

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

문서번호 : KDAS-2511000

시행일자 : 2025. 11. 27.

(경유)

수 신 : 전 회원

참 조 :

선결			지	
	일자		시	
접 수	시간		결 재	
	번호			
담당자				

제목: 한국자료분석학회 2025년 동계 학술논문발표대회

- 귀 회원님의 무궁한 발전을 기원하며, 본 학회의 발전을 위한 회원 여러분의 관심과 노고에 깊이 감사를 드립니다.
- 본 학회 2025년도 동계 학술논문발표대회를 다음과 같이 개최함을 알려 드리오니 많은 참여 바랍니다.
 - 개최일자 : 2025년 11월 27일(목) ~ 28일(금)
 - 장 소 : 충북대학교(충북 청주시 서원구 충대로 1)

한국자료분석학회장 한상태



PROCEEDINGS OF THE KOREAN
DATA ANALYSIS
SOCIETY

November 27-28, 2025

at

Department of Information Statistics,
Chungbuk National University, Korea.

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

THE KOREAN DATA ANALYSIS SOCIETY Founded 1998

한국자료분석학회 2025년도 통계 학술논문발표대회 준비위원

조직위원회

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

운영위원회

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

한국자료분석학회

2025년도 동계 학술논문발표대회

- 일시 : 2025년 11월 27일(목) ~ 11월 28일(금)
- 장소 : 충북대학교(충북 청주시 서원구 충대로 1)
- 공동주최 : 한국자료분석학회, 충북대학교 정보통계학과
- 주관 : 충북대학교 정보통계학과
- 후원 : 한국연구재단, (주)케이스탯리서치, 트랜드리서치, (주)GDS 컨설팅, 데이터솔루션, 나이스평가정보, 글로벌리서치, 베가스, 마케팅크, 충북대학교, DS플러스 사업단, 충북대학교 디지털바이오 인력양성사업단, 충북대학교 GLAMP 사업단

● 11월 27일 목요일

17:00 - 19:00 편집위원회 및 이사회

● 11월 28일 금요일

09:00 - 18:00 등록(자연과학대학 5호관 S1-5동, 123호 앞)
포스터발표(9:00 - 16:30, 자연과학대학 4호관 S1-4동, 1층)

09:30 - 10:50 논문발표 A1 - A4(자연과학대학 5호관 S1-5동 114호, 118호, 123호, 127호)

11:10 - 12:10 튜토리얼(자연과학대학 4호관 S1-4동, 1층 사이언스홀)
좌장: 신승준 교수(학술이사, 고려대학교)

제 목: 시계열 분석을 위한 인공지능 기반 방법론
강연자: 박태영 교수(연세대학교 응용통계학과)

12:10 - 12:50 총회(자연과학대학 4호관 S1-4동, 1층 사이언스홀)
진행: 이성건 교수(총무이사, 성신여자대학교)

12:50 - 14:00 증식시간 (충북대 제1학생회관 2층 별빛식당)

14:00 - 15:00 학회장 초청 강연(자연과학대학 4호관 S1-4동, 1층 사이언스홀)
좌장: 한상태 교수(학회장, 호서대학교)

제 목: 우리들의 학문, 그 너머의 무엇 (What I learned from what I taught and studied)
강연자: 허명희 명예교수(고려대학교)

15:10 - 16:30 논문발표 B1 - B4(자연과학대학 5호관 S1-5동 114호, 118호, 123호, 127호)

16:40 - 18:00 논문발표 C1 - C4(자연과학대학 5호관 S1-5동 114호, 118호, 123호, 127호)

논문발표 A: 09:30 - 10:50

논문발표 A1(114호)

보험/금융 통계

좌장 : 송 준(고려대학교)

09:30	1. Calculation of Enterprise Capital via Least-squares Monte Carlo:
	Regress Now or Later?
10:50	1 하홍준*(고려대학교 통계학과), Daniel Bauer(University of Wisconsin-Madison)
	2. Integration of Traditional and Telematics Data for
	Efficient Insurance Claims Prediction
	3 Hashan Peiris, Himchan Jeong*(Simon Fraser University), Jae-kwang Kim (Iowa State University), Hangsuck Lee (Sungkyunkwan University)
	3. Counting Data의 보험료 순서 모델링을 위한 코플라 상태모형
	5 안재윤*(이화여자대학교 통계학과)

논문발표 A2(118호)

통계 방법론

좌장 : 최호식(서울시립대학교)

09:30	1. Learning Robust Tabular Models under Unseen Missingness Shifts
	7 이지혜, 강민서, 김동하*(성신여자대학교 통계학과)
10:50	2. L0-regularized complex censored rank regression and its application
	9 선수연(성신여자대학교 수리통계데이터사이언스학부), Dipankar Bandyopadhyay(Virginia Commonwealth University), 최태화*(성신여자대학교 수리통계데이터사이언스학부)
	3. Self-Supervised Representation Learning for Tabular Data via Binning
	Tokenization
	11 박범진*(경상국립대학교 정보통계학과), 최호식(서울시립대학교 도시빅데이터융합학과)
	4. Quantile and rank estimation for encrypted data
	13 어수행, 김재선(크린토타입), 최호식*(서울시립대학교 도시빅데이터융합학과)

논문발표 A3(123호)

통계(응용)

좌장 : 이동희(경기대학교)

09:30	1. Bayesian Online Change-Point Detection	15
10:50	양대원*(충남대학교 정보통계학과), 최태련(고려대학교 통계학과)	
	2. 순서형 반응변수에 대한 비모형적 회귀 연관성 탐색: 소방공무원 PTSD 사례 ...	17
	전세봄*(목원대학교 마케팅빅데이터학과), 김대영(메사추세츠대학 수학과통계학과)	
	3. Moment Condition Models with Conditional Dirichlet Process Priors	19
	조성일*, 이재귀, 이경재, 이광민, 이재용(인하대학교 통계학과)	
	4. Hybrid Bayesian network models using mixed data in panel survey	21
	이동희*(경기대학교 경영학부)	

논문발표 A4(127호)

통계 (응용) - 학생세션

좌장 : 이은지(충남대학교)

09:30	1. Cumulative and Lagged Nonlinear Associations Between Particulate Matter	23
10:50	Exposure and Lung Function	
	이은지*, 조나래, 허태영(충북대학교 정보통계학과),	
	김영열, 권은진, 강민구(질병관리청 국립보건연구원),	
	정재환(충북대학교 정보통계학과)	
	2. Analyzing spatial effects using GMM estimators for binary dependent variable :	
	A case study of the the National Consortium on Violence Research data.	25
	이다호*, 김희영(고려대학교 빅데이터사이언스)	
	3. Big Data Analysis for AI-Generated Social Media Threads	27
	Kin Lok Wong*, 이재우(고려대학교 빅데이터사이언스학과)	
	4. 감소성 고장 검출률 함수를 활용한 소프트웨어 신뢰성 모형에 관한 연구	29
	공영선*(조선대학교 전산통계학과),	
	장인홍(조선대학교 컴퓨터통계학과), 김윤수(조선대학교 G-LAMP사업단),	
	송광윤(조선대학교 컴퓨터통계학과, G-LAMP사업단)	

논문발표 B: 15:10 - 16:30

논문발표 B1(114호)

기계학습

좌장 : 박관영(성신여자대학교)

-
- 15:10
|
16:30
1. Adaptive RKHS for Nonlinear Modeling with Interactions 31
장세훈, 송준*(고려대학교 통계학과)
 2. Valid Asymptotic Inference after Sufficient Dimension Reduction in
A Single-Index Framework 33
김경원*(연세대학교 응용통계학과), 송준(고려대학교 통계학과),
유재근(이화여자대학교 통계학과)
 3. Spatial covariance estimation in massive raster data 35
신준호*(성신여자대학교 수리통계데이터사이언스학부),
Ying Wu(Department of Applied Mathematics National Dong Hwa University),
임채영(서울대학교 통계학과)
 4. Spheresmooth: an R package for penalized piecewise geodesic curve
fitting on a sphere 37
정재환(충북대학교 정보통계학과), 이세영(성신여자대학교 통계학과),
구자용(고려대학교 통계학과), 박관영*(성신여자대학교 통계학과)

논문발표 B2(118호)

통계 분석 방법론

좌장 : 이재우(고려대학교)

-
- 15:10
|
16:30
1. 회전성을 갖는 제2종 중심합성계획의 직교블로킹 39
김혁주*(원광대학교 빅데이터금융통계학부)
 2. Deep Learning-Based Audio Deepfake Detection System for
a Competitive Challenge 43
한소율*(한남대학교 빅데이터응용학과)
 3. Multimodal AI 및 GAN을 활용한 국가승인통계의
격자단위 기반 데이터 증폭 45
이동찬*(고려대학교 경제통계학과),
김기환(고려대학교 빅데이터사이언스학부)
 4. Comprehensive Overview of Real-World Artificial Intelligence Applications 47
이재우*(고려대학교 빅데이터사이언스학과)

논문발표 B3(123호)

통계 (AI) - 학생세션

좌장 : 광일엽(중앙대학교)

15:10	1. Deep Learning for Censored Quantile Regression with An Adapted Loss Function	49
16:30		
	2. Multimodal Deep Learning for Glaucoma Diagnosis using Fundus and Vascular Features	51
		이도은*, 최상범(고려대학교 통계학과)
	3. ResNet-BiGRU with Conditioned Query-Based Cross-Attention and Weighted Loss for Automated Chagas Disease Detection from 12-Lead ECG	53
		강태영*, 임현오, 이나현, 김태환(중앙대학교), 김동건, 이동규, 오승상(고려대학교), Wuming Gong(University of Minnesota), 광일엽(중앙대학교)
	4. Car Accident Detection using CCTV data	55
		김현주*, 광일엽(중앙대학교)
	5. Transformer 기반 모델을 이용한 낚시성 기사 탐지 및 대안 제목 생성 연구	57
		문수연*, 한상태(호서대학교)

논문발표 B4(127호)

통계방법론 - 학생세션

좌장 : 강현철(호서대학교)

15:10	1. Hierarchical Extension of Latent Space Item Response Model	59
16:30		
		Soyeon Park*, Seoyoung Shin(Yonsei University), Minjeong Jeon(School of Education and Information Studies), Hyoun Kyoung Kim, Ick Hoon Jin(Yonsei University)
	2. Interaction-Only Q-learning: A Simplified and Robust Approach to Optimal Dynamic Treatment Regimes	61
		송수지*(고려대학교 통계학과)
	3. Bootstrap based score tests for overdispersion in a bivariate zero-inflated negative binomial regression model	63
		신지은, 장동민*, 정병철(서울시립대학교)
	4. Prediction of Scale Score based on Item Response Theory in case of Two Heterogeneous Groups	65
		김재은*, 김경희(고려대학교 통계학과)
	5. Hybrid Distance-Based Clustering and Dirichlet Process Mixture Modeling for Mass Spectral Analysis	67
		손민수*, 김재오(인하대학교 통계데이터사이언스학과)

논문발표 C: 16:40 - 18:00

논문발표 C1(114호)

보건/의생명

좌장 : 정재환(충북대학교)

16:40	1. Age-grouped analysis of influenza transmission influence to predict influenza spread based on mathematical model	69
18:00		
	김정현, 김정석, 김진솔, 안시현, 김민수*, 박찬근(국립한국해양대학교 전자전기정보공학부)	
	2. Graph Neural Network 기반 2차원 Contact Map 학습을 이용한 단백질 - 리간드 결합 및 상호작용 예측	71
	안윤성, 강일석(충북대학교 정보통계학과), 조효제*(충북대학교 생화학과)	
	3. Big data-driven genetic and epigenetic biomarkers in aging	73
	송난*(충북대학교 약학과)	
	4. 병원 의료데이터를 활용한 심뇌혈관질환 진단 AI 기술	75
	박승*(충북대학교 의과대학 의학과)	

논문발표 C2(118호)

기계학습

좌장 : 김성환(건국대학교)

16:40	1. Scaling Up ROC-Optimizing Support Vector Machines	77
18:00		
	배기문*, 신승준(고려대학교 통계학과)	
	2. A Mobile LiDAR-Based Deep Learning Approach for Real-Time 3D Body Measurement	83
	정용호*, 노태욱, 이용학, 이선재, 최광일, 정수진, 김성환(건국대학교 응용통계학과)	
	3. Toward Full Automation of the Revised NIOSH Lifting Equation "Using Computer Vision"	85
	정세희*, Xu Xu(North Carolina State University)	
	4. Cloud2CAD: AI-Driven High-Fidelity CAD Reconstruction of Industrial Components from 3D Point Clouds	87
	강준우(광운대학교 컴퓨터공학과), 이채환(건국대학교 데이터사이언스학과), 이용학, 김성환, 양성민*(건국대학교 응용통계학과)	

논문발표 C3(123호)

기계학습 - 학생세션

좌장 : 신승준(고려대학교)

16:40	1. Distributed Reduced-Rank Regression for Large-Scale Data	89
18:00	최홍신*, 박세영(연세대학교 통계데이터사이언스학과)	
	2. Adaptive Fused Quantile Regression with BAR Penalty	91
	은지현*(고려대학교 통계학과)	
	3. Memorize Early, Then Query:	
	Inlier-Memorization-Guided Active Outlier Detection	93
	강민서*(성신여자대학교 통계학과),	
	박승환(강원대학교 정보통계학부), 김동하(성신여자대학교 통계학과)	
	4. Sufficient KCI: Kernel-based Conditional Independence Test with	
	Sufficient Dimension Reduction	95
	류채현*(이화여자대학교 통계학과), 김경원(연세대학교)	
	5. Analyzing Classroom Interaction Data Using Prompt Engineering and	
	Network Analysis	97
	김광희*(연세대학교 통계데이터사이언스학과),	
	전민정(University of California), 진익훈(연세대학교)	

논문발표 C4(127호)

DS플러스 사업단 - 학생세션

좌장 : 강일석(충북대학교)

16:40	1. 위성 데이터 기반 글로벌 산불 확산 예측 딥러닝 모델의 개발	99
18:00	송영승*, 신성훈, 김도윤, 박혜민, 허태영(충북대학교 정보통계학과)	
	2. 대전 행사 축제 운영 효율성 분석 및 성과 개선을 위한 정책 제언:	
	DEA-Tobit 2단계 접근법을 중심으로	101
	김승연*, 김서연, 문유정, 홍지윤(충남대학교 정보통계학과)	
	3. Optimal Location Selection for Community Sports Facilities in	
	Cheongju using GTF and MCLP	103
	이동영*, 이주성(충북대학교 정보통계학과)	
	4. 무인점포 안전관리를 위한 Skeleton - Context 기반	
	실시간 이상행동 탐지 모델	105
	임지원*, 정윤채(충남대학교 통계데이터사이언스학과)	
	5. 대전 관광의 편중 완화를 위한 데이터 기반 추천 시스템	107
	박진규*(충남대학교 정보통계학과), 이소민(충남대학교 통계데이터사이언스학과)	

1. CNN과 LSTM 결합을 통한 작물 생산량 예측 정확도 향상 연구 109
강소라*, 나명환(전남대학교 수학과/통계학과)
2. AutoEncoder 사전 학습 임베딩을 활용한
그래프 신경망 기반 다변량 시계열 이상 탐지 111
고경민*(고려대학교 빅데이터융합학과)
3. Deep Learning-based Analysis of Cocaine-induced Behavioral Patterns in Marmosets 115
구선미*, 이철규, 최병모(충북대학교 약학과), 유대운, 엄희중,
이도현(오송첨단의료산업진흥재단 비임상지원센터), 윤재석(충북대학교 약학과)
4. 서울시 동작구 재정비촉진구역 주택재개발사업의 단계별 지연 요인에 관한 연구 117
김남재*(인하대학교 도시계획학과), 변병설(인하대학교 행정학과)
5. Physiological Curve-Inspired Feature Engineering for Fructose Prediction in Niitaka 119
김선호*(전남대학교 데이터사이언스학과), 김민수(전남대학교 통계학과)
6. The Effect of Social Capital on Economic Performance 121
김순호*(부경대학교 경영학부)
7. 경기도 학교급식용 친환경 배·당근 소득 및 경영비 분석 123
김연진*, 이원석, 이정명, 이수연(경기도농업기술원)
8. B2C 캠페인 Agent 오케스트레이션 : Function Calling-RAG 결합 성능 분석과
강화학습 기반 최적화 125
김영교*(고려대학교 SW·AI융합대학원 인공지능융합학과)
9. 도시열섬지수(UHI)에 관한 연구: 군집분석 · 회귀분석 중심으로 127
김지현*, 서은영(조선대학교 전산통계학과), 송광윤(조선대학교 컴퓨터통계학과)
10. NASA 연구개발 펀딩의 기관 생태계 변화 분석 129
노경란*(한국과학기술정보연구원)
11. 브라우징 관여가 과업 복잡성 및 의사결정 의존성 간의 관계에 미치는 매개 영향에
대하여: 온라인 쇼핑에서의 베이지안 매개 분석 131
노민정*(부경대학교 경영학부)
12. 남자 대학생의 인유두종바이러스(HPV)에 대한 지식, 백신접종에 대한
태도 및 수용도 133
마준용*, 하운후, 주환성, 최현서, 남궁현우, 김건우, 진은미(배재대학교 간호학과)
13. Association Between Physical Activity Timing and Metabolic Syndrome in Korea:
A Functional Principal Component Approach 135
박수아*(충북대학교 통계학과), 지희정(충북대학교 정보통계학과)

14. 딥러닝 기반 시계열 예측 경량화 모델의 비교 분석	137
박현준*(광운대학교 인공지능응용학과), 김성용(호서대학교 빅데이터AI학부), 임동혁(광운대학교 정보융합부)	
15. 1인 가구의 디지털 이해력이 웰에이징에 미치는 영향	139
성순임*, 김종숙(건양대학교 웰다잉융합연구소), 김광환(건양대학교 병원경영학과)	
16. 사용자 입력 데이터의 정형화 및 피드백 자동화를 위한 챗봇 설계	141
신준호*, 장원혁, 정상훈, 김동재(부경대학교), 김동현(주식회사 뉴아이), 문형빈(부경대학교)	
17. The Taylor rule and house price volatility	143
심성훈*(부경대학교 경제학과)	
18. Distributional Effects of the Cultural Voucher Program	147
오주현*(가천대학교 경영학과)	
19. 자기효능감이 성공적노화에 미치는 영향: 인지적 유연성과 심리적 탄력성의 매개효과	149
왕흔이*, 한현석(전북대학교 심리학과)	
20. 층화 직접질문 조건부 양적속성 무관질문모형	151
이기성*(우석대학교), 홍기학(동신대학교), 손창균(동국대학교), 박근화(한국문화관광연구원), 홍성준(동국대학교)	
21. Adaptive Weighted Total Variation Penalty for Precise Change Point Detection	155
이동영*(충북대학교 통계학과), 박관영(성신여자대학교 통계학과), 정재환(충북대학교 통계학과)	
22. 가속도계 기반 수면시간 추정을 위한 결측치 대체 방법 비교	157
이민영*(충북대학교 통계학과), 윤나영, 최형윤, 지희정(충북대학교 정보통계학과)	
23. 베이저안 신경망 기반 독성 예측에 대한 불확실성 정량화	159
이석진*, 김정연(서울시립대학교 인공지능학과)	
24. OECD 국가의 성평등 태도 비교: 세계가치조사(WVS) 자료 기반 분석	161
이아름*(창원대학교 의류학과)	
25. 느타리 수확후배지의 재활용 가치 평가: 축사갈래 활용을 중심으로	163
이정명*, 이원석, 김연진, 이수연(경기도농업기술원)	
26. 청소년기 사회적 관계 다중궤적 분석: Group-based multi-trajectory model의 적용	165
장은아*, 정혜원(충남대학교 교육학과)	
27. 고혈압이 있는 중년여성의 주관적 건강상태에 미치는 영향요인	167
전은미*(배재대학교 간호학과)	

28. Wavelets Scattering Meet Transformers: Detecting Anomalies in Robotic Trajectories 169
정윤채*, 임지원, 류하은, 박민수(충남대학교)
29. Sequence Kernel Association Test (SKAT) for
Genetic Variant Identification Based on MMSE Scores 171
정진경*, 이은지(충남대학교 정보통계학과)
30. UN 지속가능발전목표(SDG) 달성을 위한 국가연구개발사업의
데이터 기반 측정 프레임워크 173
최봉기*, 노경란(한국과학기술정보연구원 글로벌R&D분석센터)
31. 국가거점국립대 신입생 중도탈락 예측 모형 개발 및 활용 방안 175
최원기*, 김민수(전남대학교 통계학과)
32. 다중 코딩 에이전트 협업 스펙트럼 분석 : 논문 필터링을 위한
다중 코딩 에이전트 구성 비교 연구 177
한혁*(한국과학기술정보연구원)
33. AOP-Wiki 그래프 데이터 기반 Key Event 경로 자동 탐색을 위한 데이터의
경로 구조 분석과 확장 가능성 탐색 179
황신정*, 강민설, 최호식(서울시립대학교 도시빅데이터융합학과)
34. Profiling the User of Real Estate Tax Tech Market 181
Dongjae Kim*, Yunhui Kim, Goeun Kang, Junho Shin(부경대학교),
Donghyeon Kim(NEWEYE Corp. Co. Ltd.), HyungBin Moon(부경대학교)
35. Streamlit-based automated analysis platform for real-time calcium imaging in
freely moving mice
Hyeokjun Kwon*, Jaesuk Yun(College of Pharmacy, Chungbuk National University)
36. Music Therapy and Gardening Enhances Post-Discharge Adherence in Psychiatric 185
Wonsik Jung, Insu Jan, Hyeonjeong Lim*(충남대학교 자연과학연구소),
Subin Jeong, Hyun Ju Lee, Gahui Han, Shin-Young Park, Hoyeong Jeong,
Yaehyeon Kim, Seo Eun Lee, Eunjee Lee, Kwang-Yeon Choi
37. Predictive Modeling of Maritime Disasters Using Automated Machine Learning 187
Xin Zan*, Jai Woo Lee(고려대학교 빅데이터사이언스학과)
38. 멜론 스마트팜 자립경영을 위한 목표소득 달성 시나리오 분석 189
김효중*, 신길호(전라남도농업기술원), 이승현, 하지영((주)지앤비)
39. 디노이징 기반 인텐트 집계 통합 모델을 이용한 실시간 사용자 의도 추론 191
선피터빈*(아주대학교 인공지능학과), 최영준(아주대학교 소프트웨어학과)
40. 일사누적속도에 따른 급액제어를 통한 완전 순환식 양액공급 시스템 개발 193
김효중, 신길호*, 홍우진, 김주혜, 김동관(전라남도농업기술원)

튜토리얼

(1) 발표자 : 박태영 교수 (연세대학교 응용통계학과)

(2) 제목 : 시계열 분석을 위한 인공지능 기반 방법론

시계열 분석은 추세와 계절성 등 신호 구조를 파악하고, 데이터를 생성하는 확률적 메커니즘을 이해하며, 필요에 따라 미래 값을 예측하거나 외부 요인의 영향을 평가하는 다양한 목적을 바탕으로 발전해왔다. 최근에는 건강검진 기록, 웨어러블 생체신호, 커넥티드 카 운행 데이터 등 여러 변수와 개체가 비동기적으로 관측되는 불규칙 다변량 시계열이 증가하면서, 예측뿐 아니라 결측 구간 보간, 상태 분류, 이상 징후 탐지, 요약, 그리고 제한된 관측에서의 신호 복원까지 다양한 분석 목적을 수행하는 것이 중요해지고 있다. 이러한 데이터 특성을 다루기 위해 전통적인 RNN(Recurrent Neural Network)과 TCN(Temporal Convolutional Network) 기반 접근에서 Transformer, Graph Neural Network, 그리고 연속 시간 상의 변화를 직접 모델링하는 Neural ODE(Neural Ordinary Differential Equation) 기반 방법론으로 연구가 확장되고 있으며, 고차원 시계열에서 유의미한 표현을 학습하는 방식이 핵심적 역할을 한다. 본 튜토리얼에서는 이러한 목적의 확장과 방법론적 전개를 살펴보고, 불규칙 다변량 시계열의 실제 응용 사례를 통해 향후 연구 방향을 논의한다.

초청강연

(1) 발표자 : 허명희 명예교수(고려대학교)

(2) 제목 : 우리들의 학문, 그 너머의 무엇 (What I learned from what I taught and studied)

우리가 가르치고 연구하는 통계학 또는 데이터과학은 당연히 과학적이어야 하므로 인문적 해석을 붙이지 않는 것이 학문적 규범이다. 그럼에도 불구하고 우리들의 학문을 한 층 위에서 내려다 보면 인간과 삶에 대한 통찰이 드러난다. 이 발표는 이제는 학문적 규범에서 자유로워진 은퇴자가 우리들 학문에 대한 철학을 14개 항목으로 나누어 발제한다.

