기기는 필요할 때 업데이트, 업그레이드된다. 예를 들어, 현 단계에서 많은 멀티미디어 스마트 실험실은 이미 수동 잠금 해제에서 스마트 안면 인식 잠금 해제 시스템으로 전환되어 실내 장비의 안전성을 보장하면서도 수동 잠금 해제로 인한 실수 단계를 피할 수 있다. 현재 주요 AI 교육 활동은 대부분 멀티미디어 교실에서 이루어지고 있으며, 멀티미디어 기기 및 설비가 갖추어져 있지 않은 장소 및 학급에 대해서는 스마트 교습이 질서 있게 이루어지지 않고 있다. 이 외에도 멀티미디어 교실 업데이트 지연, 스마트 환경 기기 노후화 훼손 등이 AI교육의 발목을 잡고 있다.

(2) 스마트 보조 교육 모델

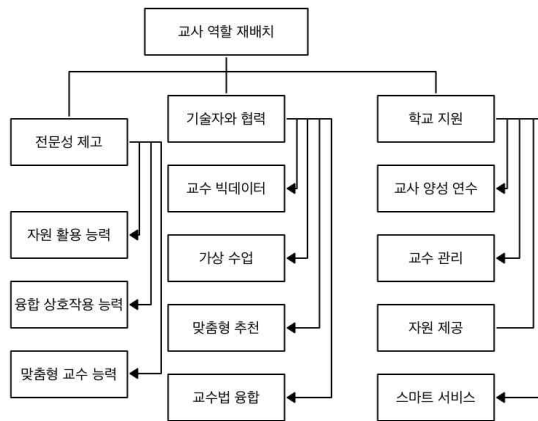
지능형 보조 교육 모델은 AI가 가르치는 가장 흔한 교육 모델이다. AI의 스마트 검색 기능은 학습 자료 찾기, 수업 완료, 사고 확장을 돕는다. AI 시스템의 지능적인 검색을 통해 학생들은 이

전의 책들을 찾는 번거로움과 달리 데이터베이스가 갖추어진 네트워크에서 짧고 정확하게 자원을 찾을 수 있다. 교사가 수업을 준비할 때도 교사의 수업 준비의 질을 향상시켰다. 둘째, AI의 정리 증명 기능은 지능 학습에서 질의응답을 가능하게 하여 학습 사고에 대한 학생들의 계층적 구축을 보조한다. 마지막으로 AI의 프로그램 자동 설계 기능은 수업 중에 자동으로 모범 사례, 카테고리 그룹화, 심지어 디자인 항목까지 열거할 수 있어 학생들이 이론을 실천으로 전환할 수 있는 기회를 제공한다. 교사가 플랫폼과 디자인 사례를 찾는 시간을 절약할 수 있을 뿐만 아니라 학생별로 특화된 사례를 설계해 맞춤형 수업을 추진할 수 있다.

(3) 휴먼 VS 머신 교육 모델

휴먼 대 머신 교육 모델은 현 단계에서 AI 학습의 고급 교육 모델이다. 인공지능의 초기 연구 방향은 인간-기계 대국의 형태를 통해 인공지능 사고의 실용성을 모니터링하는 것이다. 1997년 미국 IBM이 개발한 인공지능 컴퓨터 DEEP BLUE가 대표적이다. 이를 기반으로 한 인공지능 로봇이 점점 더 많이 등장하고 있다. 학생과 AI의 휴먼 머신 대국을 통해 수업을 풍성하게 하면서 학생들의 경기 의식을 높이고 경진대회로부터 지식을 습득하고 사고를 개발한다. 한편, 일부 학교의 문과와 이과 수업 및 취미 수업(바둑, 장기) 등은 모두 휴먼 머신 대국 수업 방식을 채택하기 시작했다. 그러나 로봇 수가 학생에 비해 적기 때문에 휴먼 머신 대국을 할 수 있는 학생 정원은 교사 지정이 필요하고 휴먼 머신 대국을 접할 수 있는 학생 수도 많지 않다. 짧은 교대로 대국을 함으로써 학생들은 소량의 대국 속에서 자신의 창조적 사고를 향상시킬 수 없다. 한편, 스포츠 경기 활동은 AI 장비 자금사정으로 인해 인간-기계의 교육모델이 본격화되지 못하고 있다.

4. 인공지능 교육 배경에서 교사의 역할 재배치



AI 기술의 활용은 교육 효율성을 효과적으로 향상시킬 수 있다. AI가 구축한 방대한 양의 지식 베이스와 편리하고 빠른 검색 엔진은 학생들에게 방대한 양의 개인화된 지식을 제공한다. 교사의 경우 교사가 적성에 맞게 가르치고 학생의 학습 상황에 따라 자동으로 문제를 출제, 채점 및 평가할 수 있도록 도울 수 있다. 교사는 번거로운 교수자료 수집, 학력기록, 지식평가 등의 업무에서 벗어나 학생의 인격형성, 대인관계, 도덕수양, 정서감정, 혁신 창의력, 고급사고력, 협동력 등의 배

양에 중점을 둘 수 있다. 사실 인공지능 교육은 교사에게 더 높은 요구 사항을 제시하며 교사는 지속적으로 새로운 지식과 기술을 배우고 자신의 인공지능 교육 소양을 함양하기 위해 노력해야 한다.

인공지능 시대의 교사는 학생의 인격을 이끄는 사람, 학생이 함께 공부하는 파트너, 학습 서비스 제공자, 교육 분야의 전문 기술자, 인공지능 교육의 평가자 되어야 한다. 따라서 교사는 협력 의식을 확립하고 협력자의 자세로 인공지능 교육 가치 사슬에 진입하여 의사소통을 잘하고 협력을 이해하는 훌륭한 파트너가 되어야 한다.

References

- LUAN Dandan, WU Han & MA Xiaoqing.(2024).Exploration into the Reform of Distance Education Teaching Methods under the Application of AI Technology. *Information & Computer* (17),220-222.
- YU Dan.(2024).Improving the Informatized Teaching Ability of University Teachers in the Context of “5G+AI”. *The world of digital communication* (07),189-191.
- Ma Ning.(2024). An attempt to improve classroom teaching methods based on AI technology. *New Curriculum Study* (06), 77-79.
- Bie Dunrong.(2024). The theoretical explanation of the application of AI technology to university education teaching. *Teaching in Chinese universities*(05),4-9+2.
- Zhang Yuchao, Xue Xiaoxiang, Zhang Yifan, Liu Wei & Kwong Kin. (2024). AI - assisted digital education teaching methods research and practice. *Computer education* (05), 118-122.

간호대학생의 자기주도학습, 학업적 자기효능감, 비판적 사고경향 및 셀프리더십의 관계

이은미¹, 오윤정²

요 약

본 연구는 간호대학생의 자기주도학습, 학업적 자기효능감, 비판적 사고경향 및 셀프리더십의 관계를 확인하기 위한 서술적 조사연구이다. 연구대상자는 G시 소재 대학에 재학 중인 간호대학생 208명을 대상으로 하였으며, 자료수집은 자가보고식 설문지를 이용하여 2024년 10월 10일에서 10월 20일까지 실시하였다. 수집된 자료는 SPSS 25.0 proram을 이용하여 기술적 통계, t-test, ANOVA, Pearson correlation coefficient을 시행하였다. 연구결과 대상자의 일반적 특성에 따른 자기주도학습은 성별($t=4.27, p=.001$), 학업성취도($F=5.48, p=.001$)에 따라, 학업적 자기효능감은 학업성취도($F=5.48, p=.001$), 전공만족도($F=33.16, p<.001$)에 따라 유의한 차이를 보였다. 비판적 사고경향은 성별($t=3.57, p<.001$), 학업성취도($F=5.94, p=.001$), 전공만족도($F=10.46, p<.001$), 선호하는 학습방법($F=2.96, p=.021$)에 따라, 셀프리더십은 성별($t=2.08, p=.006$), 학업성취도($F=6.45, p<.001$)에 따라 유의한 차이가 있는 것으로 나타났다. 대상자의 학업적 자기효능감은 자기주도학습($r=.304, p<.001$)과 유의한 관계를 보였고, 비판적 사고경향은 자기주도학습($r=.522, p<.001$)과 학업적 자기효능감($r=.577, p<.001$), 셀프리더십은 자기주도학습($r=.698, p<.001$), 학업적 자기효능감($r=.383, p<.001$), 비판적 사고경향($r=.609, p<.001$)과 유의한 상관관계를 보였다. 따라서 급변하는 의료환경 속에서 수준 높은 간호가 요구되는 상황에 적응하기 위해 간호대학생이 자기주도학습능력, 학업적 자기효능감과 비판적 사고성향을 적극적으로 활용하여 셀프리더십 수준을 함양할 수 있는 교과목 및 교수학습방법 개발 등 간호교육측면에서의 관심과 노력이 요구된다. 또한 간호대비교과 프로그램 개발 및 적용을 통해 간호대학생의 셀프리더십 향상을 돕는 과정이 필요하다.

주요 용어 : 간호대학생, 자기주도학습, 학업적 자기효능감, 비판적 사고경향, 셀프리더십.