1. 비편향성(unbiasedness): $E(T; \theta) = \eta(\theta)$ 일 때 통계량 T 가 $\eta(\theta)$ 를 비편향적으로 추정한다고 한다. 좌우의 치우침 없이 균형감을 갖고 판단한다는 의미인 ‘비편향성’은 통계학도로서 제1의 궁극적 목표이다.
2. 최소분산(minimum variance): 분산은 산포(散布), 즉 분포의 퍼짐(spread)을 나타낸다. 산포는 내재적 잡음, 불안정, 변동스러움이다. 이것을 없앨 수는 없다. 그러나 이것이 최소로 통제된 일상을 지향한다.
3. 로버스트(robust)하다: 흔히 ‘강건하다’로 번역되지만 ‘튼실하다’이 더 낫다. 비바람을 견디어내는 느티나무가 튼실함의 표상이다. 위기에 버티는 바닥 힘이 진짜 실력이다.
4. 도박문제(gambler's ruin): 초기자산 k 에서 시작한 gambler가 매번 1을 걸고 성공 확률 p 인 플레이를 거듭하는 과정을 생각하자. p 가 0.5인 공정 게임에서도 목표액 n 이 사전에 설정되지 않으면 gambler가 파산할 확률은 1이다. 과욕 금지!
- 4.1 켈리 법칙(Kelly's rule): 매번, 자산의 일정 비율 $r < 1$ 을 걸고 성공 확률 p 인 플레이를 거듭하는 과정을 생각한다. $p \geq 0.5$ 인 경우, r 의 최적값 r^* 는 $p - q$ 이다. 프로야구에서 상위 팀의 정규시즌 승률은 0.55 근처이다. 그때의 r^* 는 $0.55 - 0.45 = 0.1$ 이다. 인생은 수십 번의 플레이로 구성된다. 과감과 소심 사이에서 베팅을 해야 한다.
5. EDA(exploratory data analysis)의 가치: 데이터에 대한 탐색(exploration)은 빈둥거리는 것처럼 보일 수 있는 비(非) 목표지향적 작업이다. 탐색적 과정에서 바른 목표가 설정되고 명료해진다. 당장의 비효율이 장기적으로 가치가 있을 수 있다. [8:2 법칙]
6. 통계적 유의성(statistical significance): 충분한 수의 사례에서 확인된 것만을 따라야 한다. 인간은 one-shot learning을 하는, 빨리 알아채는 존재이다. 그러나 오판도 잦다. 새 주장은 통계적 유의성이 확인되는 때에 받는다.

7. 베이즈 철학(Bayesian philosophy): 경험적 사실엔 잡음이 섞여 있다. 원인과 결과는 1대 1 대응이 아니다. 그런 이유로, 관측 사실에 대한 해석에 복수의 틀(frame)을 고려할 필요가 있다. 각 틀에 대한 믿음에는 개인 간 차이가 있다. 그러나 베이즈적 업데이트 후에는 차이가 좁혀진다. [선순환이 작동하는 시스템]
8. 회귀(regression): 모평균보다 우수한 부모의 자녀들은 부모보다 덜 우수할 확률이 크다. 부모들은 자녀가 자신의 기대에 차지 않는다고 압박해서는 안 된다. 자신들의 잣대를 자녀에 가져다대지 않아야 한다. 성장기 자녀는 각자 다른 방식으로 우수하다. 그것이 찾아내고 키워지는 환경을 만들어주는 것이 부모가 할 일이다.
9. 과잉학습(over-training): 데이터의 세세한 부분까지 맞춘 모형은 정교해 보이지만 재현성이 떨어진다. 일반화에 취약하기 때문이다. 과잉학습이 단기적 목표 달성에는 효과적이지만 장기적 성장을 방해한다.
10. 자유(freedom) vs. 규제(regularization): 모형은 유연하여야 하지만 지나친 자유는 해가 된다. 적당히 규제된, 그러나 유연한 모형이 일반화에서 성능이 좋다. 자유와 규제는 튜닝(tuning)의 건(件)이다.
11. 다양성과 독립성: 복잡한 상황에서는 우수한 모형 하나보다 덜 우수한 여러 모형들의 총합이 더 우수하다 (모형 앙상블, model ensemble). 전제 조건은 다양성과 독립성이다.
12. 신경망: 심층 신경망의 성능은 여러 겹으로 쌓인 잠재층 구조에서 나온다. 드러난 것의 내부에 그것을 결정하는 숨은 요인들이 있다. 그것을 통찰해야 한다.
13. 강화학습: Balance between Exploitation and Exploration. 우연을 회피하면 기회도 없다. 실패로부터 배운다. 물론 실패는 쓰라리다. 균형이 중요하다.
14. GPT: Generative Pre-trained Transformer. Pre-training 대 Fine-tuning의 비율은 8:2가 맞다. 특별한 목적 없는 ‘읽고 쓰고 말하기’가 기본이다.

Calculation of Enterprise Capital via Least-squares Monte Carlo: Regress Now or Later?

하홍준¹, Daniel Bauer²

Abstract

Interest in estimating risk capital within enterprise risk models has been steadily growing, particularly with the application of Monte Carlo methods. A key challenge in this domain involves accurately establishing the distribution of a company's available capital, which depends on the conditional expectation of future cash flows. Two prominent approaches are employed for this purpose: the regress-now method, which projects cash flows and applies regression to their discounted values on basis functions, and the regress-later method, which approximates cash flows through realizations of manageable processes and then computes the conditional expectation in closed form. This paper demonstrates that the left- and right-singular functions of the valuation operator serve as robust approximators in both approaches, enhancing their overall performance. Furthermore, we delineate scenarios where each method outperforms the other, with the regress-later method showing superior performance under certain conditions, while the regress-now method generally exhibits greater robustness.

¹고려대학교 통계학과

²University of Wisconsin-Madison

Integration of Traditional and Telematics Data for Efficient Insurance Claims Prediction

Hashan Peiris¹, Himchan Jeong¹, Jae-kwang Kim², Hangsuck Lee³

Abstract

While driver telematics has gained attention for risk classification in auto insurance, scarcity of observations with telematics features has been problematic, which could be owing to either privacy concerns or favorable selection compared to the data points with traditional features. To handle this issue, we apply a data integration technique based on calibration weights for usage-based insurance with multiple sources of data. It is shown that the proposed framework can efficiently integrate traditional data and telematics data and can also deal with possible favorable selection issues related to telematics data availability. Our findings are supported by a simulation study and empirical analysis in a synthetic telematics dataset.

¹Simon Fraser University

²Iowa State University

³Sungkyunkwan University

Counting Data의 보험료 순서 모델링을 위한 코플라 상태모형

*Jae Youn Ahn*¹

Abstract

In insurance ratemaking, it is natural to charge higher premiums to policyholders with larger claim histories. Violations of this ordering principle could undermine the stability of the insurance system. In this study, we propose a copula-based state-space model specifically tailored for counting data, ensuring that this critical ordering is maintained. Due to the intrinsic complexity of copula models for counting data, computational costs become prohibitively high, particularly as the number of observation periods increases. To overcome this computational challenge, we further introduce a statistical method utilizing MCMC sampling with latent variables.

Keywords: Poisson - hurdle mixture; comonotonic random effects; credibility ordering; weak stationarity.

JEL: C22; C58

MSC: 62M10; 62P05

¹Department of Statistics, Ewha Womans University, Seoul, Republic of Korea.

Learning Robust Tabular Models under Unseen Missingness Shifts

Jihye Lee¹, Miseo Kang², Dongha Kim³

Abstract

The presence of missing values often reflects variations in data collection policies, which may shift across time or locations, even when the underlying feature distribution remains stable. Such shifts in the missingness distribution between training and test inputs pose a significant challenge to achieving robust predictive performance. In this study, we propose a novel deep learning framework designed to address this challenge, particularly in the common yet challenging scenario where the test-time dataset is unseen. We begin by introducing a set of mutual information-based conditions, called MI robustness conditions, which guide the prediction model to extract label-relevant information. This promotes robustness against distributional shifts in missingness at test-time. To enforce these conditions, we design simple yet effective loss terms that collectively define our final objective, called MIRRAMS. Importantly, our method does not rely on any specific missingness assumption such as MCAR, MAR, or MNAR, making it applicable to a broad range of scenarios. Furthermore, it can naturally extend to cases where labels are also missing in training data, by generalizing the framework to a semi-supervised learning setting. Extensive experiments across multiple benchmark tabular datasets demonstrate that MIRRAMS consistently outperforms existing state-of-the-art baselines and maintains stable performance under diverse missingness conditions. Moreover, it achieves superior performance even in fully observed settings, highlighting MIRRAMS as a powerful, off-the-shelf framework for general-purpose tabular learning.

Keywords : Missingness shifts, MI robustness conditions, Deep learning.

¹Master's student, Department of Statistics, Sungshin Women's University, 34 Da-gil, Bomun-ro, Seongbuk-gu, Seoul 02844, Republic of Korea. E-mail: jihyelee987@gmail.com

²Master's degree, Department of Statistics, Sungshin Women's University, 34 Da-gil, Bomun-ro, Seongbuk-gu, Seoul 02844, Republic of Korea. E-mail: krystalms0306@gmail.com

³(Corresponding Author) Assistant Professor, School of Mathematics, Statistics and Data Science, Sungshin Women's University, 34 Da-gil, Bomun-ro, Seongbuk-gu, Seoul 02844, Republic of Korea.
E-mail: dongha0718@sungshin.ac.kr

L_0 -regularized complex censored rank regression and its application*

선수연¹, Dipankar Bandyopadhyay², 최태화³

Abstract

Broken adaptive ridge (BAR) penalty approximates L_0 regularization through iterative reweighting of L_2 penalties. This penalty possesses both the oracle property and the grouping effect for highly correlated covariates which are the main desirable theoretical properties in penalized regression. In this study, we propose the BAR-penalized linear rank regression via induced smoothing for the semiparametric accelerated failure time model with complex censored data. To achieve the tractable computation and the reliable inference, we adopt an induced smoothing approach to the Gehan-type rank estimator. For scalable penalization, we develop a cyclic coordinate descent algorithm that minimizes the penalized objective function and estimates the regression coefficients in a coordinate-wise manner. We further extend the proposed method to more complex survival data, such as multivariate partly interval-censored data. Under mild conditions, the proposed estimator satisfies both the oracle property and the grouping effect. Numerical studies compare our approach with several well-known penalties and demonstrate its superior selection accuracy and estimation efficiency across various scenarios. An application to real clinical data further confirms the practical usefulness of the proposed method.

Keywords : Broken adaptive ridge regression, Induced smoothing, Accelerated failure time model, Coordinate descent, Interval censoring, Clustered data.

*This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (RS-2024-00340298).

¹Graduate Student, School of Mathematics, Statistics and Data Science, 2, Bomun-ro 34dagil, Seongbuk-gu, Seoul 02844, South Korea.

²Professor, Virginia Commonwealth University.

³(Corresponding Author) Assistant Professor, School of Mathematics, Statistics and Data Science, 2, Bomun-ro 34da-gil, Seongbuk-gu, Seoul 02844, South Korea. E-mail: tchoi@sungshin.ac.kr

Conditional Dependency Modeling in Tabular Pretraining

Beomjin Park¹, Hosik Cho²

Abstract

Tabular data remain one of the most fundamental and widely used formats for representing structured information. Despite the growing prominence of unstructured and semi-structured data sources, tabular datasets continue to serve as a cornerstone across numerous application domains. Traditional machine learning algorithms, such as linear regression, support vector machines, and decision trees, have long been optimized for tabular analysis and have achieved substantial success in identifying predictive relationships and constructing accurate models from structured records. Recently, conditional dependency learning approaches, including masked language modeling, have extended the pretraining paradigms of language models, enabling embedding-based representations to be effectively integrated with downstream tasks. Building upon this line of research, we propose conditional dependency modeling for tabular data via BERT, which employs binning-based tokenization of tabular variables and learns embedding representations that capture conditional dependencies and feature interactions. Empirical evaluations demonstrate that our method consistently improves downstream performance, underscoring the potential of BERT-style pretraining for understanding structured data.

Keywords : Conditional dependency modeling, Masked language modeling, Downstream tasks, Binning-based tokenization.

¹Assistant Professor, Department of Information and Statistics, Gyeongsang National University, 501 Jinju-daero, Jinju-city, Gyeongnam, 52585, Korea. E-mail: bbbeomjin@gmail.com

²(Corresponding Author) Professor, Department of Urban Big Data Convergence, University of Seoul, 163 Seoulsiripdae-ro, Seoul, 02504, Korea. E-mail: choi.hosik@uos.ac.kr

Quantile and Rank Estimation for Encrypted Data^{*}

어수행¹, 김재선², 최호식³

요약

본 발표는 동형암호(Homomorphic Encryption, HE) 환경에서의 데이터 분석과 통계 추론의 기초 개념을 소개하고, 특히 암호화된 자료로부터 분위수(quantile)와 순위(rank)를 효율적으로 추정하는 대표적 접근법을 다룬다. 기존의 정렬 기반 통계량 계산은 암호문 간 비교 연산의 높은 계산 복잡도로 인해 비효율적이므로, 이를 완화하기 위한 근사 최적화 기법과 비모수적 추론 방법이 활발히 연구되고 있다. 발표에서는 스무딩된 quantile loss 함수를 이용한 분위수 추정과, knot 기반 근사 누적분포함수(approximate ECDF)를 활용한 순위 추정 원리를 비교·설명한다. 아울러 이러한 방법들이 Spearman 상관계수, 중앙값, 사분위수 등 주요 통계량 계산에 적용된 사례를 통해 실험적으로 검증된 결과를 소개한다. 마지막으로, 공공데이터·의료·교통 등 개인정보 보호가 요구되는 분야에서의 활용 가능성을 논의하고, 도로공사 안심구역 내 구현된 암호분석 플랫폼 사례를 통해 프라이버시 보존 데이터분석의 실증적 가능성을 제시한다.

주요 용어 : 동형암호, 분위수, 순위 통계량, Privacy-Preserving Machine Learning, CKKS Scheme.

^{*}[국토교통 데이터 산업 생태계 플랫폼 구축(R&D) 과제번호 : RS-2022-00144012]에 의해 지원받았음.

¹08826 서울시 관악구 관악로 1, 서울대학교 27-404, (주)크립토크, 데이터사업개발실 팀장.

²08826 서울시 관악구 관악로 1, 서울대학교 27-404, (주)크립토크, 데이터사업개발실 책임연구원.

³02504 (교신저자) 서울시 동대문구 시립대로 163, 서울시립대학교 일반대학원 도시빅데이터융합학과.

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

Bayesian Online Change-point detection^{*}

Daewon Yang¹, Taeryon Choe²

Abstract

We develop an online change-point detection method that enforces a no-return constraint on regime indices via a probit stick-breaking construction. This yields contiguous, forward-only regimes suitable for time-ordered data while preserving flexible, data-driven allocation through “rich-get-richer” weights with bounded growth. A fast sequential update identifies candidate changes as data arrive, and a retroactive correction step revisits early assignments using prior-and-evidence - based comparisons to refine change-point locations with low computational cost. For robustness, we include a scale-mixture formulation that induces multivariate t-likelihoods without sacrificing efficient updates. Simulations and illustrative applications indicate accurate localization and competitive runtime in both online and offline use.ifts in functional data across various applications.

Keywords : Change-point model; Bayesian Nonparametrics; Statistical process control.

^{*}This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (RS-2022-NR068754).

¹Department of Information and Statistics, Chungnam National University, Daejeon, South Korea.

²Department of Statistics, Korea University, Seoul, South Korea.

순서형 반응변수에 대한 비모형적 회귀 연관성 탐색: 소방공무원 PTSD 사례

전세봄¹, 김대영²

요약

본 연구는 순서형 반응변수를 갖는 다변량 범주형 자료를 이용하여 모형 가정을 최소화한 비모형적(non-model-based) 회귀 연관성 탐색 절차를 제안한다. 2023년 소방공무원 정신건강 실태조사 자료를 이용하여, 외상후 스트레스 장애(Post-Traumatic Stress Disorder, PTSD)와 관련된 다양한 예측변수를 다차원 분할표를 이용하여 연관성을 파악하였다. 제안된 절차는 Scaled Checkerboard Copula Regression Association Measure (SCCRAM)과 비모수적 부트스트랩 및 퍼뮤테이션 방법을 통합하여, 다차원 교차표에서의 회귀적 연관성을 다각적인 측면 즉 전반적(전체), 주변적(마진), 조건적(컨디셔널) 수준에서 평가한다. 분석 결과, 우울(Depression)과 불면(Insomnia) 요인이 PTSD 심각도와 가장 높은 연관성을 보였으며, 직무 및 인구통계학적 요인의 효과는 정신건강 요인을 통제할 때 더 분명하게 나타났다. 본 연구는 순서형 범주형 자료에서의 탐색적 회귀 구조 분석을 위한 통합적 데이터 분석 프레임워크를 제시하며, 정신건강 빅데이터와 유사한 다차원 범주형 자료의 탐색 연구에 실질적인 통계적 기반을 제공한다.

¹목원대학교 마케팅빅데이터학과.

²메사추세츠대학 수확통계학과.

Moment Condition Models with Conditional Dirichlet Process Priors

Jaegui Lee¹, Seongil Jo², Kyoungjae Lee³, Kwangmin Lee⁴, Jaeyong Lee⁵

Abstract

A moment condition model is a type of nonparametric model in which the parameter of interest is characterized through a set of moment equations. Such models allow one to focus on finite-dimensional parameters of interest without fully specifying the infinite-dimensional aspects of the data-generating process. The class of moment condition models is broad, encompassing standard moment models, linear regression models, nonlinear regression models, and instrumental variable models. In this paper, we propose a method for computing the posterior distribution of a moment condition model under the conditional Dirichlet process prior introduced by Lee (2025). Previous approaches have relied on importance sampling based on the Bayesian bootstrap posterior, which can be computationally efficient but becomes challenging to implement in high-dimensional parameter spaces. We address this limitation by developing a posterior sampling scheme using the constrained Hamiltonian Monte Carlo method. The proposed approach is applicable to a wide range of moment condition models and scales effectively to high-dimensional settings. We demonstrate the method through applications to multiple datasets across various moment condition model specifications.

Keywords : Dirichlet process; conditional Dirichlet process; functional condition model; moment condition model.

¹SK Innovation, Republic of Korea.

²Department of Statistics, Inha University, Republic of Korea.

³Department of Statistics, Sungkyunkwan University, Republic of Korea.

⁴Department of Big Data Convergence, Chonnam National University, Republic of Korea.

⁵Department of Statistics, Seoul National University, Republic of Korea.

Hybrid Bayesian Network Models using Mixed Data in Panel Survey

*Dong-Hee Lee*¹

Abstract

Mixed data refer to data that contain both discrete and continuous data. By this definition, most data are mixed data. Nevertheless, the term "mixed data" is used because statistical inference processes that handle both discrete and continuous random variables are challenging. To achieve this, the joint probability distribution between discrete and continuous random variables must be modeled. This can be accomplished using various methods, including conditional distribution models, mixture distribution models using latent variables, and methods that combine different distributions such as Copula or Sarmanov family. The hybrid Bayesian network model discussed in this study utilizes conditional distributions among these methods. Bayesian network models can derive the joint probability distribution by reflecting the conditional independence and dependence between variables using directed acyclic graph models. Hybrid Bayesian networks extend it to include both discrete and continuous variables. This study examines the characteristics and limitations of hybrid Bayesian network models and proposes an improvement by applying alternative methods for handling mixed data. In particular, we propose an application method of an improved hybrid Bayesian network model for imputing missing values in panel data and examine the analysis results using the Korea Media Panel Survey data.

Keywords : Bayesian network, EM algorithm, missing imputation, mixed data, Sarmanov family.

¹16227 경기도 수원시 영통구 광고산로 154-42, 경기대학교 경영학부 교수. E-mail: dhl@kyonggi.ac.kr

Cumulative and Lagged Non-linear Associations Between Particulate Matter Exposure and Lung Function

*Eun-Ji Lee¹, Narae Jo¹, Tae-Young Heo¹, Kim Young-Youl²,
Eunjin Kwon², Min Gu Kang², Jae-Hwan Jhong³*

Abstract

In this study, we evaluate the association between ambient air pollution exposure and lung function using a Distributed Lagged Nonlinear Model (DLNM), which allows for quantification of nonlinear lagged and cumulative effects over time. The dataset used in this study is the 2018 Korea National Health and Nutrition Examination Survey (KNHANES) linked with national air pollution databases. In this dataset, $PM_{2.5}$, SO_2 , NO_2 , O_3 , and CO are considered as exposure variables and chronic obstructive pulmonary disease (COPD) status (binary) is used as the response variable. Because air pollutants accumulate in the lungs over time, DLNM allows us to characterize delayed and nonlinear impacts simultaneously. Higher $PM_{2.5}$ concentrations are associated with increased COPD risk, with the strongest effect observed three days after exposure, and larger effect sizes as cumulative exposure lengthens. These findings underscore the cumulative risk of air pollution on respiratory health and provide evidence to support targeted public health interventions and improvements in air quality policy in Korea. Future work includes extending the analysis to evaluate the relationship between air pollution and forced expiratory volume in 1 second among individuals with COPD.

Keywords : Air pollution; chronic obstructive pulmonary disease; distributed lagged nonlinear model; delayed effect; KNHANES.

¹Department of Information Statistics, 1, Chungdae-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do 28644, Republic of Korea.

²Division of Allergy and Chronic Respiratory Diseases, 187, Osongsaengmyeong2(i)-ro, Osong-eup, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do 28159, Republic of Korea.

³(Corresponding Author) Associate Professor, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea. E-mail: jjh25@chungbuk.ac.kr

Analyzing spatial effects using GMM estimators for binary dependent variable: A case study of the National Consortium on Violence Research data^{*}

Daho Lee¹, Hee-Young Kim²

Abstract

The term spatial econometrics was coined by the Belgian economist Jean Paelinck in 1974 during an address to the Dutch statistical association. In recent years, spatial econometrics has gained interest and importance in econometrics. The proliferation of applications in various areas has been supported by the growth of available software. Currently, spatial econometrics routines to estimate spatial models are available from many commercial and noncommercial software, such as MATLAB, Stata, and Pysal, among others. R is with no doubt the open source environment that contains the richest variety of options. Social agents are faced with decisions that are intrinsically discrete. There are several situations in which the variable that needs to be explained takes only two different values. Despite the huge availability of software to estimate cross-sectional spatial models, there are only few functions to estimate models dealing with spatial categorical dependent variable. Recently, Sarrias and Piras (2023) introduced the R package called *spldv* offering accessible tools for those researchers who need to model spatial binary dependent variable with GMM estimator. In this study, we apply these GMM estimators to a case study analyzing the National Consortium on Violence Research (NCOVR) dataset, which includes data for 3,085 US counties. For this analysis, we first construct the binary dependent variable, *CRIME_HIGH*. This variable is derived from the 1980 homicide rate (*HR80*) and takes a value of '1' if a county's rate is above the dataset's median, and '0' otherwise. Using the spatial Durbin binary model, we examine the impact of two key socio-economic independent variables: *P580* and *UE80*. This approach allows us to estimate both the direct effects of these variables and their spatial spillover effects on the probability of a county experiencing a high homicide rate.

Keywords : Binary; GMM; R; Spatial Durbin binary model; Spatial econometrics.

^{*}This research is supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07045707, NRF-2021R1F1A1048309). Also this article is financially supported by the 2025 College of Public Policy at Korea University.

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

²(Corresponding Author) Professor, Bigdata Science, Korea University, Sejong-Street 2511 Sejong-City 339-700, E-mail : starkim@korea.ac.kr

Big Data Analysis for AI-Generated Social Media Threads

Kin Lok Wong¹, Jai Woo Lee¹

Abstract

Artificial Intelligence (AI) has been used to develop many fields including social network service like X, Facebook, Instagram, etc. and AI-based service such as ChatGPT, Gemini, etc. while products generated by AI may affect users with confounding messages or fake news which are common in the social media. The objective of this study is to improve AI techniques which can be used to analyze large datasets in fast-growing social media environments. To effectively analyze large datasets, sophisticated AI techniques, including long short-term memory networks, have proven proficiency in monitoring opinion trends and handling misinformation. Generative AI can provide avenues for the creation of online contents and increasing the possibility of misinformation. The proposed approaches can be utilized to prevent the abuse of user information by specific APIs and offer a comprehensive overview of the applications of AI in social media environment. The extensive effectiveness and efficiency of this interdisciplinary link will have a great impact on strategic decision-making and future research in the era of big data.

Keywords : Artificial Intelligence, Big Data, Natural Language Processing, API, Social Media.

¹고려대학교 빅데이터사이언스학과

감소성 고장 검출률 함수를 활용한 소프트웨어 신뢰성 모형에 관한 연구*

공영선¹, 장인홍², 송광윤³, 김운수⁴

요 약

소프트웨어는 현대 사회의 다양한 산업과 일상 영역에서 핵심적인 역할을 수행하고 있다. 그러나 소프트웨어에 오류가 발생하면 단순한 불편을 넘어 심각한 인명 피해나 경제적 손실로 이어질 수 있다. 예를 들어, 최근 호주에서는 소프트웨어 오류로 인해 긴급전화 연결되지 않아 최소 3명이 사망하였으며, 드론 함정이 소프트웨어 고장으로 인해 정지하는 사건 등 다양한 사고가 보고되고 있다. 이러한 문제를 예방하기 위해 소프트웨어 신뢰성을 분석하고 예측하는 연구가 활발히 진행되고 있다. 전통적인 소프트웨어 신뢰성 성장모형(Software Reliability Growth Model; SRGM)은 주로 비동질 포아송 과정(Non-Homogeneous Poisson Process; NHPP)을 기반으로 평균값 함수를 모형화해 왔다. 그러나 기존의 많은 NHPP 기반 모형들은 시간이 지남에 따라 감소하는 고장 검출률을 충분히 반영하지 못하는 한계가 있다. 이에 본 논문에서는 이러한 한계를 극복하기 위하여 시간에 따라 감소하는 고장 검출률을 고려한 새로운 NHPP 기반 신뢰성 성장모형을 제안하였다. 제안한 모형의 모수는 실제 소프트웨어 고장 데이터를 이용하여 최소제곱법으로 추정하였으며, 기존의 10개 SRGM과 비교하여 9개의 평가 지표(MSE, RMSPE, PRV, TS 등)를 통해 모형의 성능을 검증하였다. 분석 결과, 제안된 모형은 두 개의 실제 데이터 세트에서 가장 낮은 오차값을 기록하였으며, 적합도와 예측력 모두에서 기존 모형들보다 우수한 성능을 나타냈다. 또한 잔차가 0을 중심으로 고르게 분포하여 모형의 안정적 적합성이 확인되었다. 이러한 결과는 제안된 NHPP 기반 모형이 시간에 따라 감소하는 고장 검출 효율의 현실적 특성을 효과적으로 반영하며, 소프트웨어 신뢰성 예측의 정확도를 향상시킬 수 있음을 보여준다.

주요 용어 : 소프트웨어 신뢰성, 고장 검출률, 비동질 포아송 과정, 소프트웨어 고장.

*이 논문은 2025학년도 조선대학교 학술연구비의 지원을 받아 연구임. 이 논문은 2024년도 정부(교육부)의 재원으로 한국연구재단의 G-램프(LAMP) 사업 지원을 받아 수행된 연구임(No. RS-2023-00285353).

¹61452 광주광역시 동구 조선대3길 30, 조선대학교 전산통계학과, 박사수료. E-mail: kongyeongseon@gmail.com

²(교신저자) 61452 광주광역시 동구 조선대3길 30, 조선대학교 컴퓨터통계학과, 교수.

E-mail: ihchang@chosun.ac.kr

³61452 광주광역시 동구 조선대3길 30, 조선대학교 컴퓨터통계학과, 조선대학교 G-LAMP 사업단, 조교수.

E-mail: cssig@chosun.ac.kr

⁴61452 광주광역시 동구 조선대3길 30, 조선대학교 G-LAMP 사업단, 박사후연구원.

E-mail: imk92315@naver.com

Adaptive RKHS for Nonlinear Modeling with Interactions

Sehun Jang¹, Jun Song²

Abstract

We present a new framework for nonlinear scalar-on-function regression that simultaneously performs functional predictor selection and captures higher-order interaction effects among multivariate functional inputs. Our approach is built on a tensor product of reproducing kernel Hilbert spaces (RKHSs), each equipped with its own kernel and hyperparameter. Unlike conventional methods that treat kernel parameters as fixed tuning constants, we estimate them directly from the data, allowing the RKHS to adaptively adjust to the underlying functional structure. This adaptive construction enables flexible modeling of complex nonlinear and interaction effects while preserving interpretability. We further propose a theoretically grounded importance measure that quantifies the contribution of each functional predictor to the response. Theoretical analysis establishes consistency and optimality properties for both the estimator and the importance measure, and simulation studies and real-world applications demonstrate substantial gains in accuracy, interpretability, and variable selection performance.

Keywords : Adaptive RKHS; Importance measure; Interaction-aware; Scalar-on-function regression; functional predictor selection.

¹Ph.D. Student, Department of Statistics, College of Political Science and Economics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: sahoon1004@korea.ac.kr

²(Corresponding Author) Associate 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

Valid Asymptotic Inference after Sufficient Dimension Reduction in a Single-Index Framework

Kyongwon Kim¹, Kyongwon Kim², Kyongwon Kim³

Abstract

Despite the explosive development of sufficient dimension reduction methods, there has been limited discussion regarding the conduct of statistical inference after dimension reduction. Dimension reduction produces sufficient predictors, but this is not the endpoint of data analysis. To ensure comprehensive analysis, confidence intervals and p-values are crucial for statistical inference. Currently, the common practice involves applying the sufficient predictors in subsequent modeling as if they were the true predictors. However, this practice often leads to overly optimistic results in statistical inference. In this paper, we demonstrate how errors resulting from dimension reduction affect the standard error of regression parameters in a two-step procedure using the von-Mises expansion. We specifically choose ordinary least squares as a dimension reduction method and employ linear and logistic regression as modeling tools. Furthermore, we investigate how to improve statistical inference when the number of covariates exceeds the sample size, utilizing the seeded regression method. We also develop asymptotic theory for seeded regression with estimated seeds in high-dimensional settings. In the numerical study, we compare coverage probabilities and the growth rate of local power before and after considering the errors induced by the dimension reduction step. Finally, we apply our approach to three real datasets, covering a variety of modeling cases.

¹연세대학교 응용통계학과

²고려대학교 통계학과

³이화여자대학교 통계학과

Spatial covariance estimation in massive raster data

Joonho Shin¹, Wei-Ying Wu², Chae Young Lim³

Abstract

With advances in technology, data collection has improved in both quality and quantity. In particular, it is now common for data sets to include location information for each individual observation. This information can be exploited to investigate the spatial correlation present in the data. Such correlation matters because applying statistical methods that assume independence to data with clear spatial dependence can lead to misleading conclusions and poor decisions. One of the most popular models for incorporating spatial correlation is the Gaussian random field (also known as a Gaussian process). In a Gaussian random field, the spatial correlation structure is specified by a covariance function, which is often assumed to follow a parametric model, or at least to be well approximated by one. The parameters of this covariance function can be estimated using several methods, including the variogram method, the method of moments, and likelihood-based methods. It is important to assess the accuracy of the estimated covariance parameters, since the performance of subsequent statistical analyses based on random field models depends critically on the quality of the estimated covariance structure. In this talk, we introduce a new spectral-domain approach for estimating spatial covariance structures when data are observed on a dense grid over a fixed bounded region. The method is based on the Whittle likelihood, a traditional spectral-domain approach that is nevertheless known to have serious drawbacks for spatial and densely observed data. Our new approach, which modifies the original Whittle likelihood, is shown to overcome these drawbacks at the cost of some additional computation. We prove that our approach can achieve an asymptotically equivalent order of accuracy to the maximum likelihood estimator (MLE), while significantly reducing computational cost by exploiting the FFT algorithm and parallelizing computations. Additionally, our method does not require a specific parametric covariance model and provides consistent estimators in general dimensions and for domains of general shape, among other advantages. Despite these strengths, the method also has limitations; an obvious one is that it cannot be applied to data observed at irregular locations. Nonetheless, we argue that our method is particularly useful for high-resolution spatial data, because it exploits the fine-scale information as effectively as possible, whereas many existing statistical methods must compromise on spatial resolution for computational feasibility.

Keywords : Spatial Statistics, Spectral analysis, Gaussian random fields

¹Assistant Professor, School of Mathematics, Statistics and Data Science, Sungshin Womens' University, 2, Bomun-ro 34da-gil, Seongbuk-gu, Seoul, 02844, Republic of Korea.

E-mail: joonho.shin@sungshin.ac.kr

²Professor, Department of Applied Mathematics, National Dong Hwa University, No. 1, Sec. 2, Da Hsueh Rd., Shoufeng, Hualien, 97401, Taiwan.

²Professor, Department of Statistics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Korea.

spheresmooth: An R Package for Penalized Piecewise Geodesic Curve Fitting on a Sphere^{*}

*Jae-Hwan Jhong*¹, *Seyoung Lee*², *Ja-Yong Koo*³, *Kwan-young Bak*⁴

Abstract

This study introduces an R package spheresmooth, which implements a penalized piecewise geodesic curve fitting method on a sphere. Spherical data observed over a continuum arise frequently in various fields including cardiology, computer vision, physiology, and geophysics. We propose an adaptive smoothing method by extending the linear spline approach to spherical data. Penalization based on differences of velocity vectors endows sparsity among control points of the spherical curve, which enables data-adaptive curve fitting. The proposed method is implemented with a Riemannian block coordinate descent algorithm. Illustrations on Triassic and Jurassic polar wander data and tropical cyclone data demonstrate practicality of the proposed method and the associated spheresmooth package.

Keywords : coordinate descent, penalization, piecewise geodesic curve, Riemannian geometry, spherical data.

^{*}This work was supported by National Research Foundation (NRF) of Korea, RS-2024-00342014.

¹Professor, Department of Information Statistics, ChungBuk National University, 1, Chungdae-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do, 28644, Republic of Korea. jjh25@chungbuk.ac.kr

²Master's student, Department of Statistics, Sungshin Women's University, 2, Bomun-ro 34 Da-gil, Seongbuk-gu, Seoul, 02844, Republic of Korea. 220254005@sungshin.ac.kr

³Professor, Department of Statistics, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, Korea.

⁴(Corresponding Author) School of Mathematics, Statistics and Data Science, Sungshin Women's University, 2, Bomun-ro 34 Da-gil, Seongbuk-gu, Seoul, 02844, Korea. E-mail: kybak@sungshin.ac.kr

회전성을 갖는 제2종 중심합성계획의 직교블로킹

김혁주¹

요 약

본 논문에서는 제2종 중심합성계획(CCD2)의 직교블로킹에 관하여 연구하였다. CCD2가 직교블로킹을 갖기 위한 조건과 특히 회전성을 갖는 CCD2가 직교블로킹을 갖기 위한 조건을 구하였으며, 회전성과 직교블로킹을 동시에 갖는 CCD2들을 구하였다.

주요 용어 : 제2종 중심합성계획, 직교블로킹, 회전성.

1. 서론

2차모형의 반응표면분석을 위한 실험을 b 개의 블록으로 실시하려 한다고 하자. 이 경우 2차모형은 다음과 같다.

$$y_u = \beta_0' + \sum_{i=1}^k \beta_i x_{iu} + \sum_{i=1}^k \beta_{ii} x_{iu}^2 + \sum_{i < j}^k \beta_{ij} x_{iu} x_{ju} + \sum_{m=1}^b \delta_m (z_{mu} - \bar{z}_m) + \epsilon_u \quad (u = 1, 2, \dots, N)$$

여기서 y 는 반응변수이고 x_1, x_2, \dots, x_k 는 설명변수이며, u 번째 관찰점이 m 번째 블록에 포함되어 있으면 $z_{mu} = 1$ 이고, 그렇지 않으면 $z_{mu} = 0$ 이다.

$\beta_0', \beta_i, \beta_{ii}, \beta_{ij}$ 의 최소제곱추정량을 각각 $b_0', b_i, b_{ii}, b_{ij}$ 라 하자. 블록효과가 $b_0', b_i, b_{ii}, b_{ij}$ 와 직교하는 성질을 직교블로킹(orthogonal blocking)이라 하는데, 1차적률과 2차혼합적률이 0인 반응표면계획들의 경우 직교블로킹이 성립하기 위한 조건은 다음과 같다(Myers, 1976, p.184).

$$\sum_{i \in \mathcal{I}_m} x_{iu} = 0 \quad (m = 1, 2, \dots, b) \quad (1)$$

$$\sum_{i \in \mathcal{I}_m} x_{iu} x_{ju} = 0 \quad (m = 1, 2, \dots, b) \quad (i, j = 1, 2, \dots, k; i \neq j) \quad (2)$$

$$\sum_{u=1}^N x_{iu}^2 z_{\mu} = \sum_{u=1}^N x_{iu}^2 \bar{z}_m \quad (3)$$

여기서 $\sum_{i \in \mathcal{I}_m}$ 이라는 표시는 m 번째 블록 안에 있는 관찰점들에 걸쳐 합을 취하는 것을 의미한다.

2차반응표면모형에 아주 많이 사용되는 계획으로 Box와 Wilson (1951)에 의해 제시된 중심합성계획(central composite design; 이하 CCD1으로 표시)이 있다. 설명변수가 두 개($k=2$)인 경우를 예로 들면, CCD1은 $2^k = 4$ 개의 요인실험점 $(1,1), (1,-1), (-1,1), (-1,-1)$ 과 $2k=4$ 개의 축점 $(\alpha, 0), (-\alpha, 0), (0, \alpha), (0, -\alpha)$ 및 n_0 개의 중심점 $(0,0)$ 으로 이루어진다. 블록으로 나누어서 실험을 하는 경우 각 블록에 주어진 중심점의 개수에 따라 CCD1이 직교블로킹을 갖기 위한 α 의 값이 정해진다.

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

한편 반응표면계획이 가질 수 있는 바람직한 성질 중 Box와 Hunter (1957)에 의해 제시된 회전성(rotatability)이 있다. 점 \underline{x} 에서 $E(y)$ 의 추정량인 $\hat{y}(\underline{x})$ 의 분산은 일반적으로 점 \underline{x} 의 위치에 따라 달라지는데, 만약 이 분산이 중심점과 점 \underline{x} 사이의 거리만의 함수라면 이러한 성질을 회전성이라고 한다. 회전성을 갖는 CCD1이 직교블로킹을 갖는 경우가 $k=2$, $k=4$ 일 때는 존재하지만 $k=3$ 일 때는 존재하지 않으며, $k=5$ 이면 요인실험점으로 $F=32$ 개를 사용할 때는 존재하지 않고 $F=16$ 개를 사용할 때는 존재한다는 것이 밝혀져 있다(Myers, 1976, p.187).

Kim (2002)은 제2종의 중심합성계획(이하 CCD2로 표시)을 제시하였다. 이 계획은 CCD1의 축점부분을 확장한 것으로, $k=2$ 인 경우를 예로 들면 $2^k=4$ 개의 요인실험점 $(1,1), (1,-1), (-1,1), (-1,-1)$ 과 $4k=8$ 개의 축점 $(\alpha_1, 0), (-\alpha_1, 0), (0, \alpha_1), (0, -\alpha_1), (\alpha_2, 0)$, 그리고 n_0 $(-\alpha_2, 0), (0, \alpha_2), (0, -\alpha_2)$,개의 중심점 $(0,0)$ 으로 이루어진다. 여기서 α_1 과 α_2 는 축점의 위치를 결정하는 값이며, $\alpha_2 > \alpha_1 > 0$ 이다.

본 논문에서는 CCD2의 직교블로킹에 관하여 연구하고자 한다. CCD2가 직교블로킹을 갖기 위한 조건과 특히 회전성을 갖는 CCD2가 직교블로킹을 갖기 위한 조건을 구하며, 회전성과 직교블로킹을 동시에 갖는 CCD2들을 구하고자 한다.

2. 요인블록 1개, 축블록 1개로 하는 경우

CCD2에서 요인실험점의 수 F 는 $F=2^k$ (설명변수의 수가 많아서 일부실험시범을 쓰는 경우에는 $F=2^{k-p}$; p 는 적절한 양의 정수)이다. 먼저 다음과 같이 2개의 블록을 사용하는 경우를 생각하자.

블록 1(요인블록): F 개의 요인실험점과 F_0 개의 중심점

블록 2(축블록): $4k$ 개의 축점과 a_0 개의 중심점

중심점의 총수는 $n_0 = F_0 + a_0$ 이며, 실험점의 총수는 $N = F + F_0 + 4k + a_0$ 이다. 이 경우 CCD2가 직교블로킹을 갖기 위한 조건을 구해 보자. 식 (1)과 (2)가 성립함은 쉽게 확인할 수 있다. 식 (3)이 성립하려면 $F = (F + 2\alpha_1^2 + 2\alpha_2^2) \left(\frac{F + F_0}{N} \right)$ 와 $2\alpha_1^2 + 2\alpha_2^2 = (F + 2\alpha_1^2 + 2\alpha_2^2) \left(\frac{4k + a_0}{N} \right)$ 가 만족되어야 한다. 위의 두 식으로부터 다음을 얻을 수 있으며, 이것이 CCD2가 직교블로킹을 갖기 위한 조건이다.

$$\alpha_1^2 + \alpha_2^2 = \frac{F(4k + a_0)}{2(F + F_0)} \quad (4)$$

CCD2가 회전성을 갖기 위한 조건은 $\alpha_1^4 + \alpha_2^4 = F$ 이다(Kim, 2002). 이것과 식 (4)를 연립으로 풀면 다음 결과를 얻는다.

$$\alpha_1 = \left[\frac{F(4k + a_0)}{4(F + F_0)} - \sqrt{\frac{F}{2} - \left\{ \frac{F(4k + a_0)}{4(F + F_0)} \right\}^2} \right]^{1/2} \quad (5)$$

$$\alpha_2 = \left[\frac{F(4k+a_0)}{4(F+F_0)} + \sqrt{\frac{F}{2} - \left\{ \frac{F(4k+a_0)}{4(F+F_0)} \right\}^2} \right]^{1/2} \quad (6)$$

식 (5)와 (6)에 의해 회전성과 직교블로킹을 동시에 갖는 CCD2들을 구하면 Table 1과 같다. $k=2$, $F=4$ 인 경우는 $F_0 \leq 5$ 인 범위에서, $k=3$, $F=8$ 인 경우는 $F_0 \leq 8$ 인 범위에서, 그리고 $k=4$, $F=16$ 와 $k=5$, $F=32$ 인 경우는 $F_0 \leq 10$ 인 범위에서 구한 것이다. 회전성과 직교블로킹을 동시에 갖는 CCD2가 존재하지 않는 경우는 식 (5)와 (6)에서 α_1 이나 α_2 가 허근이 되는 경우이다.

Table 1. CCD2's having both rotatability and orthogonal blocking (1 axial block)

(a) $k=2, F=4$									
F_0	2	3	3	4	4	4	5	5	5
a_0	0	0	1	1	2	3	2	3	4
α_1	0.9284	0.5567	0.8347	0.5177	0.7672	1.0219	0.4860	0.7147	0.9284
α_2	1.3434	1.4056	1.3692	1.4078	1.3825	1.3060	1.4093	1.3906	1.3434
(b) $k=3, F=8$									
F_0	5	6	6	7	7	7	8	8	8
a_0	0	0	1	0	1	2	0	1	2
α_1	1.0377	0.8271	1.0559	0.6325	0.8576	1.0717	0.4208	0.6776	0.8842
α_2	1.6172	1.6566	1.6123	1.6733	1.6526	1.6077	1.6801	1.6706	1.6487
(c) $k=4, F=16$									
F_0	7	8	9	9	10	10	10	10	10
a_0	0	0	0	1	0	1	1	2	2
α_1	1.5085	1.3130	1.1650	1.3944	1.0336	1.2403	1.2403	1.4811	1.4811
α_2	1.8137	1.8999	1.9398	1.8697	1.9633	1.9215	1.9215	1.8289	1.8289
(d) $k=5, F=32$									
F_0		9					10		
a_0		0					0		
α_1		1.7391					1.6093		
α_2		2.1864					2.2426		

3. 요인블록 1개, 축블록 2개로 하는 경우

이번에는 다음과 같이 3개의 블록을 사용하는 경우를 생각하자.

블록 1(요인블록): F 개의 요인실험점과 F_0 개의 중심점

블록 2(내축블록): $2k$ 개의 내축점과 a_{10} 개의 중심점

블록 3(외축블록): $2k$ 개의 외축점과 a_{20} 개의 중심점

여기서 내축점과 외축점은 각각 α_1 과 α_2 에 의해 정해지는 축점을 말한다.

이 경우 CCD2가 직교블로킹을 갖기 위한 조건을 $\alpha_1 = \sqrt{\frac{F(2k+a_{10})}{2(F+F_0)}}$, $\alpha_2 = \sqrt{\frac{F(2k+a_{20})}{2(F+F_0)}}$ 로 얻게 되며, 회전성과 직교블로킹을 동시에 갖기 위한 조건을 다음과 같이 얻는다.

$$F\{(2k+a_{10})^2+(2k+a_{20})^2\}=4(F+F_0)^2 \quad (7)$$

식 (7)에 의해 회전성과 직교블로킹을 동시에 갖는 CCD2들을 구하면 Table 2와 같다.

Table 2. CCD2's having both rotatability and orthogonal blocking (2 axial blocks)

(a) $k=2, F=4$					
F_0	6	9	11	13	21
a_{10}	2	1	5	4	3
a_{20}	4	8	8	11	20
α_1	1.0954	0.8771	1.0954	0.9701	0.7483
α_2	1.2649	1.3587	1.2649	1.3284	1.3856
(b) $k=4, F=16$					
F_0	14		18		24
a_{10}	1		0		4
a_{20}	4		7		8
α_1	1.5492		1.3720		1.5492
α_2	1.7889		1.8787		1.7889
(c) $k=5, F=16$					
F_0	6			9	
a_{10}	2			1	
a_{20}	4			8	
α_1	1.0954			0.8771	
α_2	1.2649			1.3587	

4. 논의

본 논문에서는 CCD2의 직교블로킹에 관하여 연구하였다. $k=3$ 인 경우와 $k=5, F=32$ 인 경우 회전성과 직교블로킹을 동시에 갖는 CCD1은 존재하지 않는데, CCD2에서는 존재한다는 것이 의미 있는 결과라고 할 수 있겠다.

References

- Box, G. E. P. and Hunter, J. S. (1957). Multifactor experimental designs for exploring response surfaces, *Annals of Mathematical Statistics*, 28, 195-241.
- Box, G. E. P. and Wilson, K. B. (1951). On the experimental attainment of optimum conditions, *Journal of the Royal Statistical Society, Series B*, 13, 1-45.
- Kim, H. J. (2002). Extended central composite designs with the axial points indicated by two numbers, *The Korean Communications in Statistics*, 9, 595-605.
- Myers, R. H. (1976). *Response Surface Methodology*, Blacksburg, VA: Author (distributed by Edwards Brothers, Ann Arbor, MI).

Deep Learning-Based Audio Deepfake Detection System for a Competitive Challenge

*Soyul Han¹, Taein Kang², Sunmook Chor³, Jaejin Seo², Sanghyeok Chung³,
Sumi Lee², Seungsang Oh⁴, Il-Youp Kwak⁵*

Abstract

This paper presents the CAU_KU team's approach to the Audio Deep Synthesis Detection (ADD) 2023 Challenge, focusing on Track 1.2 (audio fake detection) and Track 3 (deepfake algorithm recognition). We developed a comprehensive detection framework by extracting acoustic features from pretrained wav2vec2 networks, constant-Q transform (CQT), and mel-spectrograms. The representation extraction component of the AASIST model was modified to process 2D spectrograms derived from wav2vec2 and CQT features. Multiple deep learning architectures were investigated, and ensemble methods were employed to enhance detection performance. For Track 1.2, the Round 1 submission utilized an ensemble of CQT-LCNN and CQT-AASIST models, while the Round 2 submission incorporated an additional W2V2-GMM model. For Track 3, we ensembled CQT-LCNN, CQT-OFD, and AASIST models, integrating the OpenMax algorithm to detect unknown deepfake attacks. Our system achieved equal error rates of 23.44% and 21.26% in Rounds 1 and 2 of Track 1.2, respectively, securing third place in the track rankings.

Keywords : audio deepfake detection, deep learning, audio synthesis detection, ensemble learning.

¹Assistant Professor, Department of Big Data Applications, Hannam University, 70, Hannam-ro, Daedeok-gu, Daejeon, 34430, Korea. E-mail: soyul5458@hnu.kr

²Department of Statistics, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul, 06974, Korea.

³Department of Mathematics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea.

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

⁵(Co-corresponding author) Associate Professor, Department of Applied Statistics, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul, 06974, Korea. E-mail: ikwak2@cau.ac.kr

Multimodal AI 및 GAN을 활용한 국가승인통계의 격자 단위 기반 데이터 증폭

이동찬¹, 김기환²

요 약

국가승인통계는 막대한 공공 재원을 투입해 생산되는 고품질 데이터이지만, 행정구역 단위로 작성된 통계는 공간 해상도의 한계로 인해 실제 정책 수립이나 민간 데이터와의 융합 활용에 제약이 존재한다. 통계 작성 후 공개까지의 시차 문제는 제도적으로 개선이 어렵지만, 공간 해상도 제약은 기술적 접근을 통해 개선 가능하다고 판단하였다. 이에 본 연구에서는 행정구역 기반의 국가승인통계를 격자단위로 세분화하는 과정을 데이터 증폭이라는 개념으로 새롭게 제시하였다. 먼저 항공사진(비정형)과 토지피복면적(정형) 데이터를 융합해 활용하는 Multimodal AI 모델을 설계하여 인구, 지가, 건물연면적의 격자 단위 예측값을 산출하고, 이를 격자단위 통계 생성을 위한 초깃값으로 활용하였다. 이후 공간적 일관성을 유지하면서 실제 통계의 분포적 특성을 반영하기 위해 공간 조건 기반 생성형 모델 ScWGAN-GP(Spatially Conditioned WGAN-GP)을 새롭게 구축하였으며, 통계 이론에 기반한 분포 보정 기법인 Peening Process를 함께 제안하였다. 선행 방법론인 대시매트릭 매핑과 함께 세 분포 조정 알고리즘의 성능을 비교 분석한 결과, 서울특별시 격자인구 생성을 사례로 진행한 실험에서 ScWGAN-GP가 가장 낮은 예측 오차와 우수한 분포 재현 성능을 보였다. 또한, 실제 재정패널조사 및 사교육비조사를 활용한 전국 단위 실험을 통해 본 접근법의 일반화 가능성을 검증하였다. 이러한 결과는 격자 단위로 확장된 국가승인통계가 공공 및 민간 데이터 연계의 기반자료로 활용될 수 있음을 보여주며, 향후 증거기반행정과 지역 맞춤형 정책 수립에 기여할 수 있을 것으로 기대된다.

주요 용어 : Multimodal AI, ScWGAN-GP, 격자, 국가승인통계, 데이터증폭, 분포보정, 항공사진.

¹30019 세종특별자치시 세종로 2511, 고려대학교 일반대학원 경제통계학과 국가통계전공 박사.

E-mail: tozjs72@korea.ac.kr

²30019 세종특별자치시 세종로 2511, 고려대학교 빅데이터사이언스학부 교수. E-mail: korpen@korea.ac.kr

Comprehensive Overview of Real-World Artificial Intelligence Applications

*Jai Woo Lee*¹

Abstract

The advancement of Artificial Intelligence (AI) techniques has transformed the mechanisms of various industries. This review solidifies a thorough insight into applications and development of AI techniques, which can address possibilities and challenges appearing in a new policy. The diverse applications of AI in sectors such as business, education, healthcare, transportation, engineering, etc. can be investigated with AI techniques such as machine learning, big data, natural language processing, computer vision, signal processing. This study covers real-world examples underscoring and assessing the future possibilities for the extensive adoption of AI. Our research integrates recent findings to provide a thorough and detailed perspective on the potential advantages and obstacles of AI. These case studies with big data analysis emphasize the significant directions of AI technologies while discussing the ethical, societal, and economic issues. This study offers suggestions which can improve the available technology to optimize the potential of AI in promoting broad impacts and shareable benefits in the diverse system of industries.