¹광주광역시 광산구 첨단중앙로 23, 남부대학교 간호학과 조교수. E-mail: jjang9312@nambu.ac.kr

²광주광역시 광산구 첨단중앙로 23, 남부대학교 간호학과 조교수. E-mail: Agnes@nambu.ac.kr

생존확률 외삽을 위한 최적 노츠 탐색 연구

정완석¹, 남진현²

요약

대부분의 임상 시험은 제한된 기간 동안 진행되므로 장기 생존 데이터를 제공하지 못한다. 따라서 생존 데이터를 외삽하는 과정이 필수적이다. 특히 암과 같은 만성 질환에서는 시간이 지남에 따라 치료 효과가 어떻게 변하는지 평가하고, 이를 기반으로 비용 효과성을 판단하기 위해 생존 외삽이 중요한 역할을 한다. 본 연구는 실제 데이터를 활용하여 생존 확률 외삽 모형을 비교한다. 모수적 모형과 표준 스플라인 모형을 사용하고 표준 스플라인 모형 같은 경우 노츠의 위치를 지정하지 않아 생존 시간의 흐름에 따른 중요한 통계적 변화를 포착할 수 없기 때문에, 적절한 노츠의 위치를 지정하는 것이 필수적이다. 본 연구에서는 변화점 탐지 방법을 적용하여, 생존 함수에서 통계적 변화가 발생하는 지점을 노츠로 설정하였다. 변화점 탐지 방법은 4가지 방법과 5가지 별점을 사용해 노츠의 위치를 탐지한다. 탐지된 노츠를 스플라인 모형에 적용하여 외삽된 생존 확률을 예측한 결과, 모수적 모형과 표준 스플라인 모형보다 더 우수한 성능을 보였다. 이는 임의로 노츠를 설정하지 않고, 데이터 기반의 통계적 변화를 반영한 결과이다. 본 연구에서 제안한 변화점 탐지 방법 기반 외삽 모델은 기존의 임의 노츠 설정 방식보다 더 정확하고 신뢰성 높은 생존 확률 예측을 제공한다. 이를 통해 장기 생존 데이터가 필요한 건강 기술 평가와 임상 연구에서 보다 정교한 치료 효과 분석과 비용 효과성 평가를 수행할 것을 제안한다.

주요 용어 : 외삽, 노츠, 스플라인 모형, 생존확률, 변화점 탐지 방법.

¹30019 세종특별자치시 세종로 2511, 고려대학교 일반대학원 응용통계학과 석사과정.

E-mail: frer9547@korea.ac.kr

²(교신 저자) 세종특별자치시 세종로 2511, 고려대학교 빅데이터사이언스학부 부교수.

E-mail: jinhnam@korea.ac.kr

서울시 공공자전거 데이터 분석을 통한 도시 문제 개선방안 연구

권선일¹, 최호식²

요약

서울시 공공자전거 시스템 '따릉이'는 2015년 서비스 개시 이래 꾸준히 성장해 도심 내 지속 가능한 교통수단으로 자리 잡으며 시민들의 편의성을 높이고 탄소 배출량 감소에 큰 도움을 주고 있지만 운영 비용 증대로 고질적인 적자 상황에 놓여있다. 따라서 '따릉이'의 관리 효율성과 지속 가능한 발전을 위해 더욱 상세하고 심층적이며 포괄적인 연구가 필요한 상황이다. 기존 '따릉이' 관련 연구들이 주로 수요 예측에 초점을 맞췄다면, 본 연구는 대여소의 공간적 배치와 구조적 조정, 그리고 시민들의 '따릉이' 사용 패턴 이해를 통해 시스템의 운영 효율성을 극대화하고자 한다. 이를 위해 서울시 전역에 위치한 자전거 대여소와 같은 공간 데이터 뿐만 아니라 기존 자전거 이용 패턴, 지역별 소득 수준, 인구 통계 등 사회 경제적 요인을 포함한 여러 차별화 요소들을 결합 분석하여 대여소의 최적 위치 및 운영 전략 관련 인사이트를 도출해 보고자 한다. 본 연구 결과로 '따릉이' 관리 시스템의 효율성 제고와 함께 시민들의 만족도, 도시 교통 혼잡 감소와 같은 교통 문제 해결과 도시 환경 개선 기대 효과를 파악하여 '따릉이'가 지속 가능한 도시 발전에 기여하는 새로운 접근법을 제시한다.

주요 용어: 서울시 공공자전거, 따릉이, 시각화, 지오코딩, 자전거 이용패턴.

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

¹02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 도시빅데이터융합학과 석사과정,
E-mail: hardid@uos.ac.kr

²(교신저자)02504 서울시 동대문구 시립대로 163, 서울시립대학교 도시빅데이터융합학과 교수,
E-mail: choi.hosik@uos.ac.kr

50세 이상 성인여성의 골관절염 유병률과 관련 요인

전은미¹

요 약

본 연구는 50세 이상 성인여성의 골관절염 유병 관련 영향요인을 확인하고 그에 따른 골관절염 관리를 위한 중재방안 개발을 위한 기초자료를 제공하고자 한다. 본 연구대상자는 2016~2020년 국민건강영양조사 자료에서 응답한 39,738명 중 50세 이상 성인여성은 9,867명이었고 이 중 골관절염 유병여부에 응답한 총 9,590명을 대상으로 하였다. 수집된 자료는 실수와 백분율, 집단(골관절염 유병 유무)간 비교는 교차분석을 이용한 Rao-Scott 카이제곱 검정을 하였다. 골관절염 유병 여부에 영향을 미치는 요인은 로지스틱회귀분석을 이용하였다. 연구결과 한국 50세 이상 여성의 골관절염 유병률은 23.8%이었다. 골관절염 유병률에 통계적 차이를 보이고 있는 요인으로는 연령, 교육수준, 배우자유무, 거주지역, 음주, 유산소 신체활동, 걷기, 비만여부, 우울여부 등이었다. 또한 골관절염 유병률에 영향을 미치는 요인으로 연령, 교육수준, 비만, 우울로 나타났다. 연령은 50대를 기준으로 했을 때 60대는 골관절염 유병 위험률이 2.3배, 70대는 3.3배, 80대는 3.5배로 높아지고, 교육수준에서는 대졸에 비해 중졸은 골관절염 유병 위험이 1.8배, 초졸이하는 2.3배로 높아짐을 알 수 있다. 한편 비만인 경우, BMI지수가 24이하인 경우를 기준으로 했을 때 BMI 25이상은 골관절염 유병 위험이 1.05배 높아지고, 우울에서는 우울이 없는 경우 골관절염 유병 위험이 0.93배 낮아지고 있음을 볼 수 있다. 이상의 결과를 바탕으로 성별에 따른 다각적이고 개별적인 간호중재프로그램 개발이 필요하다.

주요 용어 : 성인여성, 골관절염, BMI, 로지스틱회귀분석.

Investigating Meteorological and Environmental Effects on COVID-19 Transmission in South Korea^{*}

Hyunmin Lee¹, Boseung Choi²

Abstract

This study comprehensively investigates the impact of meteorological and environmental factors on the spread of COVID-19 across 17 regions in South Korea using a linear regression model. Using the number of confirmed COVID-19 cases from January 2020 to November 2021 as the dependent variable, key meteorological and environmental factors such as average temperature, average relative humidity, nitrogen dioxide (NO), carbon monoxide (CO), fine particulate matter (PM_{2.5}), ozone (O), and ultraviolet (UV) index were selected as explanatory variables to analyze their influence on the spread of COVID-19. To enhance the reliability of the analysis, data from the influenza period, spanning from September 2017 to August 2019, was analyzed using the same methodology in addition to the COVID-19 data during the 2020 pandemic. This allowed for the confirmation that the results from the COVID-19 analysis were not confined to a specific situation but could be generalized. A heatmap analysis using t-values revealed that most meteorological and environmental factors exhibited consistent directional influences across South Korea, with statistically significant results. These findings indicated regional differences, with the eastern regions of South Korea showing a more pronounced influence from meteorological and environmental factors compared to the western regions. This analysis demonstrated that these factors exert complex and multidirectional effects on the transmission of COVID-19.

keywords : COVID-19, Meteorological factors, Environmental factors, Influenza, Simple linear regression.

^{*}This research was supported by a grant of the project for The Government-wide R&D to Advance Infectious Disease Prevention and Control, Republic of Korea (grant number: HG23C1629)

¹Graduate student, Department of Economic Statistics, Korea University, Sejong, 30019, Korea.
E-mail:railami98@korea.ac.kr

²(Corresponding Author) Professor, Division of big data science, Korea University, Sejong, 30019, Korea.
E-mail:cbskust@korea.ac.kr

네트워크 모델링 접근에 기반한 확률적 전염병 모델 개발*

박경호¹, 최보승²

요약

SIR 모형은 감염병의 확산 과정을 수리적인 방법으로 설명하는 대표적인 모형으로, 전체 집단을 세 그룹(S, I, R)으로 나누어 감염병 전파 과정을 설명한다. 사람과 사람 간의 접촉에 의한 감염병 확산은 네트워크 모형을 이용하여 효과적으로 모형을 구축할 수 있다. 본 연구는 감염병 확산 모형의 정확성을 높이기 위해 네트워크 모델링을 활용한 확률적 전염병 모형에 대한 통계적 추론 방법을 제안한다. 엣지 기반 구획 모형(EBCM; Edge Based Compartment model)과 pairwise 모형을 사용하여 네트워크 상에서 전염병의 확산 과정을 모형화하였다. 구성 네트워크 모델(Configuration Model)을 활용하여 개체 간의 랜덤한 접촉을 설명하며, 확률분포를 통해 연결 구조를 분석한다. 확률 과정은 동적 생존 분석(DSA) 모델에 적용되어 감염 시간, 감염 기간과 회복 시간을 추정하기 위한 우도 함수(likelihood function)를 구성하는데 활용되었다. 또한, 베이지안 추론 및 MCMC 기법을 통해 모형의 주요 모수인 감염률과 회복률, 그리고 기초감염재생산수(R_0)를 추정하였다. 시뮬레이션을 과정에서 차수(degree) 분포를 음이항 분포와 포아송 분포로 따르는 데이터를 생성하고, EBCM과 pairwise 모형으로 적합시켰다. 추가적으로 EBCM에서 유행률의 정점을 설정하여 부분적으로 모형을 최적화하였다. 결과적으로, 모수 추정과 R_0 추정을 통해 다양한 모델 간 성능을 비교하였으며, 최적의 모델을 개발하였습니다. 이를 통해 보다 현실적이고 신뢰성 있는 감염병 확산 추정 모형을 구축하였으며, 새로운 질병 확산 추정방법을 제시하였다.