Keywords : Artificial Intelligence, Big Data, Autonomous System, Algorithm, Industry.

¹고려대학교 빅데이터사이언스학과

Deep Learning for Censored Quantile Regression with an Adapted Loss Function^{*}

Doeun Lee¹, Sangbum Chor²

Abstract

In various real world datasets, censoring frequently occurs, and cases that require inference on the distributional characteristics of the data are no exception. Studies on quantile regression under such censoring situations have mainly been based on linear regression, but these approaches have limitations in adequately reflecting the nonlinear structure of real-world data. To overcome this, there have been several studies on nonlinear expansion using NN, but there are not many. Therefore, this study proposes a new model, ACQRNN (Adapted Censored Quantile Regression Neural Network), which extends the adapted loss function proposed by De Backer (2019) to a neural network by applying the MM algorithm. The proposed model can directly correct for censoring at the level of the loss function and enhances convergence stability and computational efficiency through the MM algorithm. The model performance is evaluated on nonlinear simulation data generated under varying censoring rates and feature dimensions. The proposed model is compared with existing studies, QuantReg, DeepQuantReg, and CQRNN, and using the evaluation metrics C-index, check loss, IPCW check loss, and censored D-calibration, to assess risk ranking ability, quantile prediction accuracy, and distributional calibration performance. Finally, the proposed model is applied to the failure data of HDD and SSD provided by backblaze, demonstrating its practical applicability in real-world reliability prediction.

Keywords : Quantile Regression, Neural Network, Survival Analysis, MM-algorithm.

^{*}This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (2022M3J6A1063595, 2022R1A2C1008514)

¹Master's Student, Department of the Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. Email: eundo0129@korea.ac.kr

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

Multimodal Deep Learning for Glaucoma Diagnosis using Fundus and Vascular Features^{*}

Hyemin Park¹, Xianchen Yan², Tae-Young Heo³

Abstract

Glaucoma is a visual disorder of the eye characterized by irreversible visual field defects due to damage to the optic nerve. Without timely intervention, glaucoma can lead to complete blindness, underscoring the critical need for early and accurate diagnosis. This study introduces an interpretable multimodal framework for glaucoma classification based on fundus imaging. The DeepLabV3+ model was employed to segment the optic disc, optic cup, and retinal vasculature, enabling the quantitative extraction of vascular network features, including the cup-to-disc ratio (CDR). These extracted features were subsequently integrated with the corresponding fundus images to construct a multimodal classification model capable of predicting the stage of glaucoma progression. To enhance the transparency of model decision-making, Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP) were applied to visualize and interpret the learned representations. Experimental evaluation demonstrated that the proposed multimodal approach achieved superior classification performance relative to image-only models, while simultaneously providing clinically meaningful interpretability.

Keywords : Glaucoma, Multimodality, Deep learning.

^{*}This research was supported by Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2024-00440787).

¹Master's student, Department of Information and Statistics, Chungbuk National University, Chungdae-ro 1, Cheongju-si, 28644, Chungbuk, Korea. E-mail: phm876503@gmail.com

²Master's student, Department of Information and Statistics, Chungbuk National University, Chungdae-ro 1, Cheongju-si, 28644, Chungbuk, Korea. E-mail: xyan240802@gmail.com

³(Corresponding author)Professor, Department of Information and Statistics, Chungbuk National University, Chungdae-ro 1, Cheongju-si, 28644, Chungbuk, Korea. E-mail: theo@cbnu.ac.kr

ResNet-BiGRU with Conditioned Query-Based Cross-Attention and Weighted Loss for Automated Chagas Disease Detection from 12-Lead ECG

*Taeyoung Kang¹, Hyuno Im², Nahyun Lee³, Taehwan Kim⁴, Donggun Kim⁵,
Donggyu Lee⁶, Seungang Oh⁷, Wuming Gong⁸, Il-Youp Kwak⁹*

Abstract

Our team CAUETUMN, a participating team in the 2025 PhysioNet/CinC Challenge, investigates whether integrating physiologically interpretable features with deep sequence representations, enhanced by conditioned Query-based cross-attention, can improve Chagas disease detection from standard 12-lead ECGs. We combine three information streams: (i) a 4-layer 1D ResNet backbone for local morphology extraction; (ii) a bidirectional GRU with gated attention for long-range temporal context; and (iii) handcrafted R-peak morphology feature and demographic features (age, sex). The auxiliary features are projected into a high-dimensional query space to conditionally attend over sequence embeddings, enabling selective integration of relevant temporal patterns. Raw ECGs undergo baseline-wander removal with an OC/CO morphological filter. The model is trained on a combination of SaMi-Trop, PTB-XL, and CODE-15\% dataset using a loss function that incorporates both class-specific weights to address label imbalance and group-specific weights to account for dataset-level distribution differences. Our final Challenge score on the hidden test set was 0.218, ranking 17th among the 41 eligible participating teams.

Keywords : ResNet-BiGRU, OC/CO morphological filter, Conditioned Query-Based Cross-Attention, Weighted Loss.

¹Undergraduate Student, Department of Applied Statistics, Chung-Ang University, Seoul 06974, Korea. E-mail: kangty0527@cau.ac.kr

²Undergraduate Student, Department of Applied Statistics, Chung-Ang University, Seoul 06974, Korea.

³Graduate Student, Department of Smart Cities, Chung-Ang University, Seoul 06974, Korea.

⁴Undergraduate Student, Department of Applied Statistics, Chung-Ang University, Seoul 06974, Korea.

⁵Undergraduate Student, Department of Mathematics, Korea University, Seoul 02841, Korea.

⁶Undergraduate Student, Department of Mathematics, Korea University, Seoul 02841, Korea.

⁷Professor, Department of Mathematics, Korea University, Seoul 02841, Korea.

⁸PhD, Department of Medicine, University of Minnesota, Minneapolis 55455, United States.

⁹(Corresponding Author) Associate Professor, Department of Applied Statistics, Chung-Ang University, Seoul 06974, Korea. E-mail: ikwak2@cau.ac.kr

CCTV 영상을 이용한 교통 사고 탐지 시스템*

김현주¹, 곽일엽²

요약

교통의 편의성이 증대함과 동시에 교통사고 발생 빈도 또한 증가함에 따라, 이에 대한 신속한 대응과 자동 탐지 기술의 중요성이 부각되고 있습니다. 본 연구에서는 사전 학습된 비전 트랜스포머(ViT) 기반의 영상 분류 모델을 활용하여 교통사고 탐지를 고도화하는 방법을 제안합니다. 기존의 교통사고 분류 모델들은 주로 단일 이미지나 영상의 일부 프레임만을 사용하여 시공간적 정보 손실이 발생하였으나, 제안된 모델은 전체 영상 시퀀스를 입력으로 사용함으로써 시공간적 특징을 효과적으로 반영합니다. 본 모델은 Kinetics-400 데이터셋으로 사전 학습된 ViT 백본을 기반으로 하며, 동영상 이해에 순수 트랜스포머 기반 모델을 성공적으로 적용한 첫 번째 사례 중 하나입니다. CADP와 UTIC에서 제공하는 실제 국내 CCTV 영상들을 사용하여 실험한 결과, 우리의 접근 방식이 우수한 분류 성능을 달성한다는 것을 보였습니다. 또한 UCF-Crime Dataset을 통한 일반화 성능 평가 역시 좋은 결과를 보였습니다. 본 연구는 교통사고 분류 문제에서 시공간적 정보의 통합 표현 학습이 성능 향상에 기여함을 실증하였으며, 향후 지능형 교통 시스템 및 실시간 사고 대응 시스템 개발에 기초 연구로 활용될 수 있습니다.

Keywords : Vision Transformer, Car Accident Detection, Deep Learning, Spatio-Temporal Modeling, Video Classification.

*본 논문은 한국대학교의 지원을 받아서 연구된 것임(2024a12345-7).

¹06974 서울특별시 동작구 흑석로 84, 중앙대학교 일반대학원. khyeonju19@gmail.com

²06974 서울특별시 동작구 흑석로 84, 중앙대학교 응용통계학과 부교수. ilyoup.kwak@gmail.com

Transformer 기반 모델을 이용한 낚시성 기사 탐지 및 대안 제목 생성 연구*

문수연¹, 한상태²

요약

낚시성 기사는 기사 본문과 달리 기사 제목을 과장하거나 자극적으로 구성하여 독자들의 호기심을 유발한다. 하지만 이는 독자들이 언론을 신뢰하지 않는 결과를 초래하므로 본 연구는 낚시성 기사를 탐지하고, 신뢰할 수 있는 제목을 생성하는 모델을 구축하고자 한다. 낚시성 기사 탐지를 위해 사전 학습된 BERT(Bidirectional Encoder Representations from Transformers)와 Sentence-BERT를 기반으로 기사 제목과 본문의 문맥적 연관성 분석하여 이를 기반으로 분류 모델을 파인튜닝 하였다. 또한, 사전학습된 T5(Text-to-Test Transfer Transformer)와 BART(Bidirectional and Auto-Regressive Transformers)를 기반으로 기사의 본문에서 핵심 내용을 한 문장 단위로 요약하도록 파인튜닝 하였고, 이를 정제된 기사 제목으로 간주하였다. 이후 모델들의 성능을 비교 분석하여 각 과제별 최적의 모델을 채택하였으며, 이를 통해 낚시성 기사로 인한 독자의 혼란을 줄이고 뉴스 소비 과정의 신뢰성과 효율성을 높이는데 기여하고자 한다.

주요 용어: 낚시성 기사 탐지, 제목 생성, 사전학습 모델, 뉴스 신뢰성

*이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (RS-2022-NR068754). 이 논문은 제1저자 문수연의 석사학위논문의 축약본입니다.

¹31499 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 데이터사이언스학과, 석사과정.

E-mail: suyeon_24@naver.com

²(교신저자) 31499 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 빅데이터AI학부, 교수.

E-mail: sthan@hoseo.edu

Hierarchical Extension of Latent Space Item Response Model^{*}

Soyeon Park¹, Seoyoung Shin¹, Minjeong Jeon², Hyoun Kyoung Kim³, Ick Hoon Jin⁴

Abstract

In item response data collection, purposive sampling from multiple upper-level groups is preferred to ensure representativeness and enable between-group comparisons. Such data naturally exhibits upper-level tendencies as respondents share both individual circumstances and common contextual environments. This study proposes a Hierarchical Latent Space Item Response Model (HLSIRM) for minimizing distortions in latent relationship inference and achieving practical representativeness across groups. The approach jointly models individual response patterns and group-level effects through direct correlation reflection in parameter estimation. By confining hierarchical structures to respondents, while preserving measurement invariance for items, the model facilitates diverse visual comparisons of higher-level interaction effects within a common low-dimensional latent space. Empirical analysis demonstrates that large-enrollment urban schools exhibit strong associations with happiness and positive peer relationships, while rural schools show closer alignment with counseling experiences and reduced anxiety. These school-level item interactions facilitate comprehensive policy design for mental health improvement. Strong external validity through consistent alignment between school latent vector positions and objective institutional characteristics, including enrollment size, regional location, and educational program emphasis. Additionally, model validation confirm the model's effective reproduction of diverse individual response patterns with superior performance.

Keywords : Item response data, Latent space model, Network model, Multiplicative model, Bayesian inference.

^{*}This work was supported by the National Research Foundation of Korea [grant number NRF 2020R1A2C1A01009881, RS-2023-00217705, and RS-2024-00333701; Basic Science Research Program awarded to IHJ].

¹Department of Statistics and Data Science, Yonsei University, Seoul, 03722, Korea.
Email: soyeon_stat@yonsei.ac.kr

²School of Education and Information Studies, University of California, Los Angeles, CA 90095, USA.
Email: mjjeon@ucla.edu

³Department of Child and Family Studies, Yonsei University, Seoul, 03722, South Korea.
Email: hyounkim@yonsei.ac.kr

⁴(Corresponding Author) Department of Applied Statistics, Yonsei University, Seoul, 03722, South Korea.
Email: ijin@yonsei.ac.kr

Interaction-Only Q-learning: A Simplified and Robust Framework for Optimal Dynamic Treatment Regimes^{*}

*Susie Song*¹

Abstract

With the growing importance of precision medicine, Dynamic Treatment Regimes (DTRs) have emerged as a key framework for deriving individualized, stage-specific optimal treatment decisions that adapt to patients' evolving conditions. Conventional Q-learning estimates optimal policies by sequentially modeling expected rewards; however, when the main effect is misspecified, bias can arise in estimating treatment-covariate interactions, leading to unstable decision rules. This study focuses on a two-stage DTR and proposes an Interaction-only Q-learning (IQL) model that removes the main effect and considers only the interaction term directly related to treatment decisions. The proposed model secures estimation stability and computational efficiency through LASSO regularization, incorporates A-learning to ensure causal validity in observational data, and extends to a GLM framework for non-Gaussian outcomes. Simulation results show that IQL reduces bias, variance, and mean squared error (MSE) compared with conventional Q-learning, and achieves comparable or even higher policy accuracy in high-dimensional settings. This study provides a theoretically and practically robust framework for stable and efficient optimal policy estimation, contributing to applications in precision medicine.

Keywords : Dynamic Treatment Regime, Q-learning, Interaction model, Causal inference, Regularized regression.

¹Master's Student, Department of Statistics, College of Political Science and Economics, Korea University, 21-1, Wolgok-ro, Seongbuk-gu, Seoul, 02800, Republic of Korea. E-mail: smi05095@naver.com

Bootstrap based score tests for overdispersion in a bivariate zero-inflated negative binomial regression mode^{*}

*Jieun Shin*¹, *Dongmin Jang*¹, *Byoungcheol Jung*²

Abstract

This study proposes three score tests for dispersion parameters in a generalized bivariate zero-inflated negative binomial (GBZINB) regression model. One advantage of the GBZINB regression model is that it enables a more flexible structure through the joint modeling of the two marginal response variables. We introduce a joint test, a conditional test, and a test of equality of two dispersion parameters, and derive score statistics for each. In Monte Carlo studies, the joint and conditional tests slightly underestimated the nominal significance level, whereas the equality test tended to overestimate it. The underestimation has been consistently observed in numerous studies on one-sided score tests. To address this, a parametric bootstrap procedure was additionally applied and it maintained the nominal significance level. Finally, the proposed tests were applied to real data from the Australian Health Survey.

Keywords : bootstrap; bivariate zero-inflated count data; overdispersion; score test

^{*}This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2022-NR068754).

¹Department of Statistics and Data Science, University of Seoul, Seoulsiripdaero 163, Seoul, 02504, South Korea.

²Department of Statistics, University of Seoul, Seoulsiripdaero 163, Seoul, 02504, South Korea

Prediction of Scale Score Based on Item Reponse Theory - in Case of Two Heterogeneous Groups^{*}

Jaceun Kim¹, Arlene K. H. Kim²

Abstract

In standardized examinations such as the CSAT, scale scores are reported alongside raw scores to provide interpretable performance information. When a subgroup of examinees with substantially higher ability than the overall population takes an assessment administered only to that group, it becomes useful to predict the scale scores that would have been obtained had the full population taken the same test. A commonly used approach under the random groups design is equipercentile equating, which matches raw scores with equal cumulative proportions across adjusted score distributions. While equipercentile equating performs well when the score distributions of the two tests are similar, its accuracy deteriorates severely when the tests differ greatly in difficulty. To address this limitation, we propose two alternative methods based on Item Response Theory (IRT) and the summed-score-based expected a posteriori estimator (sEAP). We treat scale scores as linear transformations of sEAP estimates and incorporate item scores when computing them. A key observation is that sEAP estimates exhibit an approximately semilinear relationship with scale scores when insignificant values are appropriately trimmed. Leveraging this structure, we construct linear and pointwise transformations that map the subgroup's sEAP values to the full population's scale scores, using anchor tests administered by KICE (Korea Institute for Curriculum and Evaluation). These transformations can then be applied to tests taken only by the subgroup. Simulation studies and analyses based on pseudo-CSAT mock exams show that the proposed methods produce lower root mean square errors in scale score prediction than equipercentile equating, particularly for high-performing examinees and tests with substantial difficulty gaps. Therefore, our approaches offer a more reliable alternative in settings where traditional equipercentile equating is prone to failure.

Keywords : Item response theory, Test equating, Scale score, CSAT.

^{*}This work is supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (MSIT) (No. RS-2023-NR076465).

¹Department of the Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea.
E-mail: eo1428@korea.ac.kr

²(Corresponding Author)Associate Professor, Department of the Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02481, Korea. E-mail: arlenent@korea.ac.kr

Hybrid Distance-Based Clustering and Dirichlet Process Mixture Modeling for Mass Spectral Analysis

손민수¹, 김재오¹

요 약

가스 크로마토그래피-질량 분석(GC-MS)에서 질량 스펙트럼 분석은 종종 측정 잡음과 기존의 유사성 기반 라이브러리 검색의 고유한 한계로 인해 정확도와 재현성이 저하되는 경우가 많습니다. 이러한 문제를 해결하기 위해, 본 연구는 스펙트럼 분석의 견고성과 확장성을 공동으로 향상시키는 두 가지 보완적인 접근 방식을 제안합니다: 거리 기반 방법과 비모수 베이저안 모델링 프레임워크입니다. 먼저, 각 질량 스펙트럼을 이산 확률 분포로 정규화하고, Jensen-Shannon divergence (JSD)과 코사인 기반 거리(CBD)를 결합하여 하이브리드 거리를 도입합니다. 이 측정은 다양한 잡음 조건에서 안정적인 성능을 제공하며 화학적으로 일관된 클러스터를 형성합니다. 이는 후속 라이브러리 검색에서 후보 선택을 안정화함으로써 기존 NIST 방법에 비해 1단계 후보 필터링의 신뢰성을 향상시킵니다. 둘째, 계층적 클러스터링의 고정 클러스터 한계를 극복하기 위해 다항-디리클레 디리클레 프로세스 혼합 모델(DPMM)을 사용합니다. DPMM은 15,973개의 스펙트럼에서 약 6,230개의 화학적으로 일관된 클러스터를 자동으로 추론하고, 새로 추가된 100개의 화합물을 새로운 클러스터로 성공적으로 분류하여 새로운 화합물 검출 능력을 입증했습니다. 이러한 접근 방식은 데이터 수준에서 노이즈 강건성을 보장하고 모델 수준에서 적응형 모델링을 통해 GC-MS 스펙트럼 해석을 향상시키는 통합 프레임워크를 구축합니다.

주요 용어: 질량 스펙트럼, 클러스터링, 하이브리드 거리, 디리클레 프로세스 혼합 모델.

Age-grouped analysis of influenza transmission influence to predict influenza spread based on mathematical model*

*Jeonghyeon Kim^{1,a}, Jeongseok Kim^{1,b}, Jinsol Kim^{1,c}, Sihyeon Ann^{1,d},
Minsoo Kim², Chankeun Park³*

Abstract

During the pandemic, the reproduction number became a critical metric for monitoring the progression of respiratory infections. The strong correlation between the transmission of respiratory diseases and the reproduction number carries significant implications for public health, particularly in forecasting outbreaks and developing control strategies. Due to its dynamic nature, accurately estimating the temporal fluctuations of the effective reproduction number in real time is a vital and extensively studied challenge within infectious disease epidemiology. Recent advancements in infectious disease forecasting research include efforts to determine age-specific reproduction numbers to better understand influenza transmission dynamics, underscoring the importance of age-structured modeling approaches. In this study, we present a novel methodology for estimating age-specific reproduction numbers. We have devised and formalized an age-stratified compartment model based on ordinary differential equations (ODEs), which leverages longitudinal infectious disease data from outbreak initiation to conclusion, along with empirically derived social contact patterns. Applying this model to the demographic context of South Korea during the 2017-2018 influenza season, our approach demonstrates enhanced effectiveness in representing influenza transmission behaviors. The findings from this study enable the quantitative forecasting of influenza spread across various age groups while accounting for heterogeneity and nonlinear interactions. Furthermore, we propose strategic measures for planning preventive interventions aimed at mitigating the transmission during outbreaks, thereby contributing to more effective epidemic response strategies.

*This work was supported by the National Korea Maritime & Ocean University Research Fund in 2025. Jeonghyeon Kim, Jeongseok Kim, Jinsol Kim, and Sihyeon Ann contributed equally to this work.

¹Student, Department of the Electronics Communications Engineering, National Maritime and Ocean University, Taejong-ro, Yeongdo-gu, Busan, 49112, Korea. E-mail: ^adnswjs35@naver.com, ^bkjs040130@gmail.com, ^cgofhfh42@gmail.com, ^dsshhy028@gmail.com

²(Corresponding Author) Assistant Professor, Department of the Electronics Communications Engineering, National Maritime and Ocean University, Taejong-ro, Yeongdo-gu, Busan, 49112, Korea.
E-mail: hikims@kmou.ac.kr

³(Corresponding Author) Professor, Department of the Electronics Communications Engineering, National Maritime and Ocean University, Taejong-ro, Yeongdo-gu, Busan, 49112, Korea.
E-mail: chapark@kmou.ac.kr

Keywords : Age-group, Influenza, Particle filter / smoother, Respiratory infectious disease, Epidemiology, Mathematical modeling

Reference

- Apolloni, A., Poletto, C., Colizza, V. (2013). Age-specific contacts and travel patterns in the spatial spread of 2009 H1N1 influenza pandemic, *BMC infectious diseases*, 13, 1-18.
- Cori, A., Ferguson, N. M., Fraser, C., Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics, *American journal of epidemiology*, 178(9), 1505-1512.
- Fraser, C. (2007). Estimating individual and household reproduction numbers in an emerging epidemic, *PloS one*, 2(8), e758.
- Gostic, K. M., McGough, L., Baskerville, E. B., Abbott, S., Joshi, K., Tedijanto, C. K. (2020). Practical considerations for measuring the effective reproductive number, R_t , *PLoS computational biology*, 16(2), 21008409.
- Son, W. S., Chae, M. K., Hwang, D. U., Nah, K. A., Kim, M. S., Kim, J. H., Lee, J. G. (2025). Social contact patterns in South Korea: an analysis of a survey conducted in 2023-2024, *BMC Infectious Diseases*, 25(1), 295.

Graph Neural Network - Based 2D Contact Map Learning for Protein - Ligand Binding and Interaction Prediction

Yoonseong An¹, IlsukKang¹, HyoJeCho²

Abstract

The precise prediction of protein - ligand interactions is a fundamental challenge and a key bottleneck in structure-based drug discovery. Although deep learning methodologies have significantly advanced this field, many computational models predominantly focus on predicting binding affinity, often overlooking the detailed structural geometry and intrinsic interaction modes essential for successful binding. Drawing inspiration from the effective use of two-dimensional (2D) contact maps in protein tertiary structure prediction, we propose a novel approach that conceptualizes the entire protein - ligand complex as a unified 2D contact map. This map is crafted to encode not only intra-molecular relationships (within the protein and within the ligand) but also, critically, inter-molecular contacts that capture the precise spatial proximity and types of chemical and physical interactions between protein residues and ligand atoms. We implement a graph neural network (GNN) architecture to extract rich, deep feature representations directly from this structured 2D contact map. GNN is capable of analyzing the complex, distance-based and interaction-based relationships encapsulated in the map to accurately predict binding possibilities and recognize the underlying interaction fingerprints. Our model is trained using experimentally resolved three-dimensional complex structures sourced from the Protein Data Bank (PDB). This GNN-based methodology provides a powerful and efficient tool for virtual screening and the rapid evaluation of drug candidates against novel protein structures. Its robust predictive capabilities make it highly applicable to de novo drug design and drug repositioning, assisting in the discovery of new targets for existing compounds and establishing a valuable data resource for studying the structural changes and dynamics that occur upon protein - ligand binding.

¹Department of Information Statistics, Chungbuk National University, Cheongju, Republic of Korea.

²Department of Biochemistry, Chungbuk National University, Cheongju, Korea.

Big data-driven genetic and epigenetic biomarkers in aging

*Nan Song*¹

Abstract

Aging is a heterogeneous process characterized by individual variation, influenced by genetic, environmental, and lifestyle factors. The risk of developing age-related diseases is also not uniform among people of the same chronological age. Different aging biomarkers can be considered to objectively measure biological age and quantify the rate of biological aging for an individual within a population. In the current study, we curated aging biomarkers, including leukocyte telomere length (LTL), DNA methylation-based epigenetic age (EA), and clonal hematopoiesis of indeterminate potential (CHIP), in human genome and epigenome datasets. We investigated that accelerated biological aging was associated with chronological age, sex, race, specific clinical and treatment exposures, unfavorable health behaviors, and health-related outcomes (quality of life and diseases). To better understand inter-individual variability, we investigated the genetic basis underlying aging acceleration and identified novel genetic susceptibility loci. However, a major limitation of existing aging biomarkers is the presence of ethnic and sociodemographic discrepancies, which may contribute to inaccurate prediction results, particularly in East Asian populations. To overcome this limitation and provide accurate age predictions for East Asian populations, we developed a novel neural network-based epigenetic aging clock, trained on Korean DNA methylation data and subsequently validated and potentially expanded for use in other East Asian populations. By rigorously investigating aging biomarkers (LTL, EA, EAA, and CHIP) and potential mechanisms underpinning accelerated aging, related exposures, and health outcomes, we will advance the field of aging research. These aging biomarkers have great potential to be used in risk prediction models to identify high-risk individuals and as targets for future intervention strategies in the general population beyond vulnerable populations.