주요 용어 : SIR 모형, 확률적 전염병 모델, 엣지 기반 구획 모형(EBCM), Pairwise 모델, 동적 생존 분석(DSA).

*본 연구는 범부처의 재원으로 범부처 감염병방역체계 고도화 R&D 사업 지원에 의하여 이루어진 것임 (HG23C1629).

¹(30019) 세종특별자치시 세종로 2511, 고려대학교 일반대학원 응용통계학과, 석사과정.

E-mail: rudgh109503@korea.ac.kr

²(30019) 세종특별자치시 세종로 2511, 고려대학교 빅데이터사이언스학부, 교수. E-mail: cbskust@korea.ac.kr

Hi-C 데이터를 활용한 희소 가중 그래프 기반 대장암 예측 모형에 관한 연구*

고민규¹, 임지원¹, 장인수^{1,2}, 박민수¹

요약

대장암은 환자들에게 심각한 신체적·심리적 영향을 미치는 주요 암 중 하나로, 조기 진단은 생존율을 높이고 삶의 질을 개선하는 데 핵심적인 역할을 한다. 본 연구에서는 유전체의 3차원 구조를 나타내는 high-throughput chromosome conformation capture (Hi-C) 데이터를 활용하여 대장암 예측을 위한 새로운 프레임워크를 제안한다. Hi-C 데이터는 유전체 내 상호작용 정보를 포함하며, 비정상적인 유전자 발현 및 암 발달과 관련된 중요한 단서를 제공한다. 제안된 프레임워크는 유전자 구역 간의 유의미한 상호작용을 선별적으로 식별하여 희소 가중 그래프(sparse weighted graph) 구조로 변환하고, 이를 기반으로 그래프 신경망(graph neural network)을 활용하여 대장암 여부를 예측한다. 실험 결과, 제안된 방법은 기존 접근법에 비해 예측 성능과 계산 효율성 면에서 향상된 결과를 보였으며, 학습 과정에서 더 안정적인 수렴 양상을 나타냈다. 이 연구는 Hi-C 데이터를 활용한 예측 모형의 가능성을 입증하며, 유전체 데이터 기반의 질병 진단 및 분석에서 새로운 접근법을 제시한다.

주요 용어 : High-throughput chromosome conformation capture, 희소 가중 그래프, 그래프 신경망, 대장암.

*본 연구는 한국 정부가 지원하는 한국연구재단의 연구비 지원을 받아 수행된 연구임(NRF-2022M3J6A1084843; RS-2024-00337752).

¹34134 대전광역시 유성구 대학로99, 충남대학교 통계데이터사이언스학과.

²34141 대전광역시 유성구 과학로125, 한국생명공학연구원.

메타데이터와 텍스트 데이터를 결합한 온라인 리뷰 유용성 예측 모델 및 해석 가능성 연구*

김유영¹, 최호식²

요약

이커머스 시장의 확장에 따라 소비자 경험을 개선하기 위해 리뷰 유용성 예측에 관한 연구가 중요해지고 있다. 온라인 리뷰 텍스트 내 특정 단어나 표현이 포함될 때 소비자들이 이를 유용하다고 느끼는지에 대해 탐구할 필요성이 있으며, 이를 통해 어떤 특성들이 소비자 경험에 긍정적인 영향을 미치는지를 밝힐 필요성이 있다. 본 연구는 소비자들이 유용하다고 평가하는 온라인 리뷰의 특성을 파악하고자, 메타데이터와 리뷰 텍스트 정보를 결합하여 리뷰 유용성을 예측하는 모델을 개발하는 것을 목표로 한다. 이를 위해 XGBoost, Random Forest, SVM, LogisticRegression 등 머신러닝 모델과 BERT, Transformer, LSTM 등 딥러닝 모델을 적용하여 예측 성능을 비교하고, 가장 우수한 성능을 보이는 모델을 선정한다. 아울러, XAI 기법을 통해 소비자들이 리뷰를 유용하다고 판단하는 요인들을 식별하고, 메타데이터와 텍스트 데이터가 유용성 평가에 미치는 영향을 해석한다. 연구를 통해, 이커머스 플랫폼에서 소비자 경험 개선을 위한 전략 수립에 기여할 것으로 예상되며, 이를 통해 상품 기획과 마케팅의 효율성을 높이는 데 중요한 시사점을 제공할 것으로 기대된다.

주요 용어: 리뷰 유용성, 머신러닝, 딥러닝, eXplainable Artificial Intelligence (XAI).

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

¹02504 서울시 동대문구 시립대로 163, 서울시립대학교 도시빅데이터융합학과, 석사과정.

E-mail: yy.kim@uos.ac.kr

²(교신저자)02504 서울시 동대문구 시립대로 163, 서울시립대학교 도시빅데이터융합학과 교수.

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

학과목 학원 금지 정책이 중국 지방 한자녀 가정에 미치는 영향 분석

도과우¹, 양정호²

요 약

2021년 7월, 중국은 ‘의무교육단계 학생의 학업부담 및 사교육 부담 경감에 관한 의견(이하 ‘쌍감정책’)’ 을 발표하였다. 이 정책의 효과 및 문제점을 세부적으로 분석해야 ‘쌍감정책’ 이 정책의도대로 성과를 낼 수 있다. 본 연구는 중국의 학과목 학원 금지 정책이 한자녀 가정에 미친 영향을 파악하기 위해, 심층 인터뷰와 자료 분석을 통한 질적 사례 연구 방법을 사용하였다. 중국 남부 지방의 한자녀 가정의 학부모 8명을 대상으로 심층 인터뷰를 진행하였다. 연구결과는 두 부분으로 나눌 수 있다. 학업 부담 경감 효과 미미, 가정 경제 부담 및 진학 불안감 증가, 교육 평등에 부정적인 영향, 둘째 출산 의향 감소 같은 소극적 영향이 있는 반면, 자녀의 취미와 특기 개발에 유리, 온라인 수업을 통한 교육 지출 및 통학 시간 감소, 교육 평등에 긍정적인 영향 등 적극적 영향도 있다. 지금 중국에 교육제도 현실에 따라 학부모의 사교육 수요가 존재하는 한 학생들의 학업 부담 경감 문제는 근본적으로 해소될 수 없다. 그래서 극단적인 일괄 금지 방식을 취하기보다는 지역별 경제 발전 상황과 교육 자원 수준에 따라 적절한 정책을 시행하는 것이 바람직하다.

주요 용어 : 사교육, 학과목 학원 금지 정책, 쌍감(双减)정책, 한자녀, 질적연구.

1. 서론

2021년 7월, 중국은 ‘의무교육단계 학생의 학업부담 및 사교육 부담 경감에 관한 의견(關於進一步減輕義務教育階段學生作業負擔和校外培訓負擔的意見)(이하 ‘쌍감정책’)’ 을 발표하였다. 해당 정책에 따르면, 중국 각지에서는 의무교육 단계 학생을 대상으로 하는 학과목 관련 학원의 신규 허가가 중단되었으며, 기존 학과 수업 학원은 모두 비영리 기관으로 등록하게 되어야 한다. 또한 미성년자 보호법의 관련 규정을 엄격히 시행하기 위해, 학원이 국가 법정 공휴일, 휴일 및 방학 기간 동안 학과 수업을 조직하는 것을 금지하고 있다. 현재 이 정책이 시행된 지 3년이 지났으며, 이 정책이 학생들의 학업 부담 경감에 긍정적인 영향을 미쳤는지, 중국의 학생 가정에 어떤 문제를 초래했는지, 중국 학부모들이 이 정책을 어떻게 바라보고 있는지, 이러한 문제를 명확히 파악해야 ‘쌍감정책’ 의 성과를 파악할 수 있다.

본 연구는 질적 사례 연구를 통해 중국 지방의 한자녀 가정을 대상으로 심층 인터뷰를 진행하였다. 인터뷰 내용을 바탕으로, 학과목 관련 학원 금지 정책이 중국의 한자녀 가정에 미친 영향을 분석하고, 이 정책의 문제점을 제시함으로써 중국 학업 부담 경감 정책의 개선방안을 제시하려고 한다.

¹03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 석사과정. E-mail: yasssiamhere@gmail.com

²(교신저자) 03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 교수. E-mail: jyang@skku.edu

2. 연구 방법

본 연구는 중국의 학과목 학원 금지 정책이 한자녀 가정에 미친 영향을 파악하기 위해, 상세한 심층 인터뷰와 자료 분석을 통한 질적 사례 연구 방법을 사용하였다.

본 연구는 중국 남부 지방의 한자녀 가정의 학부모 8명을 대상으로 심층 인터뷰를 진행하였다. 총 8명의 학부모가 심층 인터뷰에 참여하였으며, 면접 참여자의 기본 정보는 다음과 같다.

Table1. 심층 인터뷰 참여자 세부정보

이름	직업	자녀 연령	자녀 학년	가정 연간 수입	사교육 참여 유무	사교육 참여 유형	사교육비
Xxx	공무원	8	초등학교 3학년	5천만	있음	1. 숙제 지도(보육) 2. 축구(학원, 취미)	1. 78만/학기, 월~금 2. 30만/학기, 월~금
Qx	교사	15	중학교 3학년	2억	있음	1. 수학(그룹 과외) 2. 물리(그룹 과외)	1. 4만/1h30 2. 4만/1h20
Zxx	회계원	12	초등학교 6학년	5억 5천만	있음	1. 국어(그룹 과외) 2. 수학(1대3) 3. 영어(1대4)	1. 58만/학기 2. 100만/여름 방학 3. 290만/년
Yxx	불꽃놀이 공장	11	초등학교 5학년	5억	있음	숙제 지도(보육)	87만/학
Xx	회계원	10	초등학교 4학년	1억	있음	피아노(1대1, 취미)	4만/h
Lx	공무원	14	중학교 2학년	6천 4백만	있음	1. 수학 올림피아드(그룹 과외) 2. 서예(학원, 취미) 3. 춤(학원, 취미)	1. 6만/h 2. 30만/15 교시 3. 50만/80 교시
Cx	공무원	10	초등학교 4학년	8천 7백만	있음	1. 춤(학원, 취미) 2. 피아노(1대1, 취미) 3. 그림(학원, 취미) 4. 서예(학원, 취미)	1. 300만/년 2. 480만/년 3. 97만/년 4. 42만/15 교시
Cxx	슈퍼마트 구매 담당자	14	중학교 2학년	2천만	있음	1. 수학(그룹 과외) 2. 영어(1대1)	1. 32만/10 교시 2. 6만5천/2h

3. 연구 결과

3.1. 소극적 영향

1) 학업 부담 경감 효과 미미

비록 학과목 학원을 금지했지만, 다른 형태의 사교육은 금지되지 않았다. 또한, 현재 중국의 교육 제도에 따르면 입시 경쟁이 여전히 치열하다. 인터뷰에 따라 이런 말이 많이 나오고 있다. “반에서 학생들이 거의 다 사교육을 받았는데, 제 아이만 안 받으면, 좋은 학교로 진학할 수 없으면 어떡하지?” 학부모들은 진학에 대한 우려를 많이 전달했다. 그래서 중국 학부모들이 자녀에게 사교육을 시키려는 수요가 있는 한, 사교육과 학생들의 과중한 학업 부담 문제는 근본적으로 해결되지 않을 것이다.

그리고 학과목 학원 금지 정책이 발표된 동시에, 중국 초등학교는 방과 후 학습 시간을 증가했다. 이에 따라 학생들이 학교에서 보내는 시간이 두 시간 더 늘어났으며, 일부 학생은 이후에도 과제 지도를 받으러 가야 한다. 따라서 학생들의 학업 부담은 여전히 존재하며, 예상했던 경감 효과를 얻지 못했다.

2) 가정 경제 부담 및 진학 불안감 증가

인터뷰에 따라, 모든 학부모는 “학원보다 일대일 등 사교육 방식은 교육 지출이 더 높다”라는 대답을 했다. 그러나 중국의 교육 제도와 교육열 때문에 대부분의 가정은 경제적 부담을 지고서라도 자녀를 사교육에 참여시키고 있다.

그리고 중국 지방 학생의 부모는 경제 발달 지역 부모에 비해 보편적으로 교육 수준이 낮다. 일부 학부모들이 “가정에서 자녀 학습을 잘 지도할 수 있을지, 그리고 자녀가 학원을 다니지 않으면 성적이 떨어지지 않을지, 다른 아이들보다 뒤처지지 않을지”에 대해 걱정한다.

3) 교육 평등에 부정적인 영향이 있음

학과목 학원은 금지되었지만, 다른 형태의 사교육은 금지되지 않았다. 인터뷰에 응한 한 아버지는 “집안이 부유하고 성적이 좋을수록 아이는 사교육을 더 많이 받다”고 말했다. 그래서 경제적으로 여유 있는 가정은 여전히 자녀를 사교육에 보내고 있으며, 경제적으로 어려운 가정은 추가적인 사교육을 제공하기 어려운 상황에 직면하게 된다. 이로 인해 중국 학생들 간 격차가 커지고 교육 불평등이 심화될 가능성이 있다.

4) 둘째 출산 의향 감소

대부분의 학부모들은 “지금 아이를 잘 키우려면 돈이 너무 많이 든다”고 표시했다. 일부 학부모들은 “현재 국가가 ‘가정-학교 공동 교육’을 실시하고 있으며 집에서 아이에게 숙제를 가르치는 것이 너무 힘들어서 아이가 숙제 지도 학원으로 보내다”고 말했다. 또 “주변 친구들이 다 둘째 출산을 원하지 않다”고 이야기했다. 그래서 교육 지출 증가로 인해 일부 가정은 두 자녀의 생활과 교육을 감당하기 어려워져 둘째 자녀 출산 의향이 줄어들었다.

3.2. 적극적 영향

1) 자녀의 취미와 특기 개발에 유리

학과목 학원이 금지됨에 따라 학생들은 휴일과 방학에 더 많은 자유 시간을 가질 수 있게 되었다. 인터뷰에 따라 일부 학부모들이 진학에 대한 좀 걱정을 하지만 여전히 “아이가 어려서부터 사회와 자연에 대한 인식을 양성할 필요가 있다”라는 의견을 전달했다. 그들은 자녀들이 이 자유 시간을 통해 자신의 취미와 특기를 개발하고, 자신의 소양을 높일 수 있다고 생각했다.

2) 온라인 수업을 통한 교육 지출 및 통학 시간 감소

오프라인 학과목 학원에 참여하지 못할 경우, 집에서 온라인 수업을 통해 학습할 수 있다. 연구 참여한 학부모 중 일부는 “온라인 수업은 교육비용이 낮고, 부모가 자녀를 데리고 다니는 데 소요

되는 통학 시간도 줄일 수 있다”고 답했다. 그래서 그들의 자녀가 오프라인 사교육 외에도 온라인 수업도 참여했다.

3) 교육 평등에 긍정적인 영향

인터뷰 자료에 따르면, 학원 금지 정책을 지지하는 학부모도 있다는 것을 나타냈다. 한 어머니는 “이 정책은 인성교육에 대한 국가의 관심을 반영했다. 하지만 이 정책이 실제로 효과가 거두기까지는 갈 길이 멀다”라는 견해를 말했다. 학원 수업을 금지함으로써 모든 학생이 학원에 참여할 수 없게 되어, 동일한 출발선에서 경쟁하게 된다. 교육 공평에 부정적인 영향을 미치는 반면, 교육 평등 발전에 긍정적인 영향을 미치고 있다.

4. 결론 및 제언

인터뷰 결과에 따르면, 학과목 학원 금지 정책이 중국 지방 한자녀 가정에 미친 부정적 영향이 긍정적 영향보다 컸다. 주요 문제는 중국의 교육 자원 배분이 불균형하며, 현재 교육 제도가 여전히 입시 위주라는 점이다. 이로 인해 경제 발달 지역에서는 학교의 교사진이 풍부하고, 학생들의 성적이나 지식 수준이 경제 낙후 지역의 학생보다 우수하며, 이러한 가정은 학과목 학원에 참여하지 않더라도 충분한 교육 자원을 활용해 다른 사교육에 참여할 수 있다.

그러나 중국 지방의 학생들은 교육 자원의 질이 낮아, 현재의 교육 제도에 맞추어 학과목 학원을 통해 성적을 올릴 수밖에 없다. 학과목 학원 금지 정책은 가정의 경제적 부담을 증가시켰을 뿐만 아니라, 어느 정도 학생들의 경쟁력을 낮췄다.

비록 정책의 출발점은 좋지만, 극단적인 일괄 금지 방식을 취하기보다는 지역별 경제 발전 상황과 교육 자원 수준에 따라 적절한 정책을 시행하는 것이 바람직하다. 예를 들어, 학생들이 일주일 동안 사교육을 받을 수 있는 시간을 정해두면, 학생들의 학업 부담을 덜어줄 수 있을 뿐만 아니라 학부모들이 자녀의 사교육을 원하는 요구도 충족시킬 수 있다.