¹Associate professor, College of Pharmacy, Chungbuk National University 194-21, Osongsaengmyeong 1-ro, Heungdeok-gu, Cheongju, Korea. E-mail: nan.song@chungbuk.ac.kr

병원 의료데이터를 활용한 심뇌혈관질환 진단 AI 기술

박승¹

요약

심뇌혈관질환은 응급 및 중환자 진료에서 분 단위 의사결정을 요구하며, 초기 증상의 비특이성과 다질환 공존, 시간적으로 비동기화된 다중 모달 정보로 인해 진단 지연과 치료 결정의 편차가 발생하기 쉽다. 이러한 한계를 줄이기 위해 전자의무기록, 검사실 지표, 시계열 생체신호, 영상 판독 정보를 체계적으로 통합하여 조기 위험도와 중증도를 정량화하는 인공지능 기반 진단 보조 기술이 필요하다. 본 연구에서는 충북대학교병원 실제 환자 데이터를 비식별화하여 후향적 코호트를 구축하고, 동일 환자 내 기록을 연계하였다. 또한, 의료진이 본 연구에 참여하여 추적 관찰 결과, 확정 진단, 영상·심전도 판독, 치료 반응에 대한 라벨링 데이터를 구축하였다. 해당 라벨을 기준으로 구조화 임상변수, 검사실 지표, 시계열 생체신호, 선택적 영상 파생 지표를 통합하여 로지스틱 회귀, 트리 기반 앙상블, LSTM, Transformer 계열 모델을 학습하였으며, 민감도, 특이도, AUROC을 주요 지표로 삼아 임상적 판별력을 확인하였다. 초기 결과는 단일 모달 대비 다중 모달 통합이 분류 성능을 일관되게 향상하고, 특히 응급실 초기 변수와 단기 생체신호 패턴의 결합이 조기 위험도 증화에 기여함을 보여주었다. 본 모델은 AI 모델이 임상 의사결정 지원 도구로 활용될 가능성을 제시하며, 향후 전향적 외부 검증과 실제 워크플로 파일럿 적용, 성능·안전성·형평성 모니터링을 통해 임상 적용성을 고도화하고자 한다.

¹충북대학교 의과대학 의학과

Scaling Up ROC-Optimizing Support Vector Machines

Gimun Bae¹, Seung Jun Shin²

Abstract

The ROC-SVM, introduced by Rakotomamonjy [2004], directly maximizes the area under the ROC curve (AUC) and is widely used for imbalanced classification. A major drawback, however, is its computational cost: training requires evaluating all $O(n^2)$ terms, which rapidly becomes infeasible for large samples. To reduce this burden, we propose a scalable algorithm based on incomplete U-statistics, which reduces the model complexity from $O(n^2)$ to $O(n)$. We further extend the method to nonlinear classification by adopting a low-rank approximation for efficient computation of the kernel matrix. As a result, training time decreases from hundreds of seconds to just a few seconds without significant loss in predictive performance, thereby making the model feasible for large-scale datasets.

Keywords : Machine Learning, Large-scale learning, Imbalanced Binary Classification, ROC Curve.

1. Introduction

In imbalanced binary classification, accuracy is often a misleading measure of performance and a common alternative is to evaluate classifiers through the receiver operating characteristic (ROC) curve. Motivated by this, Rakotomamonjy [2004] proposed ROC-SVM, which directly maximizes the area under the ROC curve (AUC) under a hinge loss.

Given data $((x_i, y_i))_{i=1}^n$ with $y_i \in \{+1, -1\}$, let $I_+ = \{i : y_i = +1\}$ and $I_- = \{i : y_i = -1\}$ denote the index sets of positive and negative samples, respectively, with sizes n_+ and n_- . Assuming the decision function $f \in H_K$, an RKHS, the ROC-SVM solves the following problem:

$$\min_{f \in H_K} \frac{1}{n_+ n_-} \sum_{i \in I_+} \sum_{j \in I_-} [1 - (f(x_i) - f(x_j))]_+ + \frac{\lambda}{2} \|f\|_{H_K}^2 \quad (1.1)$$

where $[z]_+ = \min(z, 0)$ and $\lambda > 0$ is a regularization parameter. We remark that the ROC-SVM (1.1) is computationally intensive as it requires evaluating $n_+ \times n_-$ terms, leading to a rapid

¹Master's Student, Department of Statistics, College of Political Science and Economics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: bgd5517@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: sjshin@korea.ac.kr

increase in computational cost as n increases. In this paper, we propose an efficient algorithm to solve (1.1). The empirical risk in ROC-SVM can be interpreted as a (generalized) U-statistic, and we leverage the incomplete U-statistic approach, proposed by Cléménçon et al. [2016], which reduces the burden without significant loss in predictive performance. While the linear ROC-SVM benefits from the approach, its kernel version remains intensive due to the large kernel matrix, so we employ a low-rank approximation to reduce storage, still preserving the statistical accuracy.

2. ROC-SVM as an Incomplete U-statistic

2.1. Generalized and incomplete U-statistics

We note that the loss of ROC-SVM (1.1) is the generalized U-statistic defined as follows.

Definition 1. (Generalized U-statistic)

Suppose we have J independent population, and we let $x_{i,j}$ denote the i -th i.i.d observations from the the j -th population where $i=1,2,\dots,n_j$ and $j=1,2,\dots,J$. Let A_j denotes all possible combination of index $(1,2,\dots,n_j)$ of a given size d_j , where $j=1,2,\dots,J$. Then, the generalized U-statistic with a kernel function $H: R^p \times \dots \times R^p \rightarrow R$, is defined as

$$U_n = \frac{1}{\prod_{j=1}^J \binom{n_j}{d_j}} \sum_{I_1 \in A_1} \dots \sum_{I_J \in A_J} H(X_{I_1,1}, \dots, X_{I_J,J}) \quad (2.1)$$

where $X_{I_j,j=x_i,j|i \in I_j}$ for a given index set $I_j = (i_{j1}, \dots, i_{jd_j}) \subset (1,2,\dots,n_j)$ for the j -th population, $j=1,2,\dots,J$, and $n = \sum_{j=1}^J n_j$.

Note that the generalized U-statistic involves sum of $\prod_{j=1}^J \binom{n_j}{d_j}$ terms, which becomes intractable as n_j gets large. This motivates the use of incomplete (generalized) U-statistics, which reduce computational cost while preserving statistical efficiency.

Definition 2. (Incomplete Generalized U-statistic)

Let Λ denote the set of all index tuples (I_1, \dots, I_J) , $\forall I_j \in \Lambda_j$, $j=1,2,\dots,J$. Given a randomly chosen subset $D_B \subseteq \Lambda$ of size B , the incomplete version of the generalized U-statistic (2.1) is defined as

$$\tilde{U}_n = \frac{1}{B} \sum_{(I_1, \dots, I_J) \in D_B} H(X_{I_1,1}, \dots, X_{I_J,J}) \quad (2.2)$$

The cardinality of D_B , B is much smaller than that of Λ . By doing that, we can efficiently approximate U_n by \tilde{U}_n while maintaining a desired level of accuracy. Notably, \tilde{U}_n

achieves the same convergence rate bound as U_n if D_B is randomly chosen index tuples from Λ with replacement and $B = O(n)$ (Clemençon et al. [2016]).

2.2. Application to Linear ROC-SVM

Suppose the classification function is linear, i.e., $f(x) = \alpha + \beta^T x$. Then the incomplete ROC-SVM solves

$$\min_{\beta} \frac{1}{B} \sum_{(i,j) \in D_B} [1 - \beta^T(x_i - x_j)]_+ + \frac{\lambda}{2} \beta^T \beta \quad (2.3)$$

where D_B a randomly selected from $\Lambda = ((i,j) | i \in I_+, j \in I_-)$ with replacement. We also set $B = n$, which is the simplest choice satisfying the condition $B = O(n)$ in Section 2.1..

3. Scaling-up Kernel ROC-SVM

3.1. Low-Rank Linearization of the Kernel ROC-SVM

By Representer Theorem, the solution of (1.1) can be expressed in the finite form: $f(x) = \alpha + \sum_{k=1}^n \theta_k K(x, x_k)$. Substituting this $f(x)$ into (1.1), the kernel ROC-SVM solves

$$\min_{\theta} \frac{1}{n_+ n_-} \sum_{i \in I_+} \sum_{j \in I_-} [1 - \theta^T(k_i - k_j)]_+ + \frac{\lambda}{2} \theta^T K \theta \quad (3.1)$$

where $\theta = (\theta_1, \theta_2, \dots, \theta_n)^T$, K is the kernel matrix whose (l, m) -th element is $K(x_l, x_m)$, and k_i is the i -th column of K .

In spite of the use of incomplete U-statistic, solving (3.1) is not trivial due to the kernel matrix K , which requires $O(n^2)$ flops to compute. To address this issue, we employed the idea of low-rank linearization (Zhang et al. [2012]). Toward this, one can first transform (3.1) into a linear ROC-SVM that is equivalent but simpler, as shown in proposition 1.

Proposition 1. For a given kernel matrix K , suppose we have $K = VV^T$, (3.1) is equivalent to solve the following linear ROC-SVM

$$\min_{\beta} \frac{1}{n_+ n_-} \sum_{i \in I_+} \sum_{j \in I_-} [1 - \beta^T(v_i - v_j)]_+ + \frac{\lambda}{2} \beta^T \beta$$

where v_i denotes the i -th row of V .

A straightforward solution of V can be obtained from the eigen decomposition of K . However, this approach still requires computing K , and the resulting V offers no computational advantage. Thus, we adopt the Nyström approximation (Williams and Seeger. [2000]) to obtain $\tilde{V} \in R^{n \times d}$, a low-rank approximation of V by randomly selecting $d (\ll n)$ samples from $x_i, i = 1, 2, 3, \dots, n$. Unless d is chosen excessively large, this becomes far more efficient.

Finally, the kernel ROC-SVM (3.1) can be approximated by the following linear ROC-SVM.

$$\min_{\beta} \frac{1}{B} \sum_{(i,j) \in D_B} [1 - \beta^T(\tilde{v}_i - \tilde{v}_j)]_+ + \frac{\lambda}{2} \beta^T \beta \quad (3.2)$$

where \tilde{v}_i denotes the i -th row of \tilde{V} .

4. Real Data Analysis

We evaluate the finite-sample performance of our approach through the Accelerometer Gyro Mobile Phone dataset from the UCI Machine Learning Repository (AlSahly, [2022]). This dataset contains 31,991 samples with 6 predictors and shows a class imbalance (98.2% positive, 1.8% negative). The data were randomly divided into training and test sets with an 8:2 ratio for 30 times. On each split, we fitted our proposed linear and kernel models, and recorded the training time and empirical AUC on the test sets. The regularization parameter λ was tuned over 30 values ranging from 10^{-6} to 10^4 , with the Gaussian kernel and the target rank fixed at $d=300$. In Table 1, the result obtained at the optimal λ are reported, with AUC values expressed as percentages and standard deviations shown without the 10^{-3} .

Table 1. Training time and test AUC on Accelerometer dataset : mean (sd)

Model	Time (s)	AUC (%)
Linear	3.54 (0.12)	88.53 (9.15)
Kernel	1.60 (0.38)	89.92 (21.61)

Both models train within a few seconds while achieving strong classification performance. In particular, the kernel model trains faster and outperforms the linear one in AUC. Overall, the results demonstrate that the proposed approach effectively handles imbalanced real-world data, achieving competitive performance within a reasonable training time.

5. Conclusion

In this work, we addressed the challenge of scaling ROC-SVM to large datasets. By leveraging incomplete U-statistics and a low-rank kernel approximation, we proposed a computationally efficient algorithm that reduces training time from hundreds of seconds to just a few seconds while maintaining predictive performance. Results on large-scale benchmark datasets demonstrate that the proposed framework enables efficient learning for large-scale imbalanced classification, a task traditionally limited by heavy computation.

References

- Rakotomamonjy A. (2004). Optimizing Area Under Roc Curve with SVMs, *In ROCAI*, 71 - 80.
- Clément, S., Colin I., Bellet A. (2016). Scaling-up Empirical Risk Minimization: Optimization of Incomplete U-statistics, *Journal of Machine Learning Research*, 17(76):1 - 36.
- Zhang K., Lan L., Wang Z., Moerchen F. (2012). Scaling up Kernel SVM on Limited Resources: A Low-rank Linearization Approach, *PMLR*, 1425-1434.
- Williams C., Seeger M. (2000). Using the Nyström Method to Speed Up Kernel Machines, *Advances in Neural Information Processing Systems*, 13.
- AlSahly A. (2022). Accelerometer Gyro Mobile Phone, UCI Machine Learning Repository,

A Mobile LiDAR-Based Deep Learning Approach for Real-Time 3D Body Measurement

YongHo Jeong^{1,}, TaeUk Noh^{2,*}, YongHak Lee^{3,a}, SeonJae Lee^{3,b},
Kwangil Choi^{3,c}, SuJin Jeong^{3,d}, SungHawn Kim⁴*

Abstract

In this study, we propose a solution for automatically measuring body circumferences by utilizing the built-in LiDAR sensor in mobile devices. Traditional body measurement methods mainly rely on 2D images or manual measurements. This research, however, utilizes 3D depth information to enhance both accuracy and efficiency. By employing HRNet-based keypoint detection and transfer learning through deep learning, the precise locations of body parts are identified and combined with depth maps to automatically calculate body circumferences. Experimental results demonstrate that the proposed method exhibits a relative error of up to 8% for major body parts such as waist, chest, hip, and buttock circumferences, with waist and buttock measurements recording low error rates below 4%. Although some models showed error rates of 7.8% and 7.4% in hip circumference measurements, this was attributed to the complexity of 3D structures and the challenges in selecting keypoint locations. Additionally, the use of depth map-based keypoint correction and regression analysis significantly improved accuracy compared to conventional 2D-based measurement methods. The real-time processing speed was also excellent, ensuring stable performance across various body types.

Keywords : LiDAR, HRNet, Deep Learning, Keypoint.

*These authors contributed equally to this work.

¹Department of Applied Statistics, Konkuk University, Seoul 05029, Korea.
E-mail: yonghojeong57@gmail.com

²Department of Applied Statistics, Konkuk University, Seoul 05029, Korea.
E-mail: ntw040307@gmail.com

³Department of Applied Statistics, Konkuk University, Seoul 05029, Korea.
E-mail: ^akayron@konkuk.ac.kr, ^blth9913@naver.com, ^ccgi2148@naver.com, ^dsujini7773@gmail.com

⁴(Corresponding Author) Department of Applied Statistics, Konkuk University, Seoul 05029, Korea.
E-mail: shkim1213@konkuk.ac.kr

Toward Full Automation of the Revised NIOSH Lifting Equation "Using Computer Vision"

SeHee Jung¹, Xu Xu²

Abstract

This study presents a computer vision-based method for automating the Revised NIOSH Lifting Equation (RNLE). Although the RNLE is a widely used tool for evaluating lifting-related biomechanical stressors, its reliance on manual measurement limits its accuracy, efficiency, and scalability. The proposed method adopts a three-stage process: (1) pose estimation to extract 3D joint coordinates, (2) lift timing identification using kinematic features, and (3) RNLE multiplier estimation to calculate the recommended weight limit (RWL). When applied to a benchmark video dataset with synchronized motion capture data, the method achieved a mean absolute RWL error of 1.58 kg with a correlation of 0.91. These results demonstrate the method's potential to automate the RNLE with high accuracy. In practice, the method enables RNLE implementation using widely available video platforms, such as mobile phones or surveillance cameras, offering a scalable, real-time ergonomic intervention to promote safer lifting practices in diverse workplaces.

Keywords : action recognition, low back injury, manual material handling, musculoskeletal disorder, risk assessment.

¹PhD Candidate, Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC, 27695, USA. E-mail: sjung9@ncsu.edu

²(Corresponding Author)Associate Professor, Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC, 27695, USA. E-mail: xxu@ncsu.edu

Cloud2CAD: AI-Driven High-Fidelity CAD Reconstruction of Industrial Components from 3D Point Clouds*

Junewoo Kang¹, Chaehwan Lee², Yonghak Lee³, Sunghwan Kim⁴, Seongmin Yang⁵

Abstract

This paper proposes and empirically validates, a pipeline for automatically reconstructing complex industrial components--such as impellers, casings, and shafts--from 3D point clouds into CAD models. The workflow integrates PCA-based alignment, PointNet-family semantic segmentation, and a hybrid surface-fitting module that combines RANSAC primitives with implicit neural representations (INRs), followed by topological reconstruction to generate a watertight B-rep model. Furthermore, a targeted training strategy is introduced to mitigate over-simplification on complex geometries. Evaluated on a dataset of real-world scans, the pipeline achieves manufacturing-grade precision while preserving free-form surfaces and maintaining topological consistency. This study validates the industrial viability of automated reverse engineering by bridging advanced academic techniques with practical applications and outlines future enhancements in robustness to imperfect data and overall processing efficiency.

Keywords : Reverse engineering, Point clouds, Computer aided design (CAD), Semantic segmentation, Signed distance functions.

*This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2023-00240936).

¹Undergraduate student, Department of Computer Information Engineering, Kwangwoon University, Seoul, Korea.

²Graduate student, Department of Data Science, Konkuk University, Seoul, Korea.

³Doctoral student, Department of Applied Statistics, Konkuk University, Seoul, Korea.

⁴Associate Professor, Department of Applied Statistics, Konkuk University, Seoul, Korea.

⁵(Corresponding Author) Doctoral student, Department of Applied Statistics, Konkuk University, Seoul, Korea.

Distributed Reduced-Rank Regression for Large-Scale Data

Hongshin Choi¹, Seyoung Park²

Abstract

Modern data analysis often involves processing large-scale, high-dimensional data distributed across multiple machines. In such settings, low-rank regression suffers from limitations in both statistical estimation and computational efficiency. Although convex nuclear-norm regularization offers computational convenience, it introduces bias for large singular values. Moreover, existing distributed optimization methods typically rely on global Lipschitz constants for step-size selection, which require iterative synchronization across nodes and lead to severe communication bottlenecks. To overcome these challenges, we propose a communication-efficient adaptive nuclear normalization method. The approach applies a first-order approximation of the nonconvex Smoothly Clipped Absolute Deviation (SCAD) penalty, transforming the problem into an iterative sequence of weighted nuclear-norm minimizations. In addition, it eliminates global aggregation at each iteration by replacing globally determined Lipschitz-based step sizes with locally computable adaptive updates using the Barzilai - Borwein rule. Extensive experiments demonstrate that the proposed method improves both prediction accuracy and estimation quality while matching or outperforming existing distributed nuclear-norm baselines, often with substantially fewer communication rounds and reduced transmission cost.

Keywords : Distributed estimation, Adaptive Nuclear Norm, SCAD, Low-rank regression.

¹Department of Statistics and Data Science, Yonsei University.

²(Corresponding Author) Department of Statistics and Data Science, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea. E-mail: ishpsy@yonsei.ac.kr

Adaptive Fused Quantile Regression with BAR Penalty

*Jihyun Eun*¹

Abstract

Quantile regression offers a comprehensive view of conditional distributions but can be inefficient in high-dimensional problems with structured covariates. We study two adaptive L2 penalties for structured quantile regression—L2-Fused (first-order differences) and L2-Pairwise (all pairwise differences)—and develop scalable ADMM solvers. The L2-Fused penalty promotes local smoothness by shrinking adjacent coefficient differences, making it suitable when the covariate order encodes structure (e.g., spline bases). The L2-Pairwise penalty enforces global grouping by shrinking all pairwise differences, driving coefficients toward nearly equal values. Both penalties are combined with adaptive BAR weights to reduce bias, and the tuning parameter is selected by BIC. Simulations and real data illustrate complementary strengths: L2-Fused excels at recovering smooth structural trends, whereas L2-Pairwise is effective at global coefficient grouping. These results show that adaptive L2 penalties offer a flexible and powerful framework for structured quantile regression.

Keywords : Quantile Regression, Fused Penalty, Adaptive Weights, ADMM, Smoothing, Grouping.

¹고려대학교 통계학과

Memorize Early, Then Query: Inlier-Memorization-Guided Active Outlier Detection

Minseo Kang¹, Seunghwan Park², Dongha Kim³

Abstract

Outlier detection (OD) aims to identify abnormal instances, known as outliers or anomalies, by learning typical patterns of normal data, or inliers. Performing OD under an unsupervised regime - without any information about anomalous instances in the training data - is challenging. A recently observed phenomenon, known as the inlier-memorization (IM) effect, where deep generative models (DGMs) tend to memorize inlier patterns during early training, provides a promising signal for distinguishing outliers. However, existing unsupervised approaches that rely solely on the IM effect still struggle when inliers and outliers are not well-separated or when outliers form dense clusters. To address these limitations, we incorporate active learning to selectively acquire informative labels, and propose IMBoost, a novel framework that explicitly reinforces the IM effect to improve outlier detection. Our method consists of two stages: 1) a warm-up phase that induces and promotes the IM effect, and 2) a polarization phase in which actively queried samples are used to maximize the discrepancy between inlier and outlier scores. In particular, we propose a novel query strategy and tailored loss function in the polarization phase to effectively identify informative samples and fully leverage the limited labeling budget. We provide a theoretical analysis showing that the IMBoost consistently decreases inlier risk while increasing outlier risk throughout training, thereby amplifying their separation. Extensive experiments on diverse benchmark datasets demonstrate that IMBoost not only significantly outperforms state-of-the-art active OD methods but also requires substantially less computational cost.

Keywords : Outlier Detection, Deep Generative Models, Active Learning, IM Effect.

¹Master's student, Department of Statistics, Sungshin Women's University, 2, Bomun-ro 34 Da-gil, Seongbuk-gu, Seoul, 02844, Republic of Korea. E-mail: krystalsms0306@gmail.com

²Associate Professor, Department of Information Statistics, Kangwon National University, 1, Gangwon-daehak-gil, Chuncheon-si, Gangwon, 24341, Republic of Korea.
E-mail: stat.shpark@kangwon.ac.kr

³(Corresponding Author) Assistant professor, School of Mathematics, Statistics and Data Science, Sungshin Women's University, 2, Bomun-ro 34 Da-gil, Seongbuk-gu, Seoul, 02844, Republic of Korea.
E-mail: dongha0718@sungshin.ac.kr

Sufficient KCI: Kernel-based Conditional Independence Test with Sufficient Dimension Reduction

*Chaehyun Ryu*¹, *Kyongwon Kim*²

Abstract

Conditional independence test between random variables is an important task in Bayesian networks and causal discovery. However, in high-dimensional settings, existing methods often suffer from the curse of dimensionality. To address this issue, we propose a new method, Sufficient KCI, which integrates nonlinear sufficient dimension reduction into the kernel-based nonparametric conditional independence test. By applying a generalized sliced inverse regression, our method effectively mitigates the curse of dimensionality while retaining the essential dependency structure among variables. Under the null hypothesis of conditional independence, we derive an appropriate test statistic based on kernel functions and establish its corresponding asymptotic distribution. Through extensive simulations and real data applications, we demonstrate that our proposed method achieves higher accuracy and efficiency compared to existing graphical model-based approaches, particularly when the data dimension is much larger than the sample size.

Keywords : graphical model, reproducing kernel Hilbert space, nonlinear sufficient dimension reduction, conditional independence test.

¹Department of Statistics, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul, 03760, Republic of Korea. Email: stella062@ewha.ac.kr

²(Corresponding author) Department of Applied Statistics, Department of Statistics and Data Science, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea.
Email: kimk@yonsei.ac.kr

Analyzing Classroom Interaction Data Using Prompt Engineering and Network Analysis^{*}

Gwanghee Kim¹, Ick hoon Jin², Minjeong Jeon³

Abstract

Classroom interactions play a vital role in developing critical thinking, collaborative problem-solving abilities, and enhanced learning outcomes. While analyzing these interactions is crucial for improving educational practices, the examination of classroom dialogues presents significant challenges due to the complexity and high-dimensionality of conversational data. This study presents an integrated framework that combines prompt engineering with network analysis to investigate classroom interactions comprehensively. Our approach automates utterance classification through prompt engineering, enabling efficient and scalable dialogue analysis without requiring pre-labeled datasets. The classified interactions are subsequently transformed into network representations, facilitating the analysis of classroom dynamics as structured social networks. To uncover complex interaction patterns and how underlying interaction structures relate to student learning, we utilize network mediation analysis. In this approach, latent interaction structures, derived from the additive and multiplicative effects network (AMEN) model that places students within a latent social space, act as mediators. In particular, we investigate how the gender gap in mathematics performance may be mediated by students' classroom interaction structures.

Keywords: Classroom Interaction, Prompt Engineering, Additive and Multiplicative Effect Network Model, Network Mediation Analysis.

^{*}This work was partially supported by the National Research Foundation of Korea [grant number NRF-2021S1A3A2A03088949, RS-2023-00217705, RS-2024-00333701; Basic Science Research Program awarded to IHJ], [grant number NRF-2025S1A5C3A0200632411; awarded to IHJ and MJ], International Joint Research Grant by Yonsei Graduate School [awarded to GK], and the ICAN (ICT Challenge and Advanced Network of HRD) support program [grant number RS-2023-00259934], supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation).