References

- Ministry of Education of the People's Republic of China. (2021). 中共中央辦公廳 國務院辦公廳印發《關於進一步減輕義務教育階段學生作業負擔和校外培訓負擔的意見》. http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202107/t20210724_546576.html
- Gao, P., Wang, B, C. (2022). Analysis of the Implementation Effect and Optimization of the “Double Reduction” Policy—Qualitative Research on Proposals and Motions from Deputies of NPC and CPPCC based on NVivo. *Social Governance Review*, 2022(7), 21-28. <https://doi.org/10.16775/j.cnki.10-1285/d.2022.07.012>. (in Chinese).
- Qi, Z, Y., F, J. (2023). Multiple Dilemmas in the Implementation of “Double Reduction” Policy from the Perspective of Stakeholders and Its Mitigation Strategies. *Journal of South China normal University (Social Science Edition)*, 2023(6): 56-71. (in Chinese)
- Zou, B, L., Zhang, M, K. (2024). The Influence of “Double Reduction” Policy Cognition on Parents’ Educational Anxiety, *Education Science Research*, 2024(7), 13-21. (in Chinese).
- Kim, E, G., Cho, E, S., Zhu, M, I. (2024). Paradoxes Unveiled: Examining the Implementation of China’s Double Reduction Policy, *The Journal of Politics of Education*, 31(1), 1-25. <https://dx.doi.org/10.52183/KSPE.2024.31.1.1>. (in Korean).

- Yoon, T. H. (2023). Research on the Impact of the Double Reduction(Shuang Jian) Policy on Private Education during the Xi Jinping Era, *Journal of Asia-Pacific Studies*, 30(4), 61-96. <https://doi.org/10.18107/japs.2023.30.4.003>. (in Korean).

불법 온라인 성매매 광고 분류 및 자동 신고를 위한 다중 모달 시스템 개발*

박민회¹, 박수민², 윤성범³, 김준철⁴, 최호식⁵, 송경우⁶

요약

온라인 성매매 홍보 불법 콘텐츠는 인터넷과 소셜 네트워크 서비스(SNS)의 확산으로 심각한 사회적 문제가 되고 있으며, 특히 아동 및 청소년 대상의 성범죄는 즉각적이고 효과적인 대응이 필요합니다. 기존 대응 방식은 주로 사후 처벌에 중점을 두어 사전 예방과 조기 탐지에 한계가 있었습니다. 본 연구는 이러한 문제를 해결하기 위해 X API를 활용하여 수집된 텍스트와 이미지 데이터를 기반으로 성범죄 징후를 탐지하고, 신고 과정을 자동화하는 AI 기반 시스템을 제안합니다. 본 시스템은 멀티모달 대형 언어 모델(Multimodal Large Language Model, MLLM)을 활용하여 텍스트와 이미지 데이터를 통합 분석하며, 성범죄의 징후를 탐지하고 자동으로 신고서 초안을 생성합니다. 신고서는 사람의 검토를 거쳐 최종 제출되어, 신속성과 정확성을 모두 확보할 수 있습니다. 실험 결과, 멀티모달 데이터의 결합이 텍스트나 이미지 단독 분석보다 더 높은 성능을 보였으며, 신고 자동화 절차를 통해 효율성을 증대시켰습니다. 본 연구는 성범죄 예방 및 피해자 보호를 위한 실질적 도구로서 온라인 성매매 홍보 불법 콘텐츠에 대한 새로운 대응 방향을 제시하며, 사회적 안전망 구축에도 기여할 수 있을 것으로 기대됩니다.

주요 용어: 성범죄 자동 신고 시스템, 멀티모달 인공지능, 이미지 및 텍스트 분석, SNS 콘텐츠 검출, 불법 콘텐츠 예방.

*이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NO.2022M3J6A1084845), 서울연구원(2024-PR-39, 서울 디지털 성범죄 통합대응체계 구축방안)의 지원과 2024년도 식품의약품안전처 연구개발비(22183MFDS431)로 수행된 연구임.

¹25504 서울특별시 동대문구 서울시립대로 163 서울시립대학교 도시빅데이터융합학과 석사과정.

E-mail: minhoi.park@uos.ac.kr

²03722 서울특별시 서대문구 연세로 50 연세대학교 통계데이터사이언스학과 석사과정.

E-mail: 2019122009@yonsei.ac.kr

³06756 서울특별시 서초구 남부순환로 340길 57 서울연구원 AI빅데이터랩 연구원 E-mail: yunsb33@si.re.kr

⁴3(교신저자) 06756 서울특별시 서초구 남부순환로 340길 57 서울연구원 AI빅데이터랩 연구위원.

E-mail: kjc@si.re.kr

⁵(교신저자) 025504 서울특별시 동대문구 서울시립대로 163 서울시립대학교 도시빅데이터융합학과 교수.

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

⁶(교신저자) 03722 서울특별시 서대문구 연세로 50 연세대학교 응용통계, 통계데이터사이언스학과 교수.

E-mail: kyungwoo.song@yonsei.ac.kr

동형암호를 통한 이중 데이터 결합 및 분석의 실용성 검토*

최민호¹, 정효혁², 최호식³, 신성철⁴

요약

데이터 활용 및 데이터 분석 기술이 발전함에 따라 다양한 산업에서 민감한 데이터를 바탕으로 한 고도화된 분석 수요가 증가하고 있다. 특히 금융, 의료, 교통 분야에서는 개인정보를 포함한 데이터를 안전하게 다루기 위한 관리가 중요하다. 하지만 현재의 가명처리 및 데이터 결합 신청 절차는 복잡하며, 민감한 개인정보 유출의 위험성으로 인해 데이터 보호와 분석 효율성을 동시에 달성하는 데는 기술적 어려움이 존재한다. 이를 해결하기 위한 대안으로 데이터가 암호화된 상태에서도 연산이 가능한 동형암호 기술이 주목받고 있으며, 동형암호를 활용한 다양한 연구가 진행 중이다. 본 연구에서는 동형암호를 활용하여 택시 위치 데이터와 택시 미터기 요금 데이터를 각각 동형암호화하여 결합하고, 이를 기반으로 통계분석을 수행하여 동형암호를 활용함으로써 기존 절차보다 결합 과정이 간소화되고, 데이터 품질을 유지하면서 효율적인 분석이 가능함을 보여주고자 한다. 실증분석 결과, 동형암호 환경에서도 평균 데이터와 동일한 결과를 산출할 수 있음을 확인하였다. 이를 통해, 동형암호를 통한 이중 데이터 결합분석의 실용성과 향후 다양한 분야에서 민감한 데이터의 개인정보 유출 위험을 줄이면서도 데이터 활용을 극대화할 수 있는 가능성을 제시한다.

주요 용어 : 동형암호, 이중 데이터 결합, 결합 분석.

*이 논문은 정부(국토교통부)의 국토교통 데이터산업 생태계 플랫폼 구축(R&D) 사업의 재원으로 국토교통과학기술진흥원의 지원을 받아 수행된 연구임 (NO.RS-2022-0144012).

¹08791 서울특별시 관악구 관악로 98, 삼성빌딩 4층 (주)크립토허 데이터사업개발실 연구원.

E-mail: minho8173@cryptolab.co.kr

²08791 서울특별시 관악구 관악로 98, 삼성빌딩 4층 (주)크립토허 데이터사업개발실 연구원.

E-mail: hyohyeokjeong@cryptolab.co.kr

³02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 도시빅데이터융합학과 교수.

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

⁴(교신저자) 08791 서울특별시 관악구 관악로 98, 삼성빌딩 4층 (주)크립토허 데이터사업개발실 이사.

E-mail: scshin73@cryptolab.co.kr

Diversity in Integrated Interventional Radiology Residency Program^{*}

Myungji Kim¹, Jinseop Hyun², Eunjee Lee³, Samir S. Jambhekar⁴

Abstract

Diversity in healthcare plays a crucial role in reducing racial health disparities and enhancing patient satisfaction and outcomes. This study underscores the importance of diversity by examining its status within integrated interventional radiology (IR) residency programs and comparing it with other medical specialties. Using data from the American Association of Medical Colleges Reports on Residents from 2018 to 2023, we analyzed samples of women, Black, Hispanic, American Indian, Alaska Native, Native Hawaiian, and Pacific Islander residents. The analysis, which employed statistical methods including chi-square tests and between-group proportion tests, revealed that the proportion of women in integrated IR programs was relatively lower compared to other medical specialties (p-value<0.001). While the proportion of White individuals remained predominant in terms of racial diversity, there was an increase in Hispanic and Black individuals across most specialties(p-value<0.05). However, this upward trend was not observed within integrated IR programs. Additionally, the proportion of graduates from international medical schools decreased across most specialties(p-value<0.05). These findings underscore the imperative for efforts to foster awareness and inclusivity within the field of radiology.

Keywords : Interventional Radiology Residency Programs, Diversity.

^{*}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. NRF2022M3J6A1084843, No. NRF-2021R1C1C1C1013936)

¹Department of information and statistics, Chungnam National University, wlaud0605@o.cnu.ac.kr

²Department of information and statistics, Chungnam National University, guswlstjq0225@gmail.com

³Department of statistics and data science, Chungnam National University, eunjee.cnu@gmail.com

⁴Division of Vascular and Interventional Radiology, Department of Radiology, The Ohio State University Wexner Medical Center, Columbus, Ohio 43240, USA

인구고령화 시대에 중국 노인교육의 수요분석

등자운¹, 양정호²

요약

중국의 인구 고령화는 빠르게 진행되고 있으며, 이에 따른 노인 교육 수요가 급증하고 있다. 본 연구는 인구 고령화의 배경에서 중국 노인 교육의 현황과 수요를 분석하는 것을 목적으로 한다. 문헌 연구와 사례 연구를 통해 노인의 건강, 생활 기능, 문화 오락 등 다양한 요구를 분석하고, 경제적 여건, 건강 상태, 디지털 기능, 사회적 지원 등이 교육 수요에 미치는 영향을 파악하였다. 또한, 현재 중국 노인 교육의 공급 부족, 제한된 정책 지원, 낮은 사회적 인식 등의 문제를 제기하고, 이를 해결하기 위한 정책적 보완과 교육 자원의 풍부화, 사회적 지원 향상, 수업 형식의 다양화를 제안하였다. 이 연구는 중국 노인 교육 체계의 최적화를 위한 이론적 근거와 실천적 지침을 제공한다.

주요 용어: 인구고령화, 노인교육, 수요분석, 정책제언, 사회적 지원.

1. 서론

중국은 급속한 고령화의 과정에 있으며, 이에 따라 노인 교육의 필요성이 점차 중요해지고 있다. 2000년 이후 65세 이상의 노인 인구 비율이 급격히 증가하며, 2030년에는 노인 인구가 2억 5천만 명에 이를 것으로 예상된다. 이로 인해 경제, 의료뿐만 아니라 노인의 교육과 학습에 대한 수요도 급증하고 있다. 그러나 현재 노인 교육 시스템은 초기 단계에 머물러 있으며, 교육 자원과 정책 지원의 부족이 문제로 지적되고 있다.

2. 이론적 배경 및 선행 연구

노인 교육의 개념은 1990년대 프랑스에서 시작되었으며, 노인의 생리적 특성과 심리적 요구를 충족하기 위해 설계된 교육 체계이다. 노인 교육은 개인의 학습 수요를 만족시키는 수단일 뿐만 아니라 사회 통합과 평생 학습을 실현하는 중요한 방법이다. 특히 중국의 사회문화적 배경에서 노인 교육의 역할이 점점 더 강조되고 있으며, 이는 노인들에게 계속 학습하고 사회에 통합할 수 있는 플랫폼을 제공한다. 기존 연구들은 노인 교육의 현황과 문제점을 다루고 있으나, 구체적인 교육 수요에 대한 세밀한 분석은 부족한 상황이다.

¹03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 석사과정. E-mail: dengzy0426@163.com

²(교신저자) 03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 교수. E-mail: jyang@skku.edu

3. 연구 방법

본 연구는 문헌 연구와 사례 연구를 결합하여 수행되었다. 문헌 연구를 통해 노인 교육과 관련된 이론적 배경을 검토하였고, 사례 연구를 통해 중국 노인 교육의 현황과 주요 문제를 분석하였다. 특히, 노인의 건강, 디지털 기능, 문화 오락 등의 다양한 수요를 분석하여 교육 수요에 영향을 미치는 주요 요인을 도출하였다.

4. 연구 결과

노인 교육에 대한 수요는 매우 다양하며, 건강, 문화 활동, 디지털 기술 습득 등 다양한 분야를 포함한다. 그러나 현재 중국의 노인 교육은 이러한 다양한 수요를 충분히 충족시키지 못하고 있다. 교육 자원의 불균형, 정책 지원의 부족, 사회적 인식의 낮음 등이 주요 문제로 지적된다. 대부분의 노인들은 건강 유지와 삶의 질 향상을 위해 교육을 받고자 하나, 실제로 참여할 수 있는 교육 기회는 제한적이다.

5. 정책 제언

인구 고령화에 대응하기 위해, 본 연구는 다음과 같은 대책을 제안한다. 첫째, 노인 교육 정책을 보완하고, 각급 정부와 지역 사회의 참여를 통해 교육 자원을 풍부하게 해야 한다. 둘째, 노인의 다양한 수요를 반영한 다원화된 교육 과정을 제공하고, 온라인 학습 플랫폼을 구축하여 노인들이 쉽게 접근할 수 있도록 해야 한다. 셋째, 지역 사회의 지원을 강화하고, 가족과 사회가 연계하여 노인 교육에 대한 관심과 지원을 확대해야 한다. 마지막으로, 유연한 교육 방식과 맞춤형 학습 계획을 통해 노인의 학습 동기와 참여를 증진시켜야 한다.

6. 결론

본 연구는 중국의 인구 고령화가 노인 교육 수요를 크게 증가시키고 있음을 보여주며, 이를 해결하기 위한 정책적, 사회적 대책의 필요성을 강조한다. 노인의 다양한 학습 수요를 충족시키기 위해서는 정책, 자원, 사회적 지원의 포괄적인 개선이 필요하며, 이를 통해 노인의 전면적인 발전과 사회 참여를 촉진할 수 있을 것이다.

Reference

- Caprara M, Molina M Á, Schettini R, et al. Active Aging Promotion: Results from the Vital Aging Program. *Current Gerontology and Geriatrics Research*, 2013, 2013(1): 817813.
- Guo Aimei, Shi Ying. Delayed Retirement and Employment for Middle-aged and Older Workers: International Experience and China's Path. *Jianghai Academic Journal*, 2016(05): 98-104.
- Liu Xinyi. Exploration of High-Quality Development Paths for Elderly Education. *Journal of Guangxi Open University*, 2024, 35(03): 39-42.

- Mu Guangzong. The Key to Successful Aging: Balancing “Gains in Aging” with “Losses in Aging”. *Journal of Southwest University for Nationalities (Humanities and Social Sciences Edition)*, 2016, 37(11): 9-15.
- Research Group on Population Aging Strategies. Research on Population Aging Trends and Development Strategies. Beijing: *Huailin Publishing House*, 2014: 123-124.
- Schmidt-Hertha B, Formosa M, Fragoso A. Active Aging, Social Inclusion, and Wellbeing: *Benefits of Learning in Later Life*. 2019.
- Shao Juan. Research on Elderly Education Based on Lifelong Education System. *Journal of Jilin Radio and Television University*, 2022(06): 43-45.
- Yang Lina. Research on Legal Protection of Elderly Education Rights from the Perspective of Active Aging. Guizhou Normal University, 2024.
- Zhang Xinhua, Luo Shiguang, Luo Changri. Research on the Real Needs and Countermeasures of Elderly Education under the Background of Active Aging. *Journal of Qingyuan Vocational and Technical College*, 2023, 16(06): 84-90.
- Ministry of Education. Enrollment Guidelines for the Fall 2024 Semester of the Elderly University.
<https://jyb.syjh.lndx.edu.cn/app/eduorg/public/index/posts/2132?postsCateId=308>

Predictive modeling of maritime disasters: a comparison of neural networks, random forest, support vector machine, XGBoost, and LASSO*

*Xin Zan*¹, *Jai Woo Lee*²

Abstract

Shipping accidents are unexpected events, often resulting in severe loss of life and property. In this study, we use the Estonia maritime disaster dataset to compare the performance of predictive models such as neural networks, random forest, support vector machine, XGBoost, and LASSO, on survival probability. These models include various factors affecting survival probability to identify the optimal algorithm. The selected model is further validated using maritime disaster datasets. Predicting survival likelihood in maritime disasters provides evidence to support emergency rescue efforts and policymaking. The research findings not only suggest guides for data-driven decision-making but also devise methodologies for optimization of predictive models.

Keywords : Machine Learning, Optimization, Survival Analysis, Decision Making, Maritime Studies.

*This article is financially supported by the 2024 College of Public Policy at Korea University.

¹Department of Big Data Science, Korea University, Sejong

²(Corresponding Author) Department of Big Data Science, Korea University, Sejong.

Email: jaiwoolee@korea.ac.kr

머신러닝 방법론을 결합한 혼합효과 모형의 성능 비교 연구*

박정연¹, 이근범², 이상인³

요약

본 연구는 혼합효과 모형과 머신러닝 기법을 결합한 하이브리드 예측 모형을 제안하여, 군집화된 데이터에서 집단 간 및 개별 수준 변동성을 정교하게 반영하고자 하였다. 혼합효과 모형은 계층적 데이터의 구조적 특성을 설명하고, 집단 간 및 개별 관측치 간의 상호작용을 반영하는데 강점을 가지지만, 비선형적 패턴을 학습하는 데는 한계가 있다. 이를 보완하기 위해 혼합효과 모형에서 계산된 잔차를 머신러닝 기법의 입력 데이터로 활용하는 결합 방식을 제안하였다. 머신러닝 기법으로는 Lasso, Ridge, Elastic Net Regression과 같은 패널티 회귀 모형, Support Vector Machine(SVM), 그리고 비선형적 패턴 학습에 강점을 가진 XGBoost를 적용하였다. 혼합효과 모형은 데이터의 계층적 구조를 설명하며 고정 효과와 랜덤 효과를 추정하고, 머신러닝 기법은 혼합효과 모형에서 설명하지 못한 비선형적 패턴과 변수 간의 복잡한 상호작용을 학습하여 잔차를 보완한다. 최종적으로, 혼합효과 모형의 예측값과 머신러닝을 통해 학습된 잔차 예측값을 결합하여 계층적 구조와 비선형성을 동시에 반영할 수 있는 모형을 구축하였다. 제안된 모형의 유효성을 검증하기 위해 다양한 조건의 시뮬레이션 데이터와 실제 데이터를 활용하였다. 시뮬레이션 데이터는 데이터 크기와 복잡성의 변화를 반영하여 계층적 구조와 비선형적 패턴을 포함하도록 설계되었으며, 실제 데이터로는 대전시 공공자전거 시스템(타슈)의 대여 이력 데이터를 사용하였다. 평균 절대 오차(MAE), 평균 제곱근 오차(RMSE), 결정 계수(R^2)를 주요 성능 지표로 활용하여, 혼합효과와 머신러닝 기법의 결합 모형이 기존 혼합효과 모형 대비 우수한 성능을 보임을 확인하였다. 특히, XGBoost를 결합한 모형은 비선형적 관계를 효과적으로 학습하여 모든 평가 지표에서 가장 높은 예측 정확도를 기록하였다. 본 연구는 혼합효과 모형과 머신러닝 기법의 장점을 결합하여, 군집화된 데이터에서 발생하는 복잡한 비선형성과 계층적 구조를 정량적으로 반영할 수 있는 예측 프레임워크를 제시하였다.