¹Department of Statistics and Data Science, Yonsei University. 50 Yonsei-ro, Seodaemun, Seoul, Republic of Korea. E-mail: musagh08@yonsei.ac.kr

²(Corresponding Author) Department of Applied Statistics, Yonsei University. 50 Yonsei-ro, Seodaemun, Seoul, Republic of Korea. E-mail: ijin@yonsei.ac.kr

³School of Education and Information Studies, University of California. Los Angeles. USA.
E-mail: mjjeon@ucla.edu

위성 데이터 기반 글로벌 산불 확산 예측 딥러닝 모델의 개발*

송영승¹, 신성훈², 김도윤³, 박혜민⁴, 허태영⁵

요약

최근 기후변화로 인한 고온·건조 현상의 심화로 전 세계적으로 대형 산불이 빈번하게 발생하고 있으며, 이에 따라 산불 확산을 사전에 예측하여 효율적으로 대응할 수 있는 기술의 필요성이 커지고 있다. 본 연구는 전 세계 어디서나 활용 가능한 위성자료(MODIS, VIIRS, ERA5-Land 등)를 이용하여 산불의 다음날 확산 양상을 예측하는 딥러닝 기반 모델을 제안한다. ResNet-UNet 구조를 기반으로 기상, 지형, 토지피복 등 21개 환경 변수를 통합하여 학습하였으며, AP 0.559를 달성하여 단순 지속성 기준모델(AP 0.287)을 크게 상회하였다. 제안한 모델은 국내 중·대규모 산불 사례에서 주요 확산 방향과 활성 화선을 올바르게 포착하며, 딥러닝 기반 산불 예측의 실질적 활용 가능성을 입증하였다. 더 나아가, 본 연구는 전 지구적 적용이 가능한 모델 구조를 기반으로, 다양한 국가와 기후대에서 발생한 대형 산불 데이터를 추가 확보하고 고빈도 위성 관측 자료를 통합함으로써, 안동 산불과 같은 초대형 산불에 대한 예측 성능을 강화하고자 한다. 이러한 확장은 국가 간 데이터 편차를 극복하고, 전 세계적으로 일반화 가능한 산불 확산 예측 모델 구축으로 이어질 것이다. 궁극적으로 본 연구는 인명·재산 피해를 최소화하기 위한 실시간 재난 대응 및 자원 배치 의사결정 지원 체계의 기반 기술로 발전할 것으로 기대된다.

주요 용어: 위성 데이터, 딥러닝, 산불 확산 예측, 글로벌 모델.

*이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (RS-2025-23525903).

¹충북 청주시 서원구 충대로1, 충북대학교 자연과학대학 정보통계학과 학부연구생. E-mail: world819@gmail.com

²충북 청주시 서원구 충대로1, 충북대학교 자연과학대학 정보통계학과 학부연구생. E-mail: ssh3537000@gmail.com

³충북 청주시 서원구 충대로1, 충북대학교 자연과학대학 정보통계학과 박사수료. E-mail: doyun2516@gmail.com

⁴충북 청주시 서원구 충대로1, 충북대학교 자연과학대학 정보통계학과 석사과정. E-mail: tphm876503@gmail.com

⁵(교신저자) 충북 청주시 서원구 충대로1, 충북대학교 자연과학대학 정보통계학과 교수.

E-mail: theo@chungbuk.ac.kr

대전 행사·축제 운영 효율성 분석 및 성과 개선을 위한 정책 제언: DEA-Tobit 2단계 접근법을 중심으로

김승연¹, 김서연¹, 문유정¹, 홍지윤¹

요약

본 연구는 2022년부터 2024년까지의 대전광역시 행사·축제 데이터를 기반으로 한다. 유사 축제 문제를 실증적으로 규명하고 운영 효율성을 정량 분석하여 실효성 있는 정책 개선 방안을 제시하는 것이 주된 목적이다. 먼저, 텍스트 마이닝을 통해 축제의 사업 목적을 7개 유형으로 분류한 결과, 예술문화·전통문화 등 특정 유형에 집중된 유사 축제가 상당수 존재함이 확인되었다. 이어서 DEA 산출지향 모형을 적용하여 195개 축제의 상대적 효율성을 측정하였으며, 전체 기술 효율성 평균은 2.220으로 나타나 현재 투입 대비 약 122%의 추가 산출이 필요한 높은 비효율성을 보였다. 특히 비효율성은 운영 효율성(PTE, 평균 1.46)과 규모 효율성(SE, 평균 1.47)이 복합적으로 작용한 결과로 분석되었다. 패널 토빗 분석을 통해 효율성 영향 요인을 규명한 결과, 재정 자립도와 주민 인구 규모가 운영 효율성을 유의하게 향상시키는 핵심 환경 요인으로 나타났다. 특히, 여름 축제는 운영 효율성(PTE)이 아닌 규모 효율성(SE) 저하로 인한 비효율성이 두드러졌다. 본 연구는 대전 축제의 유사성 문제를 데이터 기반으로 입증하고 축제별 효율성 수준 및 개선량을 제시했다. 이는 지역 간 운영 역량 격차 해소, 유형·계절 특성에 따른 규모 재설계, 벤치마킹 기반 축제 관리 등 실효성 있는 정책 시사점을 제공한다는 데 의의가 있다.

주요 용어: 대전광역시 축제, 자료포락분석(DEA), 효율성 분석, 패널토빗 모형.

Optimal Location Selection for Community Sports Facilities in Cheongju using GTF and MCLP

Dong-Young Lee¹, Ju-Seong Lee²

Abstract

Cheongju City faces a significant spatial imbalance in community sports facilities between its original downtown and newly developed areas like Osong. This study proposes a two-stage data-driven framework to identify optimal locations for new facilities. First, we use Graph Trend Filtering (GTF) on a 500m grid-based supply-demand imbalance index to identify both large-scale demand hotspots and localized hidden demand pockets. Second, a Maximal Covering Location Problem (MCLP) is applied to these potential areas to select 5 optimal sites that maximize population coverage. The model's ability to address both structural and localized imbalances is validated by the resulting locations, which include four sites in the underserved downtown hotspots and one in the Osong new town. This approach provides a strategic solution for balanced urban development.

Keywords : Optimal Location, Graph Trend Filtering (GTF), Maximal Covering Location Problem (MCLP), Spatial Data Analysis, Sports Facility, Cheongju.

¹Master's student, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea.
E-mail: dtsy5891@gmail.com

²Master's student, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea.
E-mail: ljs000603@gmail.com

무인점포 안전관리를 위한 Skeleton - Context 기반 실시간 이상행동 탐지 모델*

임지원¹, 정윤채²

요 약

전국적으로 무인점포가 빠르게 확산되면서 관련 범죄가 증가하고 있으며, 특히 대전·충청 지역에서는 절도, 파손, 흡연과 같은 무인점포 내 위해 행동이 높은 수준으로 보고되고 있다. 그러나 순찰과 사후 모니터링 중심의 대응은 사회적 비용 대비 실질적 효과가 낮은 한계를 보인다. 이에 본 연구는 신체 움직임 기반의 스켈레톤 정보와 장면 내 상황적 특징을 함께 활용하는 Skeleton - Context 융합 기반의 실시간 이상행동 탐지 모델을 제안한다. 연구에는 AI-Hub에서 제공하는 실내 사람 구매행동 데이터와 이상행동 CCTV 데이터를 활용하였다. 분석 과정은 YOLOv8-Pose를 이용해 얼굴을 제외한 17개의 관절 좌표를 추출해 핵심 움직임 정보를 확보하고, CTR-GCN을 통해 관절 간 구조적 관계와 시간적 패턴을 학습하는 방식으로 구성하였다. 여기에 절도 상황의 물건 이동, 흡연 시의 연기 등 장면 기반의 상황적 feature를 결합해 단순 관절 정보만으로는 설명하기 어려운 행동의 맥락을 보완하였다. 또한 실제 무인점포 환경의 변동성을 고려해 데이터 불균형 조정, 조명 변화와 흔들림을 반영한 노이즈 기반 증강, 다양한 행동 변형을 생성하는 Mix-up을 적용해 모델의 안정성과 일반화 성능을 강화하였다. 제안된 Skeleton - Context 융합 모델은 정상 및 이상행동을 안정적으로 구분하는 경향을 보였으며, 약 0.30ms의 추론 속도로 실시간 처리 요건을 충족해 다양한 무인점포 환경에서 활용 가능한 AI 기반 안전관리 체계로서의 잠재력을 확인하였다.

주요 용어 : 무인점포, 이상행동 탐지, Skeleton - Context 융합, YOLOv8-Pose, CTR-GCN.

¹34134 대한민국 대전광역시 유성구 대학로 99, 충남대학교 통계데이터사이언스학과 석사과정.

E-mail: jiwonlim0492@gmail.com

²34134 대한민국 대전광역시 유성구 대학로 99, 충남대학교 통계데이터사이언스학과 석사과정.

E-mail: bella55560@gmail.com

대전 관광의 편중 완화를 위한 여행 동선 추천 시스템

박진규¹, 이소민²

요 약

대전 지역의 관광 수요가 특정 장소와 동선에 집중되는 편중 현상이 지속적으로 지적되고 있으며, 이를 완화하기 위해 관광객의 실제 이동 기반 데이터를 활용한 정밀 분석의 중요성이 커지고 있다. 이러한 맥락에서 본 연구는 대전 지역 관광객 행동 데이터를 활용하여 관광객 행동 패턴을 체계적으로 파악하고, 편중 완화에 기여할 수 있는 데이터 기반 추천 체계를 구축하는 것을 목표로 한다. 이를 위해 사전소비, 이동정보, 방문지 정보 등 다양한 이종(異種) 데이터로 구성된 관광객 데이터셋을 정규화, 전처리하여 분석 목적에 맞게 재구성하였다. 초기 탐색적 데이터 분석을 통해 지역 관광 현황과 편중성을 파악한 뒤, 기존 분류 방식의 한계를 보완하기 위해 LDA(Latent Dirichlet Allocation) 기반 여행객 페르소나를 도출하였다. 또한 카카오맵·네이버 지도 등 외부 지도 기반 POI 데이터를 확장하여 로컬 관광지의 특성을 보강하고, 페르소나 간 이동 전이확률 행렬을 설계하여 POI 클러스터와 연계함으로써 실제 이동 맥락을 반영한 경로 기반 추천 로직을 고도화하였다. 이를 통해 숨겨진 지역 관광지를 적절히 노출·추천할 수 있는 개인화된 경로 추천이 가능하도록 구현하였다. 아울러 본 연구에서는 이러한 분석·추천 구조를 실제로 동작하는 형태로 시연 가능한 프로토타입 시스템으로 제작하여 적용 가능성과 실효성을 검증하였다. 본 연구는 대전 지역 관광 편중 문제를 완화하고 지역 기반 맞춤형 관광경로 추천 모델을 확장하는 데 기여할 수 있는 기반을 제시한다. 향후 외부 API 연계 확대와 사용자 로그 분석을 통해 페르소나를 더욱 정교화하고, 이동 시간·공간 제약을 반영한 최적화 모델을 도입함으로써 실용성을 강화할 계획이다

주요 용어: 관광 데이터 분석, 페르소나(LDA), POI 확장, 전이확률, 관광 추천 시스템.

¹ 34143 대한민국 대전광역시 유성구 대학로99 충남대학교 자연과학대 2호관 429호, 정보통계학과 학부과정, E-mail: dasapjg2@gmail.com

²34143 대한민국 대전광역시 유성구 대학로99 충남대학교 자연과학대 2호관 431호, 통계데이터사이언스학과 학부수료후 연구생. E-mail: 0813somin@naver.com

CNN과 LSTM 결합을 통한 작물 생산량 예측 정확도 향상 연구*

강소라¹, 나명환²

요약

농작물의 수확량 예측은 국가적 수급 조절과 농가의 경영 안정성 확보를 위해 매우 중요한 과제이다. 본 연구에서는 2021년부터 2023년까지 경상북도 지역의 단동 비닐하우스에서 토경제 배 방식으로 재배된 참외의 생육 및 환경 자료를 활용하여 수확량 예측 모델을 구축하였다. 대상 농가는 스마트팜 기반 농가로, ICT 환경제어 시스템을 통해 1작기(11월 정식~이듬해 7월 수확) 동안 온도, 습도, 이산화탄소(CO₂), 일사량 등 주요 생육 환경 요인을 수집하였다. 본 연구는 이러한 시계열 데이터를 효율적으로 분석하기 위해 1차원 합성곱 신경망(1D-Convolutional Neural Network, 1D-CNN)과 장단기 메모리 네트워크(Long Short-Term Memory, LSTM)를 결합한 하이브리드 구조를 제안하였다. 1D-CNN은 시계열 내 특정 시점에서 발생하는 지역적 패턴을 감지하고, 시간 축의 이동에 강인하게 반응하도록 설계되어 있다. 반면, LSTM은 시간의 흐름에 따른 정보의 연속성과 장기 의존성을 학습함으로써 과거 상태가 현재 및 미래 수확량에 미치는 영향을 반영한다. 두 모델의 결합을 통해 단기적 생육 반응과 장기적 성장 추세를 동시에 고려할 수 있으며, 이는 참외 수확량 예측의 정밀도를 향상시킬 수 있는 기반이 된다. 이러한 접근은 작물의 생육 환경 변화에 따른 반응을 정량적으로 모델링할 수 있는 가능성을 제시하며, 향후 다양한 작물 및 재배 조건으로의 확장 연구에도 기여할 것으로 기대된다.

주요 용어: 참외, 수확량 예측, 합성곱 신경망(Convolutional Neural Network, CNN), 장단기 메모리 네트워크(Long Short-Term Memory, LSTM).

*본 논문은 농촌진흥청 공동연구사업의 지원에 의해 이루어진 것임(RS-2025-02243084).

¹(제1저자) 61186, 광주광역시 북구 용봉동 용봉로 77, 전남대학교 수학과/통계학과 박사과정.

Email: sc12love@gmail.com

²(교신저자) 61186 광주광역시 북구 용봉로 77, 전남대학교 통계학과 교수. Email: nmh@chonnam.ac.kr

AutoEncoder 사전 학습 임베딩을 활용한 그래프 신경망 기반 다변량 시계열 이상 탐지

고경민¹

요약

산업 환경에서는 여러 센서를 통해 실시간으로 데이터를 수집하고, 이를 통해 조기에 이상을 탐지하여 운영 안정성과 품질을 유지한다. 다만 수집되는 데이터의 규모와 복잡성이 기술 발전에 따라 증가하고 있어 이와 관련된 이상 탐지 기술의 발전도 요구되고 있으며, 최근에는 다변량 시계열을 효과적으로 학습하는 그래프 신경망 기반의 이상 탐지 기술이 주목받고 있다. 그러나 그래프 신경망의 초기 학습 단계에서는 시계열 특성이 적절히 반영되지 않아 그래프 구조의 왜곡에 취약하다는 한계가 존재하며, 이를 개선하기 위해 본 연구에서는 사전 학습을 활용한 그래프 신경망 구조를 제안한다. 먼저 AutoEncoder를 통해 단변량 시계열을 사전 학습하고, 그래프 신경망의 초기 학습 단계에 AutoEncoder로부터 추출된 Latent Vector를 반영하여 그래프 구조의 품질과 학습 안정성이 개선되도록 하였다. 제시한 방안은 여러 벤치마크 데이터 세트에서 반복 실험하여 재현성을 검증하고 개선 효과와 우수성을 확인할 수 있었다.

주요 용어 : 그래프 신경망, AutoEncoder, 사전 학습, 이상 탐지.

1. 서론

현대 산업 환경은 4차 산업 혁명의 진전에 따라 데이터의 규모와 복잡성이 기하급수적으로 증가하고 있으며, 제조 공정/IT 인프라 등의 환경에서는 수백 개의 센서를 통해 설비의 상태나 자원 사용률 등의 데이터를 실시간으로 수집하여 여러 분야에 활용하고 있다. 특히 데이터로부터 비정상적인 동작을 검출하는 이상 탐지 기술은 신속한 대응을 통해 운영 안정성을 확보하고 품질 사고를 최소화할 수 있어 산업 환경에서 필수적으로 요구되는 기술 중 하나이다. 수집되는 시계열을 분석하기 위해 Box & Pierce(1970)가 제안한 통계 기반의 분석 모델인 ARIMA와 Ester et al.(1996)가 제시한 클러스터링 기반의 DBSCAN 등을 전통적으로 사용해왔으나, 기존 방안들은 현대 산업 환경에서 발생하는 다변량 시계열의 비선형성과 센서 간 상호작용 같은 복합적인 특성을 적절히 반영하지 못한다는 한계를 가진다. 이를 극복하기 위해 그래프 신경망을 활용해 시계열 간 관계성을 학습하는 방안이 제안되었으며 이 중 Deng & Hooi(2021)가 제시한 학습을 통해 그래프 구조를 모델링할 수 있는 Graph Deviation Network(GDN)은 여러 벤치마크에서 성능을 입증하고 많은 분야에서 활용되고 있는 모델이다. 그러나 GDN은 데이터에 대한 사전 정보 없이 학습을 진행하기 때문에 초기 학습 단계에서의 불안정성과 노이즈로 인한 그래프 구조의 왜곡에 취약하다는 단점이

¹02481 서울특별시 성북구 안암로 145 고려대학교 SW·AI융합대학원 빅데이터융합학과 석사과정.

E-mail: xcv3549@gmail.com

존재한다. 본 연구에서는 이를 개선하기 위해 각각의 단변량 시계열에 대해 **AutoEncoder** 로 사전 학습한 정보를 그래프 구조 모델링 과정에 반영하여 초기 그래프 품질 개선을 통해 학습 안정성을 확보하고 모델의 최종 성능을 개선하는 방안을 제안한다.

2. 본론

2.1. GDN의 구조 및 한계

GDN은 시계열을 그래프 구조로 변환하여 시계열 간 상호작용을 명시적으로 모델링 한 뒤 그래프 신경망을 통해 시계열 내의 복합적인 특성을 학습하는 방식으로, 효과적인 다변량 시계열 이상 탐지 방안으로 주목 받은 연구이다. **GDN**에서는 먼저 각각의 시계열을 특정 차원으로 임베딩하여 이를 그래프 구조의 노드로 정의하고 노드 간 유사도가 높은 경우에는 엣지를 생성한다. 이렇게 형성된 그래프 구조를 **Attention** 방식이 적용된 그래프 신경망을 통해 노드 간의 복합적인 관계와 전체 데이터의 패턴을 학습한다. 이후 학습이 완료된 모델의 출력값과 정상 시계열 간의 편차를 기반으로 **Anomaly Score**를 산출하고, 이를 이상 시계열 여부를 판별하는데에 사용한다.

이러한 **GDN**의 학습 과정을 도식화한 것이 **Figure 1**의 회색 영역이며, 정상 시계열과 이상 시계열을 구분할 수 있도록 학습된다. 다만 **GDN**에서는 각각의 시계열에 대한 임베딩이 무작위로 초기화 된 상태에서 학습을 시작하기 때문에 초기 학습 단계에서는 주어진 시계열의 특징이 충분히 반영되지 않는다. 이로 인해 초기 그래프 구조의 품질 저하와 학습 불안정성이 야기되며, 그래프 구조가 왜곡되어 최종 성능에 제한이 발생할 수 있다는 단점이 존재한다.

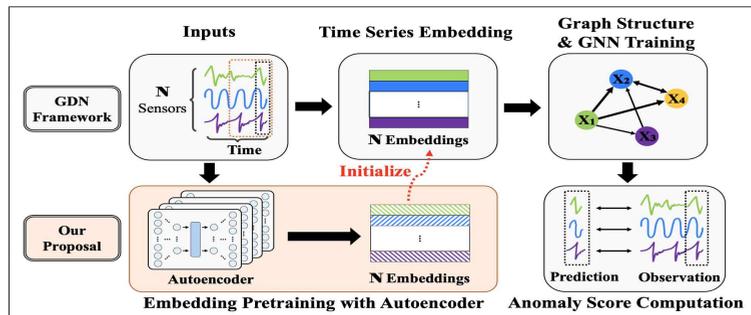


Figure 1. Architecture of the graph deviation network and the proposed pre-training scheme

2.2. AutoEncoder 기반 사전 학습

본 연구에서는 앞서 기술한 **GDN**의 초기 학습 단계에서 발생하는 문제점을 개선하기 위해 **Auto Encoder** 기반 사전 학습을 활용하여 시계열 특성이 반영된 **Latent Vector**를 추출하고, 이를 사용하여 시계열 임베딩을 초기화하는 방식을 제안한다. **AutoEncoder**는 입력 데이터를 저차원 **Latent Space**로 압축한 이후, 이를 입력 데이터로 다시 복원하는 비지도 학습 방식의 신경망이다. **Encoder**는 데이터를 함축적으로 표현하는 **Latent Vector**를 생성하고, **Decoder**는 이를 다시 원래 입력으로 복원하도록 학습한다. 이는 데이터 내의 주요 특징을 포함하면서 노이즈는 제거된 표현을 효과적으로 학습할 수 있으며, **Kortmann et al.(2021)**의 연구를 통해 이상 탐지나 시각화 등 다양한 분야

에서 AutoEncoder의 Latent Vector를 통해 시계열 특징을 추출하여 활용하는 것을 확인할 수 있었다. Figure 1의 주황색 구역은 본 연구에서 AutoEncoder를 활용한 사전 학습 과정을 도식화하여 보여주는데, 사전 학습은 크게 2가지 단계로 구성되어 있다. 먼저 AutoEncoder로 시계열을 학습할 때 사용하는 Window Size를 GDN 보다 더 크게 설정하여, 장기적인 구간에서의 시계열 특징을 학습할 수 있게 한다. 다음은 학습이 완료된 AutoEncoder로부터 Latent Vector를 추출하고 이를 그래프 구조에서 노드로 표현되는 임베딩의 초깃값으로 사용하여, GDN의 초기 학습 단계에서부터 시계열의 특징이 반영된 그래프 구조를 학습할 수 있도록 하였다. 이러한 사전 학습 과정을 통하여 기존의 무작위 초기화 방식 대비 초기 그래프 구조의 품질을 개선과 학습 안정성을 높여 GDN의 최종 성능을 향상시키고자 하였다.

2.3. 데이터 세트와 실험 조건

제안하는 사전 학습 방안의 검증을 위해 다변량 시계열 이상 탐지에 많이 활용되는 벤치마크 데이터 세트 3개를 사용하여 실험을 수행하였다. Su et al.(2019) 제시한 Server Machine Dataset(SMD)는 메모리/CPU 사용량 등 서버 환경에서 수집된 데이터 세트이며, 수자원 처리 시스템을 모형화한 Mathur & Tippenhauer(2016)의 Secure Water Treatment(SWaT)와 Ahmed et al.(2017)의 Water Distribution(WADI)는 여러 이상 시나리오에서 펌프, 센서 등으로부터 수집된 데이터 세트이다. 본 실험에서는 모델의 학습 속도를 높이기 위해 원본에서 10초마다 1개의 측정값으로 다운샘플링 하였고, Table 1은 실험에서 사용된 데이터 세트별 특징을 나타낸다.

실험에 주로 사용되는 하이퍼 파라미터로 시계열에 Window Size 100, Window Stride 10을 적용하였고, 각각의 시계열에 대한 Embedding Size는 128을 사용하였다. 이를 포함해 모든 데이터 세트에 대부분 동일한 조건을 사용하였으나 GDN은 그래프 구조를 형성할 때 임베딩 간 유사도가 높은 Top-K개의 노드 쌍에 대해서만 엣지를 생성하기 때문에, 데이터 세트별 특징 개수가 다른 점을 고려하여 Top-K 설정값은 20(SMD), 25(SWaT), 30(WADI)로 차등 적용하였다.

2.4. 실험 결과

먼저 AutoEncoder의 사전 학습 과정에서 Window Size를 각각 200/400/800으로 다르게 적용하여 Latent Vector를 생성하였고, 이를 사용하여 노드 임베딩을 초기화한 뒤 GDN 모델을 학습하였다. 실험은 각 데이터 세트마다 10회씩 반복 수행하여 F1 Score의 평균과 표준 편차를 기록하였고, Table 2에서 확인할 수 있듯이 모든 데이터 세트에서 사전 학습을 적용한 경우 성능이 향상됨을 관측할 수 있었다. 또한 반복 실험 시의 성능 편차가 개선되는 것을 통해 제시한 방안이 학습 안정성 향상에 기여함을 나타낸다. 추가로 데이터 샘플 수가 적은 SMD 데이터 세트에서는 사전 학습 시 Window Size 400, SWaT/WADI 데이터 세트에서는 Window Size 800을 적용했을 때 최대 성능이 관측되었는데, 이는 효과적인 사전 학습을 위해서는 시계열 길이에 따라 적절한 Window Size를 선택하는 것이 중요함을 시사한다.

Table 1. Statistics of the datasets used in experiments

Dataset	# Features	# Train	# Test	Anomaly Rate
SMD	38	23688	23688	4.20%
SWaT	51	47515	44986	11.97%
WADI	127	118795	17275	5.99%

Table 2. Performance comparison by embedding initialization method (The best results are bolded)

Dataset	Mertics	Random Initialize	Pretrain(200)	Pretrain(400)	Pretrain(800)
SMD	Precision	87.0±15.4	88.4±12.6	93.9±7.6	84.0±23.3
	Recall	60.5±16.5	66.4±6.6	72.1±6.2	64.6±15.3
	F1-score	71.0±17.2	75.7±8.5	81.4±5.6	72.8±18.3
SWaT	Precision	81.9±21.4	68.3±27.5	80.7±23.3	94.6±3.4
	Recall	63.1±5.6	67.3±5.8	63.5±5.2	60.4±5.8
	F1-score	70.1±11.3	64.2±13.2	69.0±11.0	73.6±4.6
WADI	Precision	57.7±11.2	72.0±18.8	65.7±17.0	61.4±14.0
	Recall	37.5±4.2	36.5±5.0	38.3±3.6	41.2±6.3
	F1-score	45.2±5.4	47.3±7.7	47.3±4.2	48.3±5.2

3. 결론

본 연구에서는 기존의 GDN 구조에서 발생할 수 있는 그래프 데이터의 왜곡을 방지하고 초기 학습 단계의 불안정성을 개선하기 위해 Auto Encoder 기반 사전 학습 방안을 제안하였다. 제안하는 방안을 통해 초기 학습 단계에서 시계열 정보를 활용할 뿐만 아니라 시계열의 장기적인 패턴을 포함하여 적절한 그래프 구조를 형성할 수 있도록 유도하였다. 이는 다변량 시계열 이상 탐지에서 활용되는 벤치마크 데이터 세트에서 반복 실험하여 연구의 효과와 실제 적용 가능성을 확인할 수 있었고, 본 연구를 통해 산업 현장에서의 실시간 이상 탐지 방안의 신뢰성과 효율성을 높이는 데 기여하고자 한다. 이후 후속 연구로는 LSTM-AE, Variational AE와 같은 AutoEncoder 변형 구조를 통해 임베딩 품질을 개선함으로써 사전 학습 효과를 극대화하고 사전 학습에 소요되는 시간을 최적화하는 것이 필요할 것으로 사료된다.