주요 용어 : 혼합효과모형, 머신러닝, 군집화 데이터, 공공자전거, 비선형성.

*본 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 연구되었음(NRF 2022M3J6A1084843)

¹34134 대전광역시 유성구 대학로99, 충남대학교 통계데이터사이언스학과, 석사과정. E-mail: wjddus3325@naver.com

²34134 대전광역시 유성구 대학로99, 충남대학교 통계데이터사이언스학과, 석사과정. E-mail: veing9856@naver.com

³(교신저자)34134 대전광역시 유성구 대학로99, 충남대학교 정보통계학과 부교수.. E-mail: sanginlee44@gamil.com

Integrative Machine Learning Techniques for Modelling Complex Systems in Alzheimer's Disease

Juwon Kim¹, Sangjin Kim¹ and Jai Woo Lee²

Abstract

The incidence of Alzheimer's disease has been increasing over time while the effective therapies remain elusive. Machine learning approaches hold the potential to reduce the substantial time and cost involved in developing solutions for disease analysis. This study utilizes data from 2,150 Alzheimer's patients, incorporating a range of variables including demographic information, lifestyle factors, medical history, clinical, cognitive function assessments, Alzheimer's diagnosis status, and information from attending physicians. Assuming that the risk and incidence of Alzheimer's increase with age, various machine learning models with age-based groupings were applied. Random Forest model resulted in the best predictive performance in analyzing Alzheimer's patients. This model leverages patient information to predict Alzheimer's and examines the significance of various predictors affecting the disease status. By including the significant variables on a group-specific analysis, the ultimate aim of this study is to propose therapeutic strategies and interventions for Alzheimer's disease. This approach is expected to provide a novel contribution to Alzheimer's research, outperforming previous approaches in terms of predictive performance.

Keywords : Machine Learning, Big data, Stratified Analysis, Clinical Information, Disease Analysis.

*This article is financially supported by the 2024 College of Public Policy at Korea University.

¹Department of Big Data Science, Korea University, Sejong.

²(Corresponding Author) Department of Big Data Science, Korea University, Sejong,

Email: jaiwoolee@korea.ac.kr

AI-Driven approaches for Analysis of socio-Demographic and health-related factors affecting Individual depression

Hyeonjun Nam¹, Donggeun Kim¹ and Jai Woo Lee²

Abstract

The aim of this study is to analyze the effect of demographic features on individual depression and develop depression measurement indicators which may have a significant effect on individuals in society. Demographic features and individual depression experience were extracted from the survey data in 2023. In the data preprocessing step, variables related to education were curated after handling missing or irrelevant values in the analysis. Based on the pre-processed data, a depression prediction model using demographic features was developed using a boosting algorithm. It is expected that this study can contribute to understanding the effect of individual experience related to depression and suggesting effective measurement tools for mental health.

Keywords : Artificial Intelligence, Big Data, Pre-Processing, Demographic Features, Mental Health.

*This article is financially supported by the 2024College of Public Policy at Korea University.

¹Department of Big Data Science, Korea University, Sejong

²(Corresponding Author) Department of Big Data Science, Korea University, Sejong,

Email: jaiwoolee@korea.ac.kr

AI-Based Recommendation SYSTEM APPLIED in Consumer Behavior Analysis for a new automobile sales marketing model

Dongha Kim¹, Juyong Ko¹ and Jai Woo Lee²

Abstract

With the development of advanced artificial intelligence technology after the COVID-19 pandemic, we have entered the era of hyper-personalization. These social changes have highlighted the importance of personalization technology, and as a representative example, companies such as YOUTUBE and NETFLIX are greatly adopting personalized recommendation systems that provide personalized benefits based on individual preferences and circumstances as a core marketing strategy. The aim of this study is to develop a personalized consumer price prediction system and devise an artificial intelligence model that predicts an appropriate price by analyzing the effect of features related to the car sales in international automobile transaction data. It is expected that development of artificial intelligence technology for automobile sales marketing can contribute to rational decision making for both companies and consumers.

Keywords :A rtificial Intelligence, Recommendation Systems, Big Data, Marketing, Automobile Sales.

*This article is financially supported by the 2024College of Public Policy at Korea University.

¹Department of Big Data Science, Korea University, Sejong

²(Corresponding Author) Department of Big Data Science, Korea University, Sejong,

Email: jaiwoolee@korea.ac.kr

Prediction of future risk to suicidal ideation based on text data from social media with a machine learning approach

Kin Lok Wong¹ and Jai Woo Lee²

Abstract

This study attempts to prevent the suicidal risk by applying text analysis to filter the users based on socio-demographic features. We can suggest the social media to spread out the message and save people who may be in suicidal risk. The dataset includes 110 Instagram accounts of 102 adolescents activities created before 2022. The dataset has first been cleaned by eliminating a number of incomplete or irrelevant data points. In order to solidify the analytical process, conditions related to users, the personalization setting and data with product selling are considered to make accurate predictions. Various artificial intelligence techniques were applied to understand the associations of socio-demographic features related to social media users and estimate the suicidal risk in different conditions. It is expected that the development of prediction models based on text data can contribute to the detection of future risk to suicidal ideation in society.

Keywords : Machine Learning, Optimization, Survival Analysis, Decision Making, Maritime Studies.

*This article is financially supported by the 2024 College of Public Policy at Korea University.

¹Department of Big Data Science, Korea University, Sejong

²(Corresponding Author) Department of Big Data Science, Korea University, Sejong,

Email: jaiwoolee@korea.ac.kr

Classification of Tourist Movement Patterns Using Mobile Communication Data *

Seongjae Nam¹, Hongjae Lee², Boseung Chor³

Abstract

This study aims to analyze the travel patterns of domestic travelers using travel log data obtained from AI Hub, contributing to the development of a systematic travel route and destination recommendation system. The research data includes a total of 2,560 travelers visiting the Seoul metropolitan area, categorized into day-trippers, one-night two-day travelers, and travelers staying for more than two nights and three days. Among them, the analysis focuses on the travel coordinate data of 161 travelers who stayed for more than two nights and three days. The original dataset contains location information recorded at every moment. This data was preprocessed to extract travelers' coordinates at 30-minute intervals. Beyond analyzing simple travel routes, a clustering method incorporating temporal sequences was applied to identify not only the travel patterns but also the order of movements. This approach provides detailed insights by reflecting the sequence of visits, enabling more precise recommendations for routes and destinations. Furthermore, frequently visited areas in the metropolitan region, such as Gangnam-gu and Songpa-gu, were grouped into a single range, and detailed travel logs were analyzed for travelers who stayed in this area for more than two hours. This methodology is expected to offer a refined understanding of travelers' preferred routes and destinations, providing foundational data for the development of personalized travel recommendation systems.

Keywords : Travel Log Data, Movement Pattern Analysis, Time-Series Clustering, Travel Route Recommendation

*This research was supported by the MSIT(Ministry of Science, ICT), Korea, under the National Program for Excellence in SW), supervised by the IITP(Institute of Information & communications Technology Planing & Evaluation) in 2024 (grant number: R2416011)

¹Undergraduate student: Department of Economic Statistics with a Major in Big Data, Korea University, Sejong, Korea. E-mail:namseongjae8870@gmail.com

²Undergraduate student: Department of Economic Statistics with a Major in Big Data, Korea University, Sejong, Korea. E-mail:ghdwo7344@naver.com

³(Corresponding author) Division of Big Data Science, Korea University, Sejong, Korea.
E-mail:cbskust@korea.ac.kr

오차항이 자기회귀 과정을 따르는 다중 시계열 회귀 모형을 이용한 코로나19 확진자수와 인구이동량 연관에 대한 연구*

이선재¹, 최보승²

요 약

본 연구는 서울특별시 25개 자치구를 대상으로, 일별 인구이동량이 코로나19 일별 확진자 수와의 연관성을 분석하고, 이러한 연관성을 모형화 한 후 모형 구축 결과를 이용하여 지역을 그룹화하는 것을 목적으로 한다. 연관성을 모형화 하기 위하여 오차항이 자기회귀(AR; autoregressive) 과정을 따르는 다중회귀모형을 구축하였다. 인구이동량 변수는 접촉과 확진까지의 시차를 고려하여 4, 5, 6일 지연된 시차를 적용하였으며 이통 데이터 형태에 따라 여가 이동량과, 업무 이동량으로 구분하였다. 총 6개의 이동량 변수를 설명 변수로 구성되었다. 각 구별로 모형을 적합하여 추정된 회귀계수(β)를 도출한 결과, 여가 이동 유형이 확진자 수 증가에 더 강한 영향을 미치는 자치구와, 업무 이동량이 상대적으로 더 큰 영향을 미치는 자치구가 구분하였다. 도출된 회귀계수를 활용한 클러스터링 결과, 자치구 간 인구이동량과 확진자 수 간의 관계에서 유의미한 패턴 차이가 관찰되었다. 본 연구는 지역별 인구이동 특성을 반영한 감염병 확산 방지 대책 마련에 기초자료를 제공할 수 있다는 점에서 의의를 가진다.

주요 용어: 코로나19, 인구이동량, 오차 자기회귀 모형, 클러스터링, 회귀분석.

*본 연구는 범부처의 재원으로 범부처 감염병방역체계 고도화 R&D 사업 지원에 의하여 이루어진 것임 (HG23C1629).

¹고려대학교 빅데이터사이언스학부 학부과정. E-mail: sunjae0111@korea.ac.kr

²세종특별자치시 세종로 2511, 고려대학교 빅데이터사이언스학부, 교수. E-mail: cbskust@korea.ac.kr

JMP

www.jmp.com



수업용 무료 라이선스 (JMP Student Edition)

코딩 없이 통계를 배우고
가르칠 수 있습니다.



- JMP는 SAS에 의해 개발된 데스크탑용 분석 소프트웨어로서 전 세계 수천 개의 기업과 교육 현장에서 사용되고 있습니다.
- JMP Student Edition은 학생, 교육자 및 학술 연구자에게 연간 단위로 무료로 제공되는 라이선스입니다.
- JMP Student Edition(JMP Pro 기반)은 데이터 시각화에서 실험 설계, 예측 모델링, 머신 러닝 및 구조 방정식뿐만 아니라 Python 통합 등 광범위한 통계적 과제를 해결할 수 있습니다.



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 포탈 사이트 회원이 아닙니다

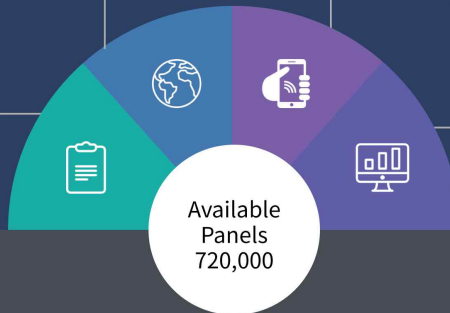
조사참여 횟수, 조사 참여일, 참여조사 종류 및 응답의 성실도를 체계적으로 관리합니다.

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



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"> · 수요 예측 · 리스크 예측 · 상품 추천 · 품질 분석 및 예지 보전 · 이미지 분석 · 텍스트 분석 · 고객 분석 	<ul style="list-style-type: none"> · MAaaS PASP · PoC · Pro · Maintenance · MAaaS by AI Model 	<ul style="list-style-type: none"> · KoreaPlus Statistics (Embaded on SPSS) · KNIME · IBM CP4D · Market Mind · Brightics AI · AI.NER 	<ul style="list-style-type: none"> · 빅데이터리닝센터 · SPSS/Python/R 교육 연간 80여개 온/오프라인 강의제공

데이터솔루션 분석 사례

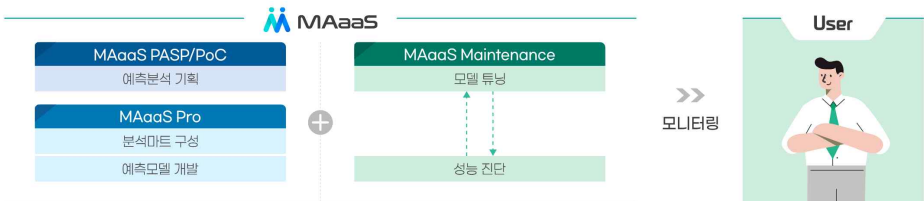
수요 예측 제조 수요예측 분석 시스템 구축 유통 판매량 예측 모델 개발 물류 물동량 예측 시스템 구축 공공 열 수요 예측 모델 개발	텍스트 분석 통신 VoC 분석 금융 텍스트 분석을 활용한 평판 분석 유통 텍스트 분석을 통한 마케팅 증권 챗봇 메시지 분석
상품 추천 유통 AI 상품 추천 프로그램 구축 유통 온라인 쇼핑을 개인화 상품 추천 모델 개발 유통 개인화 추천 서비스 모델 구축 금융 고객관리 시스템을 통한 상품 추천	품질 분석 및 예지보전 제조 취약 부위 예측 모델 개발 제조 특정 공정이상 예측 분석 제조 공정 최적화 및 설비 고장 모델 개발 화학 수율 예측 시뮬레이터 개발
이미지 분석 제조 이미지 데이터 기반의 양불 판정 제조 비전 검사를 통한 양불 판정	

AI 예측서비스, 이제 구독하세요!

No 플랫폼 No 데이터수입 No 데이터/가이던트스트 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>회사명 ㈜지디에스컨설팅그룹</p> <p>설립일 2015년 2월</p> <p>대표이사 김은석</p> <p>업종 데이터베이스, 소프트웨어 자문, 개발 및 공급, 전자상거래업</p> <p>주소 서울 성동구 광나루로 228, 9층 (성수동2가, 렉스모터스)</p> <p>연락처 Tel: 02-2135-8895(대표) Fax: 02-2135-8896</p>	<p>2018.07 고려대 '데이터 사이언스 센터' 업무 제휴 (교육 및 인턴십)</p> <p>2017.07 KB국민카드 '빅데이터 센터' 업무 제휴 (카드 데이터 상품개발)</p> <p>2016.09 KT '빅데이터 지원단' 업무 제휴 (빅데이터 가공, 중계, 서비스 분야)</p> <p>2016.08 데이터 거래 중개 사업 개시 (미래창조과학부 ICT 기금사업) 부설연구소 '한국공공정보개발원' 설립 (산기협 2016113799호)</p> <p>2016.07 빅데이터 분석 사업 개시 (지속가능한 빅데이터 사업, 경기도)</p> <p>2015.04 국가 DB 구축 사업 개시 (유동인구 실사 DB 구축, NIA)</p> <p>2015.02 ㈜지디에스컨설팅그룹 설립 (㈜지디에스케이 內 컨설팅 사업본부의 법인 분할)</p> <p>2004.10 벤처기업 등록 (041134221-1-01033호)</p> <p>2003.02 ㈜지디에스케이 설립 (대표이사 김은석 차)</p>
---	---

Business : 사업 영역

데이터 활용 및 유통의 일관된 서비스 제공!!

사업영역	빅데이터 분석 컨설팅	데이터 가공	빅데이터 플랫폼 구축	데이터 판매/유통
사업영역		자세히 보기 >	자세히 보기 >	자세히 보기 >
상품 및 서비스	<ul style="list-style-type: none"> · 데이터관련 ISP 및 계획수립 · 주요 빅데이터 (통신, 카드) 공급 및 활용 · 공공데이터 활용 서비스 · 데이터 분석 및 시각화 서비스 	<ul style="list-style-type: none"> · 공공 개방 정보의 정제, 가공 및 제공 · 통신, 카드 등 빅데이터 정제, 결합 및 제공 · 비정형 데이터 가공 및 제공 	<ul style="list-style-type: none"> · 통계분석 도구 (압트릭스) 판매 · 데이터 용복합 및 통합 시스템 구축 · 빅데이터 분석 시스템 구축 	<ul style="list-style-type: none"> · 데이터 사용영역별 데이터 판매 · 업무 영역별 맞춤형 데이터 구축 및 제공 · 주제별 맞춤형 데이터 제공
사업사례	<ul style="list-style-type: none"> · 국방기술제품개발, 빅데이터 시범사업 (17-) · 경기도, 지속가능한 빅데이터 분석 (16) · 서울시, GDP 분석 모델링 구축 (16) · 건강보험공단 빅데이터-노인자원 모델링 구축(14) · 서울시, 유동인구 DB 구축 및 모델링 (15) 	<ul style="list-style-type: none"> · 삼성전자, 법인영업 데이터 제공 (15-) · 한샘, 가구판매 고객 데이터 제공 (15-) · 코웨이, 고객정보 관리 데이터 제공 (06-) · KT, 지사 관리 데이터 제공 (10-) · 기업은행, 영업점 관리 데이터 제공 (10-) 	<ul style="list-style-type: none"> · 창원시, 빅데이터 분석시스템 구축 (17) · KT, AI 조류독감 시각화 시스템 구축 (16) · 경상북도, 통계정보시스템 구축 (15) · KT, 지사 관리 데이터 제공 (10-현재) · 서울시, 도시통계 시스템 구축 (14, 15) 	<ul style="list-style-type: none"> · 한국정보화진흥원, 데이터 거래 중개 플랫폼 구축 (16) · 한국정보화진흥원, 소상공인 데이터 제공 플랫폼 (17) · 경상북도, 사학사 연구자를 위한 데이터 (16) · 코웨이, 고객정보관리를 위한 데이터 (16)외다수