References

- Box, G. E., Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *Journal of the American statistical Association*, 65(332), 1509-1526.
- Ester, M., Kriegel, H. P., Sander, J., Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise, *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 96(34), 226-231.
- Deng, A., Hooi, B. (2021). Graph neural network-based anomaly detection in multivariate time series, *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5), 4027-4035.
- K.-P. Kortmann, M. Fehsenfeld, M. Wielitzka. (2021). Autoencoder-based representation learning from heterogeneous multivariate time series data of mechatronic systems, *arXiv preprint*, arXiv:2104.02784.
- Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., & Pei, D. (2019). Robust anomaly detection for multivariate time series through stochastic recurrent neural network, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2828-2837.
- Mathur, A. P., & Tippenhauer, N. O. (2016). SWaT: a water treatment testbed for research and training on ICS security, *2016 International Workshop on Cyber-Physical Systems for Smart Water Networks*, 31-36.
- Ahmed, C. M., Palleti, V. R., & Mathur, A. P. (2017). WADI: a water distribution testbed for research in the design of secure cyber physical systems, *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks*, 25-28.

Deep Learning-based Analysis of Cocaine-induced Behavioral Patterns in Marmosets

*Sun Mi Gu¹, ChulKyu Lee², Byoung Mo Cho³, Taeyun Yoo⁴,
Heejong Eom⁵, Dohyun Lee⁶, Jaesuk Yun⁷*

Abstract

Video-based behavior recognition plays a critical role in neuroscience research by enabling the objective analysis of diverse and complex behavioral patterns. However, current methods for behavior recognition in non-human primates (NHPs), such as marmosets, often require labor-intensive manual analysis, are time-consuming, and lack standardized evaluation protocols. This presents a major barrier, particularly in addiction research, where understanding drug-induced behavioral alterations is essential for developing pharmacological interventions. In this study, we utilized deep learning algorithms to examine behavioral responses in common marmosets (*Callithrix jacchus*) following cocaine administration. We focused on detecting rapid and repetitive psychomotor behaviors, especially head shaking, which is commonly observed in response to psychostimulants. To improve both accuracy and scalability in behavioral assessment, we utilized an automatic behavior classification algorithm, an open-source tool capable of analyzing video-based behavioral data without human supervision. The automatic behavior classification algorithm employs both pattern image extraction and animation-based features to capture spatiotemporal characteristics, enabling precise deep neural network-based behavior classification. Our algorithm demonstrated strong performance, achieving 90% accuracy, 94% sensitivity, and 92% F1

¹Post-Doc, College of Pharmacy, Chungbuk National University, 194-31 Osongsaengmyeong 1-ro, Osong-eup, Heungdeok-gu, Cheongju-si, Chungbuk 28160, Korea. E-mail: g09010327@nate.com

²Graduate student, College of Pharmacy, Chungbuk National University, 194-31 Osongsaengmyeong 1-ro, Osong-eup, Heungdeok-gu, Cheongju-si, Chungbuk 28160, Korea. E-mail: chlee970716@chungbuk.ac.kr

³Graduate student, College of Pharmacy, Chungbuk National University, 194-31 Osongsaengmyeong 1-ro, Osong-eup, Heungdeok-gu, Cheongju-si, Chungbuk 28160, Korea. E-mail: chlqudah97@naver.com

⁴Assistant Research Engineer, Non-clinical Center, Osong Medical Innovation Foundation, Cheongju-si, Chungcheongbuk-do, Korea. E-mail: youteayun@naver.com

⁵Research Engineer, Non-clinical Center, Osong Medical Innovation Foundation, Cheongju-si, Chungcheongbuk-do, Korea. E-mail: eomheejong@kbiohealth.kr

⁶Senior Research Engineer, Non-clinical Center, Osong Medical Innovation Foundation, Cheongju-si, Chungcheongbuk-do, Korea. E-mail: soohanlee@kbiohealth.kr

⁷(Corresponding Author)Associate Professor, College of Pharmacy, Chungbuk National University, 194-31 Osongsaengmyeong 1-ro, Osong-eup, Heungdeok-gu, Cheongju-si, Chungbuk 28160, Korea.
E-mail: jyun@chungbuk.ac.kr

score in correctly identifying cocaine-induced behavior categories. These research results highlight the potential of AI-driven behavior analysis tools for advancing research in behavioral neuroscience, especially in the context of drug-induced behavioral phenotyping, and establish a scalable framework for investigating substance-induced behaviors in neuroscience research.

Keywords : Marmoset, Deep learning, Cocaine, Stereotyped behavior, Drug-induced behavior.

서울시 동작구 재정비촉진구역 주택재개발사업의 단계별 지연 요인에 관한 연구

김남재¹, 변병설²

요 약

본 연구는 정비촉진구역 내 주택재개발사업의 지연 원인을 다각적으로 분석하고, 사업 단계별 주요 영향요인을 실증적으로 규명하였다. 기존 연구들이 사업 지연을 일반적 차원에서 접근한 것과 달리, 본 연구는 주택재개발사업의 제도적·절차적 특성을 고려하여 사업을 계획 및 구역지정, 조합설립 및 사업인가, 관리처분계획 인가, 착공 및 준공의 네 단계로 구분하였다. 학계 전문가, 공무원, 사업참여자, 부동산 전문가 등 167명을 대상으로 온라인 및 현장 설문조사를 실시하였으며, 수집된 자료는 SPSS 25.0을 활용하여 신뢰도 분석, 타당도 분석, 다중회귀분석 등을 수행하였다. 분석 결과, 경제적·사회적·행정적 요인이 사업 지연에 유의한 영향을 미치는 것으로 나타났으며, 특히 반복적인 행정서류 요구와 조합 청산 지연 등 행정적 복잡성이 최종 단계에서 핵심적인 지연 요인으로 작용함을 확인하였다. 또한 조합 임원의 전문성 부족이 주요한 영향요인으로 나타났고, 이해관계자 간 인식 차이와 지연 요인의 조절효과 역시 검증되었다. 본 연구는 도시재개발사업 지연 메커니즘에 대한 미시적 이해를 제고하고, 단계별 지연 완화 전략, 행정 절차의 간소화 및 투명성 제고, 거버넌스 역량 강화, 소통체계 개선 등을 통한 효율성 제고 및 지속가능한 도시재생을 위한 정책적 시사점을 제공한다.

주요 용어 : 주택재개발, 재정비촉진구역, 사업 지연 요인, 도시재생, 이해관계자 인식.

Physiological Curve-Inspired Feature Engineering for Fructose Prediction in Niitaka*

김선호¹, 김민수²

요약

배 ‘니타카(Niitaka)’의 과실 내 당 축적 과정은 복잡한 생리적 조절과 환경 요인에 의해 영향을 받으며, 이는 수확 전 품질 예측을 어렵게 한다. 본 연구는 과실을 채취하지 않고도 프럭토스(Fructose) 함량을 정량적으로 예측할 수 있는 비파괴적 모델을 개발하는 것을 목표로 하였다. 2018년부터 2022년까지 전라남도 나주시 왕곡면의 ‘신고’ 품종 배나무 농가에서 기상 및 생체 정보를 수집하였다. 기상자료는 일별 기온, 습도, 일사량, 강수량, 풍속 등으로 구성되며 DAFB(Days After Full Bloom) 구간별 평균값으로 집계하였다. 생체 데이터는 연도별로 무작위로 선정된 5주에서 각 1개 과실을 채취하여 11개 DAFB 구간의 생체중량, 수분함량, 건물중량 및 주요 당(Fructose, Glucose, Sucrose, Sorbitol)을 측정하여 총 275개의 관측치를 구축하였다. 예측 모델은 XGBoost 회귀 기반으로 설계하였으며, 생리적 과정을 모사하기 위해 DAFB 정규화값을 입력으로 하는 Logistic, Richards, Gaussian 곡선을 이용해 시간적 축적 및 전환 패턴을 근사하고, 이를 피쳐 엔지니어링 단계에서 명시적으로 반영하였다. LOYO(Leave-One-Year-Out) 교차검증 결과, 본 모델은 $R^2 = 0.90$, $RMSE \approx 42.4 \text{ mg} \cdot 100 \text{ g}^{-1} \text{ f.w.}$ 로 기상 단독 모델과 자기회귀 모델보다 향상된 성능을 보였다. 이러한 접근은 생리적 곡선 기반 피쳐 설계를 통해 과실 내 당 변화의 비선형적 특성을 효과적으로 포착할 수 있음을 입증하며, 향후 과수 품질 예측 및 생육 모형화 연구의 정밀도를 높이는 데 기여할 것으로 기대된다.

주요 용어 : Niitaka, 프럭토스(Fructose) 예측, XGBoost, Curve-based feature engineering, LOYO 교차검증.

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

¹61186 광주광역시 북구 용봉로 77, 전남대학교 데이터사이언스학과 석·박사과정. E-mail: ssuno2576@jnu.ac.kr

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

The Effect of Social Capital on Economic Performance

*Kim, Soonho*¹

Abstract

This study investigates the impact of social capital on economic performance and examines how this relationship is conditioned by sovereign Environmental, Social, and Governance (ESG) factors and globalization. Using an unbalanced panel dataset of 173 countries from 2010 to 2022, we estimate random-effects regressions to analyze both direct and interactive effects. The results reveal that social capital is positively associated with GDP growth, reflecting its role in facilitating trust, cooperation, and institutional efficiency. Sovereign ESG performance also contributes positively to economic outcomes. However, the interaction terms between social capital and ESG dimensions yield negative or diminishing coefficients, suggesting potential overlaps or institutional frictions when both factors are simultaneously strong. Similarly, globalization enhances growth independently but may weaken the marginal contribution of social capital, indicating crowding effects in highly open economies. Additional analyses show that these patterns differ between common-law and civil-law countries, where governance plays a more critical role in the latter. Overall, the findings underscore the complex, context-dependent role of social capital in promoting sustainable economic development amid ESG and globalization dynamics.

Keywords : social capital, ESG, globalization, economic performance, legal origin.

¹Professor, Division of Business Administration, Pukyong National University, Yongso-ro 45, Busan, Korea. E-mail: kimsoonho@pknu.ac.kr

경기도 학교급식용 친환경 배·당근 소득 및 경영비 분석

김연진¹, 이원석², 이정명³, 이수연⁴

요약

친환경 인증관리 정보시스템을 통해 추출된 모집단(당근 213호, 배 41호) 중 재배면적을 기준으로 5계층 단계로 구분하여 각 품목별 30농가를 단순임의표본추출하였다. 2024년 경기도 친환경 경농산물 채소 소득조사표를 활용하여 총수입, 경영비, 생산비, 노동시간 등을 조사하였다. 경영진단지표는 델파이 전문가조사를 통해 재배관리, 수확 및 판매관리, 경영관리 항목의 가중치를 설정하고, 5단계 세부수준으로 평가기준을 개발하였다. 상·하위 소득률 기준 각 20% 농가(6농가)를 선정하여 생산성 및 경영성과를 비교 분석하였다. 친환경 당근의 2024년 기준 10a당 생산량은 1,673kg, 총수입 4,856천원, 소득 2,051천원으로 소득률은 42.2%를 기록하였다. 생산비는 3,754천원으로 일반재배(3,227천원) 대비 16.4% 높았으며, 특히 노동비(고용+자가)가 전체 생산비의 51.9%를 차지하여 노동집약적 특성을 보였다. 경영진단 결과 관수 및 안전성검사 항목은 우수하였으나, 종자선택(23%), 시비(37%), 파종(30%), 제초·숙음(30%) 등 기초 재배관리와 경영기록관리(20%), 자금관리(17%) 등이 미흡한 것으로 분석되었다. 친환경 배의 2024년 10a당 생산량은 1,598kg, 총수입 10,030천원, 소득 4,937천원으로 소득률 49.2%를 나타냈다. 단가는 6,277원/kg으로 일반재배(3,122원/kg) 대비 약 2배 높았으며, 생산비는 8,097천원으로 일반재배(6,905천원)보다 17.3% 높았다. 경영진단 분석 결과 관수방법($r=0.425$, $p=0.019$)과 적화·적과·적심 관리가 소득과 유의한 상관관계를 나타냈다. 덕시설 및 적화·적과·적심 등 생산기술은 우수하였으나, 잡초관리(43%) 및 전정·전지 등 과원관리 기술이 상대적으로 미흡하였다. 경영진단지표는 경영성과, 재배기술, 수확 및 판매관리, 경영관리 등 4개 영역으로 구성되었으며, 각 항목별로 I~V단계의 세부평가 수준을 설정하였다. 개발된 진단지표를 당근 및 배 각 30농가 이상에 현장 적용한 결과, 문항 이해도와 타당도가 높은 것으로 평가되었다. 또한 농가별 경영진단 결과를 자동으로 분석하고 피드백할 수 있는 경영분석 프로그램을 개발하여, 맞춤형 컨설팅 도구로 활용 가능성을 확인하였다.

주요 용어 : 친환경농업, 소득조사, 경영진단지표, 생산성 분석, 당근, 배.

¹18388 경기도 화성시 병점중앙로 283-33, 경기도농업기술원 농업연구사. E-mail: boong0323@gg.go.kr

²18388 경기도 화성시 병점중앙로 283-33, 경기도농업기술원 농업연구관. E-mail: born815@gg.go.kr

³18388 경기도 화성시 병점중앙로 283-33, 경기도농업기술원 농업연구사. E-mail: jmlee@gg.go.kr

⁴18388 경기도 화성시 병점중앙로 283-33, 경기도농업기술원 농업연구관. E-mail: onpickle@gg.go.kr

B2C 캠페인 Agent 오케스트레이션 : Function Calling-RAG 결합 성능 분석과 강화학습 기반 최적화

김영교¹

요 약

본 연구는 오프라인 B2C(외식·리테일·렌탈·호텔 등) 캠페인·CRM 의사결정 업무를 지원하는 대화형 에이전트에서, SQL 기반 Function Calling과 RAG(Retrieval-Augmented Generation) 결합 방식의 품질을 비교하고, 이어 강화학습(Contextual Multi-Armed Bandit, CMAB) 기반 정책으로 자동 선택을 수행하는 방법을 제안한다. 동일한 데이터베이스·매뉴얼 코퍼스·모델(GPT-4o)·채점 스크립트 환경에서 (1) ToolOnly, (2) SimpleFusion, (3) CoordinatedFusion 세 가지 오케스트레이션을 구현하고 정확도(Accuracy), F1 score, 근거충실성(Faithfulness), 정책정합도(Policy Align), 지연시간(Latency)으로 평가하였다. 이후 CMAB 정책 학습을 통해 질의의 컨텍스트(질의 유형, 길이, 정책 키워드 포함 여부 등)에 따라 오케스트레이션 모드를 자동 선택하도록 하였다. 30문항 실험 결과, 강화학습 적용 후 정책정합도가 향상되었으며, 정확도·F1·근거충실성은 기존 수준을 유지하였다. 본 연구는 LLM + SQL + RAG 오케스트레이션의 재현 가능한 비교 프레임워크와 정책 중심 자동화의 실증 결과를 제시한다.

주요 용어 : LLM, AI Agent, RAG, Text-to-SQL, 강화학습.

¹02481 서울특별시 성북구 안암로 145 고려대학교 SW·AI융합대학원 인공지능융합학과 석사과정.
E-mail: titf7200@hanmail.net.com

도시열섬지수(UHI)에 관한 연구: 군집분석 · 회귀분석 중심으로*

김지현¹, 서은영², 송광윤³

요약

도시열섬 현상(Urban Heat Island, UHI)은 도심의 기온이 주변 지역보다 2°C 이상 높게 나타나는 현상으로 정의되며, 이는 에너지 소비 증가와 대기오염 심화 등 다양한 도시 환경 문제를 초래한다. 본 연구는 경기도 광주, 용인, 과천 중에서 광주를 기준 지역으로 선정하였으며, 선행 연구에서 AWS(Automatic Weather Station)로 측정된 결과를 참고하여 서울시 9개 자치구를 분석 대상으로 선정하였다. 2018년부터 최근 5년간 데이터를 활용하였으며, 시간별 기온자료 약 39만 건과 기상, 환경, 인구, 에너지 변수를 포함하였다. 연구 목적은 대규모 데이터를 기반으로 군집분석과 회귀분석을 중심으로 도시열섬 지수를 설명하는 주요 요인을 규명하는 것이다. 먼저, 계층적 군집분석을 통해 3개의 지역과 4개의 지역으로 분류된 그룹을 확인하였다. 이후 한강을 기준으로 구분된 3개 지역 모형을 최종 분석에 활용하였다. 그 결과, 그룹 1은 상업·업무 중심, 그룹 2는 중간 특성, 그룹 3은 주거 중심 지역으로 구분되었으며, 특히 주거 지역에서 녹지율이 상대적으로 높게 나타났다. 또한, 다중공선성 문제를 확인하여 이를 보완하기 위해 정규화 회귀분석 기법인 Lasso, Ridge, Elastic Net을 적용하였다. 추가로, 랜덤포레스트 회귀분석(Random Forest Regression)을 수행하였고 오존(O3), 풍속, 총 생활인구 수, 전기 사용량, 이산화질소(NO2)가 공통으로 도시열섬 지수에 영향을 미치는 주요 요인으로 도출되었다. 본 연구는 군집분석과 회귀분석 기법을 결합하여 도시열섬 현상을 통계적·기계 학습적으로 규명하였다는 점에서 의의가 있으며, 향후 도시 구조 및 환경 관리 정책 수립을 위한 기초자료로 활용될 수 있다.

주요 용어 : 도시열섬, 군집분석, 정규화 회귀분석, 랜덤포레스트 회귀분석, 변수중요도.

*이 논문은 2025학년도 조선대학교 학술연구비의 지원을 받아 연구되었음.

¹61452 광주 동구 조선대 3길 30, 조선대학교 전산통계학과, 석사과정. E-mail: kjm67899@chosun.ac.kr

²61452 광주 동구 조선대 3길 30, 조선대학교 전산통계학과, 석사과정. E-mail: khej4996@chosun.ac.kr

³(교신저자) 61452 광주 동구 조선대 3길 30, 조선대학교 컴퓨터통계학과 조교수. E-mail: ccssig@chosun.ac.kr

NASA 연구개발 펀딩데이터를 활용한 기관 생태계 변화 분석

노경란¹

요약

본 연구는 NASA의 연구개발 펀딩 데이터를 이용하여 NASA R&D를 수행한 기관의 생태계 구조 변화를 계량적으로 분석하였다. NASA Procurement Data View(NPDV)와 USAspending에서 수집한 NASA Grant 데이터를 통합하여 분석 데이터를 구축하였다. 이 데이터는 기관 단위의 연구개발 과제수, 펀딩 금액, 어워드 유형 정보가 포함하고 있다. 기관명 정제 및 동일기관 식별 등 데이터 전처리 과정을 거쳐 데이터 일관성을 확보하고자 하였다. 이를 기반으로 기관 유형별 과제 수, 총 투자금액, 과제당 평균 펀딩 규모, 주요 연구기관의 연도별 펀딩 추이, 그리고 어워드 유형(Grant, Cooperative Agreement)의 포트폴리오 차이를 살펴보았다. 분석 결과, NASA의 연구개발 펀딩은 과제수와 투자규모 측면에서 전체의 절반 이상이 공립·주립대학을 중심으로 이루어졌고, 대학 중심의 연구 생태계가 NASA의 핵심 구조를 형성하는 것으로 나타났다. 그리고 비영리기관과 중소기업의 참여가 확대되면서, NASA의 펀딩 구조가 다핵화·다층화되는 것으로 보였다. 특히 비영리기관은 대규모 전략 과제를 수행하면서 NASA의 전문 연구 파트너로 기능하는 것으로 나타났다. 대학은 기초과학과 인력양성을 담당하는 기반 역할을 주도하였다. 종합하면, NASA가 연구 생태계의 폭과 깊이를 동시에 확장하며, NASA Center가 위치한 지리적 조건을 이용하면서 대학·산업체·비영리기관이 공존하는 균형적인 연구개발 구조를 구축하는 것으로 여겨진다.

주요 용어 : NASA, 연구생태계, 펀딩 데이터, 연구개발 투자, 연구 포트폴리오.

*본 논문은 2025년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다
(과제번호: K25L4M2C3-01).

¹02456 서울시 동대문구 회기로 66, 한국과학기술정보연구원 글로벌 R&D분석센터, 책임연구원.

E-mail: infor@kisti.re.kr

브라우저 관여가 과업 복잡성 및 의사결정 의존성 간의 관계에 미치는 매개 영향에 대하여: 온라인 쇼핑에서의 페이지안 매개 분석

노민정¹

요 약

본 연구는 과업 복잡성이 전자상거래 플랫폼에 대한 의사결정 의존성에 영향을 미침에 있어 브라우저 관여가 어떻게 매개 변수로 역할하였는지 살펴보았다. 쿠팡과 알리익스프레스를 이용하는 한국 사용자 600명을 대상으로 페이지안 분석 프레임워크를 활용하여 설문조사를 실시했으며, 페이지안 분산분석 확장 모형을 사용하여 과업 복잡성과 의사결정 의존성 간 브라우저 관여의 매개 효과를 모델링했다. 연구 결과, 과업 복잡성이 증가하면 브라우저 관여도가 높아지고, 이는 다시 플랫폼 기능에 대한 의사결정 의존성을 강화하는 것으로 나타났다. 복잡한 과업의 인지적 요구에 의해 유발되는 탐색적 브라우징 행동은 사용자의 정신적 부담을 줄이는 동시에 플랫폼 기능에 대한 습관적 참여를 촉진함으로써 궁극적으로 구매 결정에 영향을 미치는 것으로 나타났다. 본 연구는 서열척도 종속변수에 적합한 페이지안 분산분석 확장 모형에 기반한 매개 프레임워크를 활용하여 기존 매개분석의 한계를 극복하고 방법론적 지평을 넓히고자 시도하였으며, 의사결정 의존성을 디지털 플랫폼이나 모바일 애플리케이션에 대한 사용자 참여의 깊이를 보여주는 측정도구로 활용함으로써 플랫폼 의존도를 체계적으로 분류하고 평가할 수 있는 새로운 관점을 제시하는 학술적 기여를 하고자 시도하였다.

¹부경대학교 경영학부

남자대학생의 인유두종바이러스(HPV)에 대한 지식, 백신접종에 대한 태도 및 수용도

마준용¹, 하운후², 주환성³, 최현서⁴, 남궁현우⁵, 김건우⁶, 전은미⁷

요 약

본 연구는 남자대학생 총 204명을 대상으로 인유두종바이러스에 대한 지식, 인유두종바이러스 백신접종에 대한 태도와 수용도를 확인하는 서술적 조사연구이다. 자료분석은 SPSS 25.0을 이용하여 빈도와 백분율, 평균과 표준편차, X²-검정, 일원분산분석과 사후검정은 Scheffe로 분석하였다. 인유두종바이러스 지식점수는 11점 만점에 평균 2.29±2.71점으로 낮게 나타났다. 인유두종바이러스 백신접종에 대한 태도는 61.3%가 자궁경부암과 생식기 사마귀를 함께 예방할 수 있는 백신을 선호하였고, 인유두종바이러스 백신 접종 시기는 53.4%가 첫 성관계 이전이 적절하다고 생각하였다. 연구대상자의 80.9%가 성파트너인 여성의 자궁경부암 예방을 위해 남성도 인유두종바이러스 예방접종을 해야 하며, 대상자의 57.9%가 자비로 인유두종바이러스 예방접종을 할 의향이 있다고 답변했다. 인유두종바이러스 백신접종 수용도에 따른 인유두종바이러스 지식의 차이는 “백신접종을 하겠다”고 응답한 대상자와 “하지 않겠다”고 응답한 군이 “모르겠다”라고 응답한 군보다 지식이 높은 것으로 나타났다.

주요 용어 : 남자대학생, 인유두종바이러스, 백신접종, 지식, 태도, 수용도.

¹35345 대전광역시 서구 배재로 155-40(도마동), 배재대학교 간호학과 학부재학생. E-mail: wnsdyd2385@naver.com

²35345 대전광역시 서구 배재로 155-40(도마동), 배재대학교 간호학과 학부재학생. E-mail: yunhoo1028@naver.com

³35345 대전광역시 서구 배재로 155-40(도마동), 배재대학교 간호학과 학부재학생. E-mail: jws021227@naver.com

⁴35345 대전광역시 서구 배재로 155-40(도마동), 배재대학교 간호학과 학부재학생. E-mail: chs34018@naver.com

⁵35345 대전광역시 서구 배재로 155-40(도마동), 배재대학교 간호학과 학부재학생. E-mail: nghu2002@naver.com

⁶35345 대전광역시 서구 배재로 155-40(도마동), 배재대학교 간호학과 학부재학생. E-mail: rjsdn200208@naver.com

⁷(교신저자)35345 대전광역시 서구 배재로 155-40(도마동), 배재대학교 간호학과 교수.

E-mail: charminggold@pcu.ac.kr

Association Between Physical Activity Timing and Metabolic Syndrome in Korea: A Functional Principal Component Approach^{*}

Suah Park¹, Hee-Jung Jee²

Abstract

Metabolic Syndrome (MetS) significantly increases the risk of cardiovascular disease and type 2 diabetes. The prevalence of MetS is steadily increasing in South Korea, further emphasizing the importance of identifying risk factors. While physical activity (PA) is beneficial, recent evidence suggests that the temporal pattern of activity, not just its total volume, may be crucial for metabolic health. This study evaluates the association between PA temporal patterns and MetS risk in Korean adults. We analyzed minute-level accelerometer data from Korean adults who participated in the 2014 - 2016 Korea National Health and Nutrition Examination Survey (KNHANES). Functional principal component analysis (FPCA) was applied to the physical activity trajectories to extract key temporal patterns. Logistic regression models, adjusted for total moderate-to-vigorous physical activity (MVPA) and confounding variables, tested the association between the resulting principal component scores and MetS. Four major components explaining the daily physical activity patterns were identified. Among these, the third principal component (PC3)—reflecting a pattern with reduced afternoon variability—was significantly associated with an increased risk of MetS in the fully adjusted model (adjusted OR = 1.117; 95% CI: 1.003 - 1.244). These findings suggest that reduced variability in afternoon physical activity may be linked to adverse metabolic outcomes. Beyond overall activity volume, the timing of physical activity should be considered in metabolic health research.

Keywords : Metabolic syndrome, Physical activity, Temporal pattern, Functional principal component analysis.

^{*}This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00354881 and RS-2024-00440787).

¹MS course student, Department of Statistics, Chungbuk National University, Cheongju 28644, Korea. E-mail: 2025221010@chungbuk.ac.kr

²(Corresponding Author) Assistant professor, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea. E-mail: heejung.jee@chungbuk.ac.kr

딥러닝 기반 시계열 예측 경량화 모델의 비교 분석*

박현준¹, 김성용², 임동혁³

요약

정확한 시계열 예측을 위해 트랜스포머 기반의 복잡한 딥러닝 모델이 사용되지만, 이는 수백만 개의 파라미터와 높은 연산 비용으로 인해 실시간 응답성이 필요한 환경에 배포되기 어렵다. 이로 인해 모델 복잡도와 연산 비용을 획기적으로 줄인 경량 모델이 주목받고 있다. 본 논문에서는 복잡한 모델의 성능을 유지하면서 효율성과 단기 예측 성능을 향상시키기 위해 세그먼트 기반의 경량 모델을 제안한다. 제안 모델은 먼저 시계열을 기간 단위의 하위 시퀀스로 분할한다. 이후 선형 레이어를 통해 기간 인덱스 간의 관계, 즉 기간 내 관계를 명시적으로 모델링한다. 다음으로, 각 하위 시퀀스를 다시 다수의 세그먼트로 분할하여 계층적 학습을 수행한다. 첫 번째 단계에서는 각 세그먼트 내부의 미세한 지역적 트렌드 패턴을 학습한다. 두 번째 단계에서는 세그먼트 간의 관계를 학습하여 전역적 트렌드를 구성한다. 이 모든 과정은 단순한 선형 레이어로 구현되어 최소한의 파라미터로 높은 표현력을 갖도록 설계된다. 벤치마크 데이터셋을 사용한 실험 결과, 목표했던 단기 예측에서 기존 경량 모델 및 복잡한 트랜스포머 기반 모델보다 우수한 예측 성능을 달성했다. 제안 모델은 실험에 사용된 모든 비교 모델 중 가장 적은 파라미터로 우수한 단기 예측 정확도를 달성했다. 또한, 비교 모델 중 가장 빠른 추론 시간을 기록했으며 학습 시간 역시 최고 수준의 효율을 보였다. 결론적으로, 리소스가 제한된 환경에서의 실시간 단기 예측을 위해, 예측 정확도와 연산 효율성 간의 균형을 효과적으로 달성한 실용적인 모델임을 입증한다.

주요 용어: 다변량 시계열 예측, 단기 시계열 예측, 경량 모델, 세그먼트 단위 학습.

*이 논문은 2025년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (RS-2024-00406796, 2025년 산업혁신인재성장지원사업).

¹01897 서울특별시 노원구 광운로 20, 광운대학교 인공지능응용학과 석사과정. E-mail: parkhj59@kw.ac.kr

²31499 충청남도 아산시 배방읍 호서로 79번길 20. 호서대학교 빅데이터AI학부 교수. E-mail: yaba96@hoseo.edu

³(교신저자) 01897 서울특별시 노원구 광운로 20, 광운대학교 정보융합학부 교수. E-mail: dhim@kw.ac.kr

1인 가구의 디지털 활용능력이 웰에이징에 미치는 영향 연구*

성순임¹, 김종숙², 김광환³

요 약

본 연구는 1인 가구의 디지털 활용능력이 웰에이징에 미치는 영향을 알아보기 위해 1인 가구의 연령대에 따른 디지털활용 역량에 따른 특성을 파악하였다. 조사대상 및 기간은 1인 가구 416명을 대상으로 온라인 설문조사를 2025년 9월 15 ~ 9월 21까지 7일간 실시하였다. 연구결과 연령대에 따른 노트북 보유와 노트북 사용이 언제나 가능한 상태가 60세 이상에서 평균 점수가 높았으며 통계적으로 유의하였다($p < 0.01$). 이용 가능한 기기를 보유한 스마트 패드와 스마트 주변기기 또한 29세 이하보다 60세 이상에서 평균 점수가 높았으며, 통계적으로 유의한 차이를 나타냈다($p < 0.001$). 연령대에 따른 디지털매체 활용 상태는 프로그램 설치 및 문서작성, 뉴스 및 정보 검색, 이메일 콘텐츠 활용, 사회관계 및 정보공유는 60세 이상보다 29세 이하의 평균이 높았으며, 통계적으로 유의하였다($p < 0.001$). 스마트기기 활용 상태는 생활정보에서 유의한 차이가 있었으며($p < 0.001$), 디지털기기 이용 대인 관계 상태는 알고 있는 지인과 인터넷 활용이 통계적으로 유의하였다($p < 0.05$). 디지털 매체 과다사용 정도는 식사 중 디지털 매체물두와 디지털 기기 사용 중 행복하다가 통계적으로 유의한 차이를 보였다($p < 0.001$). 코로나19 이후 비대면 생활이 일상화되면서 외로움 등 정서적 욕구가 미디어 이용의 주요 동기가 되었으며, 일과 삶의 균형을 이루고 건강하고 행복하게 나이 들어가려는 ‘웰에이징(Well-aging)’에 대한 관심이 높아지고 있다. 이에 정부에서는 단순한 복지 지원보다는 소득 유지 및 취업 지속 지원, 생활권 기반의 소규모 커뮤니티 활성화, 심리·정서 지원 체계의 강화가 필요할 것으로 사료된다.

주요 용어 : 1인 가구, 매체활용, 웰에이징, 디지털리터러시, 디지털 사용수준.

*본 논문은 한국연구재단의 지원을 받아서 연구된 것임(NRF-2020S1A5C2A04092504).

¹35365 대전광역시 서구 관저동로 158, 건양대학교 웰다잉융합연구소 연구원. E-mail: ssi0307@naver.com

²35365 대전광역시 서구 관저동로 158, 건양대학교 웰다잉융합연구소 연구원. E-mail: ball-123@kakao.com

³35365 대전광역시 서구 관저동로 158, 건양대학교 병원경영학과 교수. E-mail: kkh@konyang.ac.kr

사용자 입력 데이터의 정형화 및 피드백 자동화를 위한 챗봇 설계*

신준호¹, 장원혁², 정상훈³, 김동재⁴, 김동현⁵, 문형빈⁶

요약

최근 인공지능과 자연어 처리(NLP) 기술이 급격히 발달함에 따라 대규모 언어 모델(Large Language Model, LLM)의 성능이 고도화되고 있다. 이에 LLM을 활용한 다운스트림 태스크(downstream task)가 활발히 연구되고 있으며, 특히 정보 추출(Information Extraction)은 LLM의 주요 응용 분야로 부각되고 있다. 본 연구는 LLM을 기반으로, 사용자 입력의 비정형 데이터를 정형화하고 이 과정에서 누락된 정보를 식별하여 사용자에게 능동적으로 피드백하는 챗봇 시스템을 설계하였다. 제안하는 시스템 아키텍처는 사용자의 의도를 동적으로 파악하여 유연한 상호작용을 수행하며, 이와 동시에 사용자 정보를 효과적으로 정형화한다는 점에서 기존 아키텍처와 차별성을 가진다. 사례 연구로서 (주)뉴아이의 세금 솔루션 시뮬레이션을 대상으로 하였으며, 자체 구축한 검증 데이터셋을 활용하여 사용자 의도를 파악하는 Intent_LLM과 정보를 정형화하는 Struct_LLM의 성능을 분류 성능 지표로 평가하였다. 실험 결과, Intent_LLM은 Macro F1-score 0.94, Struct_LLM은 F1-Score 0.9583을 달성하며 제안 모델의 높은 신뢰성과 실효성을 확인하였다. 본 연구는 복잡한 정보를 다루는 도메인에서 LLM이 사용자의 의도를 유연하게 추론하고, 동시에 데이터 정형화를 수행할 수 있는 고도화된 대화형 시스템의 실질적인 구현 가능성을 제시했다는 점에서 그 의의를 가진다.

주요 용어: 대규모 언어 모델, 정보 추출, 데이터 정형화, 챗봇.

*본 연구는 2025 과학기술정통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음 (2024-0-00018).

¹부산광역시 용소로 45, 국립부경대학교 데이터공학과 석사과정. E-mail: 202455077@pukyong.ac.kr

²부산광역시 용소로 45, 국립부경대학교 데이터공학과 석사과정. E-mail: wonhyeok1104@gmail.com

³부산광역시 용소로 45, 국립부경대학교 컴퓨터공학과 학부재학생. E-mail: 1230david@naver.com

⁴부산광역시 용소로 45, 국립부경대학교 데이터공학과 석사과정. E-mail: dongjaekim@pukyong.ac.kr

⁵부산광역시 남구 신선로 365, (주) 뉴아이 대표. E-mail: kdh@taxai.co.kr

⁶부산광역시 용소로 45, 국립부경대학교 빅데이터융합전공/데이터공학과 부교수. E-mail: hbmoon@pknu.ac.kr

The Taylor rule and house price volatility

*Sunghoon Sim*¹

Abstract

This study aims to determine whether house price uncertainty has been an important determinant of the Taylor rule-based interest rate during the recent booming period in Korean housing markets. An EGARCH-based specification has been used to produce a time-varying measure of volatility. Our empirical results indicate that the house price uncertainty has had a significant negative effect on the policy interest rate, but that its addition only produces a slightly better fit to the actual interest rate.

Keywords : Taylor rule, house price, interest rates, inflation gap, output gap.

1. Introduction

In the context of the global financial crisis in 2008, market participants debated actively the role of asset price volatility on determining the monetary policy. Economists have not currently agreed the issue on whether the central bank should react to the movement in asset prices. According to Bernanke and Gertler (1999), the disagreement comes from the fact that the central bank cannot make a distinction between asset price bubbles and normal movements. Goodhart and Hofmann (2001) have also asserted the important role of asset price volatility in determining monetary policy as follows. First, asset prices and volatility can be affected by the future expected returns, which relate to the expected economic conditions, inflation and monetary policy. Second, since persistent asset price volatility will result in financial instability, asset price stability should be considered as an objective if the central bank wants to maintain economic stability. The aim of this study is to determine the effect of house price uncertainty on monetary policy using a Taylor rule approach and a measure of volatility to proxy uncertainty.

2. The Model and Empirical Results

Our analysis is based on the conventional formulation of the Taylor rule as presented in Taylor (1993).

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

$$i_t = r^* + \pi_t + \beta_\pi (\pi_t - \pi_t^*) + \beta_y (y_t - y_t^*) \quad (1)$$

The above Taylor rule model is augmented as in equation (2) to check whether house price uncertainty has been an important determinant of the Taylor rule-based interest rate during our sample period.

$$i_t = r^* + \pi_t + \beta_\pi (\pi_t - \pi_t^*) + \beta_y (y_t - y_t^*) + \beta_v \sigma_t^2 \quad (2)$$

where i_t denotes the nominal policy rate at time t , π_t is the inflation rate, and y_t is the growth rate of real GDP. r_t^* indicates a long-run equilibrium real interest rate, π_t^* is a target inflation rate, and y_t^* is a growth rate of potential GDP. The last two variables are unobservable. Meanwhile, σ^2 represents the volatility of the house price returns. β_π , β_y and β_v denote response of monetary policy to inflation, output gap and the volatility of the housing market, respectively. The model in this study is estimated using Ordinary Least Square method. The data used in this study are inflation based on the CPI, growth rate of GDP which are obtained from the Bank of Korea (BOK). The national apartment price index is used for the house price and is obtained from Kookmin bank. Following the previous studies, the long-run equilibrium real interest rate is assumed to be 2% from 1999 through 2008 (the period of global financial crisis) and is to be 1% after the financial crisis. Moreover, following the research studies issued by Bank of Korea, the potential output growth in this study is set to be the range from 6.1% in 1999 to 2% in 2025. The target inflation rate in this study is based on that of BOK. Finally, we measure the volatility of the house price based on the Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) model developed by Nelson (1991).

Table 1 displays the results of the EGARCH model with all parameter estimates being significant. To test whether there are any remaining Autoregressive Conditional Heteroskedasticity (ARCH) effects in the EGARCH model, we carried out the Lagrange Multiplier (LM) test to determine whether the standardized residuals exhibit additional ARCH. The results suggest that the null of no remaining ARCH is accepted, so there is no ARCH left in the standardized residuals.

Table 1. EGARCH results for the house price volatility

Parameters	Coefficients	Probability
β_π	-1.503***	0.0003
β_y	0.736***	0.0000
β_σ	0.329***	0.0003
$adj - R^2$	0.906***	0.0000
ARCH effect	0.0643	0.5181

Note: *** indicates 1% significance level.

Table 2 presents the estimated results based on the traditional Taylor rule and augmented Taylor rule. We assume a long-run equilibrium real interest rate to be positive as usual. Further, we assume that the parameters for response of monetary policy to inflation and output gap are positive as in Taylor (1993). We also assume the policy rate reacts to the house price uncertainty negatively, as the central banks usually try to boost the economy by easing monetary policy and increasing the supply of credit to the market in the event of economic uncertainty. Therefore, if there is asset price volatility and it is likely to increase in the future, the central bank will stimulate the economy by decreasing the policy rate.

For both models, the long-run equilibrium real interest rates are found to positively and strongly affect the policy interest rate as expected. For the augmented model, the coefficient of inflation is 0.17, which is consistent with the Taylor principle and implies that including house price volatility in the Taylor rule suggests central bank has a stronger reaction to inflation. In spite of being insignificant, the response of the interest rate to the output gap is 0.024 and increases compared to the first model. The coefficient of house price volatility is -14.15 and significant at 1% confidence level, which means that if the volatility in the housing market increases, central banks will ease monetary policy. The intuition behind this finding is that the more volatile house prices are, the more uncertainty there will be in the market. In a more uncertain environment, there is less incentive to invest. With less investment and less liquidity, the aggregate demand will be adversely affected, motivating a central bank to react by decreasing the policy rate. This will encourage investors to invest and it will also inject more liquidity into the market. This result supports the fact that the conventional Taylor rule has not been the only factor affecting the Federal Reserve's decision making.

By comparing the augmented model with the standard model, it is apparent that the reaction of the interest rate to inflation is larger with house price volatility included than that in the first model, implying that the volatility of house prices may also create a threat to inflation and therefore the central bank reacts more strongly in order to control it. This results suggests that the augmented Taylor rule outperforms the benchmark model, considering the huge reaction of the central bank to the house price volatility. Finally, both Figure 1 (a) and (b) provide a comparison between the actual value and the fitted value for two models. For both Figures, the fitted values are a good approximation with the actual rates fitting very well from 2008 to 2010, which coincide with the recession period of global financial crisis. During the soaring period of house prices from 2017 to 2023, the actual value is lower by about 1% point to 2% point and the monetary policy is relatively loose. In Figure 1 (b), the difference between the actual rate and the fitted value is smaller than the conventional model (Figure 1 (a)), suggesting overall a better fit to the data. According to our result, the Bank of Korea should have raised the policy interest rate up to 1% point or 2% point in order to soothe the bubble-suspected housing markets in Korea.

Table 2. Results of traditional Taylor rule and augmented Taylor rule

Coefficients	Model 1	Model 2
r^*	2.195***	2.374***
β_π	0.142***	0.172**
β_y	0.007	0.024
β_σ	—	-14.156***
$adj-R^2$	0.665	0.694

Note: ***, ** indicate 1%, 5% significance level, respectively.

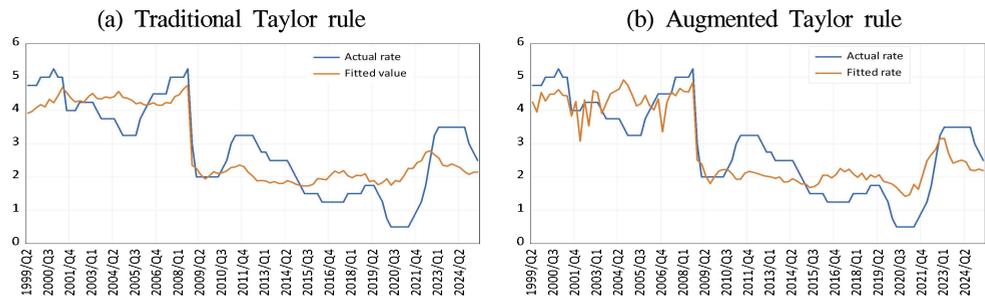


Figure 1. Actual and fitted interest rates for traditional Taylor rule and augmented Taylor rule

3. Conclusion

This paper provides evidence that house price volatility has had a strongly negative effect on interest rates over the past two decades, indicating the Bank of Korea should have considered asset price uncertainty when deciding on interest rate policy. The augmented Taylor rule shows a better goodness of fit compared to the result by the traditional Taylor rule, although both rules show that monetary policy during the booming period in housing markets from 2017 to 2023 is too loose. However, even when considering asset price volatility, the interest rates were arguably too low since 2015, leading up to the housing bubbles in Korean housing markets. In the future, the Bank of Korea need to consider including a more formal measure of asset uncertainty such as house price volatility in the Taylor rule.

References

- Bernanke, B. S., Gertler, M. (1999). Monetary policy and asset price volatility, *Federal Reserve Bank of Kansas City Economic Reviews*, 84, 17-51.
- Goodhart, C., Hofmann, B. (2001). Asset prices, financial conditions and the transmission of monetary policy, Paper presented on Asset prices, exchange rates and monetary policy, Stanford University, Stanford, CA.
- Nelson, D. B. (1991). Conditional Heteroskedasticity in asset returns: a new approach, *Econometrica*, 59, 347-370.
- Taylor, J. (1993). Discretion versus policy rules in practice. Carnegie-Rochester Conference Series on Public Policy, 39, 195-214.

Distributional Effects of the Cultural Voucher Program

*Juhyun Oh*¹

Abstract

This study investigates the distributional effects of the Cultural Voucher Program on household income distribution in Korea. Previous research on cultural welfare policies has mainly relied on survey-based analyses focusing on cultural participation or satisfaction, with limited empirical evidence on how cultural fiscal expenditure affects income inequality and cultural disparities. To fill this gap, this study employs data from the Korean Welfare Panel Study, which enables precise identification of program beneficiaries, to analyze the program's impact on income distribution. By applying the decomposition of the Gini coefficient, the study quantifies the within- and between-group contributions to inequality, thereby assessing how the Cultural Voucher Program influences the structure of income distribution across households. The findings provide empirical evidence that cultural fiscal expenditure, beyond serving as a simple welfare subsidy, contributes to mitigating social inequality and promoting inclusive growth. These results offer practical implications for the effective design of cultural welfare programs and the formulation of long-term fiscal strategies.

Keywords : Cultural fiscal expenditure, Cultural Voucher Program, Income distribution, Gini decomposition, Cultural welfare policy.

¹College of Business, Gachon University, Gyeonggi-do 13120, Korea. Visiting Professor.
Email: ohjh7883@gachon.ac.kr

자기효능감이 성공적 노화에 미치는 영향: 인지적 유연성과 심리적 탄력성의 매개효과*

왕흔이¹, 한현석²

요 약

인구 고령화가 심화되는 사회적 맥락에서, 성공적 노화를 증진하기 위한 방안은 현대 사회에서 중요한 과제로 대두되고 있다. 본 연구의 목적은 한국의 고령층의 자기효능감과 성공적 노화 간의 관계에서 인지적 유연성과 심리적 탄력성의 이중 매개효과를 검증하는 데 있다. 이를 위해 국내에 거주하는 65세 이상 노인 300명을 대상으로 자료를 수집하였고, 노인의 자기효능감, 성공적 노화, 인지적 유연성, 심리적 탄력성 수준을 측정하였으며, 자료 분석에는 SPSS 27.0과 PROCESS macro를 사용하였다. 분석을 통한 주요 결과는 다음과 같다. 자기효능감은 성공적 노화에 정적 영향을 미치는 것으로 나타났다. 인지적 유연성과 심리적 탄력성은 각각 자기효능감과 성공적 노화와의 관계에서 매개하는 것으로 나타났다. 특히 자기효능감이 인지적 유연성과 심리적 탄력성을 순차적으로 거쳐 성공적 노화에 이르는 이중매개효과 통계적으로 유의미한 것으로 확인되었다. 이 결과를 통해 성공적 노화를 촉진하기 위해 인지적 유연성과 심리적 탄력성 같은 내적 심리 자원의 영향을 실증적으로 규명하였으며, 이에 노인의 자기효능감을 강화하는 구체적인 개입 방안 마련에 필요한 이론적·실천적 토대를 제공하였다.

주요용어 : 자기효능감, 성공적 노화, 인지적 유연성, 심리적 탄력성

*본 논문은 전북대학교의 지원을 받아서 연구된 것임.

¹54896 전북특별자치도 전주시 덕진구 백제대로 567, 전북대학교 심리학과 석사과정. E-mail: mollyzxc@jbnu.ac.kr

²(교신저자) 54896 전북특별자치도 전주시 덕진구 백제대로 567, 전북대학교 심리학과 부교수.

E-mail: hysh@jbnu.ac.kr

층화 직접질문 조건부 양적속성 무관질문모형*

이기성¹, 홍기학², 손창균³, 박근화⁴, 홍성준⁵

요 약

본 논문에서는 매우 민감한 조사에서 모집단이 층으로 구성되어 있을 때 덜 민감한 속성에 대하여 응답자들에게 직접질문을 하여 “예”라고 응답한 사람들에게만 민감한 변수와 무관한 변수를 포함하고 있는 양적속성 무관질문모형을 사용하도록 하여 민감한 변수에 대하여 층화 추정할 수 있는 층화 직접질문 조건부 양적속성 무관질문모형을 제안하였고, 제안한 층화 직접질문 조건부 양적속성 무관질문모형과 Greenberg et al.(1971)의 양적속성 무관질문모형에 층화추출법을 적용한 층화 양적속성 무관질문모형과의 효율성을 비교하였다.

주요 용어 : 덜 민감한 속성, 직접질문, 양적 무관질문모형, 층화추출, 조건부 모형.

1. 서론

Warner(1965)는 민감한 질문과, 민감한 질문과 배반되는 질문으로 구성된 확률장치를 사용하여 민감한 속성의 모비율을 추정하고 있으며, Greenberg et al.(1971)은 민감한 양적속성의 정보를 얻을 수 있는 양적속성 무관질문모형을 제안하여 민감한 변수의 모평균을 추정하였다. 그리고 Loynes (1976)는 민감한 질문과 강요된 응답 “예”로 구성된 확률장치를 이용하여 민감한 속성의 모비율을 추정하는 강요질문모형을 제안하였으며, Carr, Marascuilo.(1982)은 덜 민감한 질문과 강요된 응답으로 구성된 확률장치를 통해 “예”라고 응답한 사람들에게만 Loynes의 강요질문모형을 사용하도록 하여 민감한 속성에 대한 모비율을 추정해 내는 조건부 확률화응답모형을 제안하였다. 또한 Ahn, Lee.(2003)는 직접질문에 의한 조건부 양적속성 무관질문모형을 제안하기도 하였다.

본 논문에서는 매우 민감한 조사에서 모집단이 층으로 구성되어 있을 때 덜 민감한 속성에 대하여 응답자들에게 직접질문을 하여 “예”라고 응답한 사람들에게만 민감한 변수 X 와 무관한 변수 Y 를 포함하고 있는 Greenberg et al.의 양적속성 무관질문모형을 사용하도록 하여 민감한 변수 X 에 대하여 층화 추정할 수 있는 층화 직접질문 조건부 양적속성 무관질문모형을 제안하고, 제안한 층화 직접질문 조건부 양적속성 무관질문모형과 Greenberg et al.의 양적속성 무관질문모형에 층화추출법을 적용한 층화 양적속성 무관질문모형과의 효율성을 비교하고자 한다.

*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 명예교수. E-mail: khhong@dsu.ac.kr

³38066 경북 경주시 동대로 123 동국대학교 보건의료정보학과 교수. E-mail: ckson85@donguk.ac.kr

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

⁵04620 서울특별시 중구 필동로 1길 30 동국대학교 융합교육원 겸임교수. E-mail: hsj8129@naver.com

2. 층화 직접질문 조건부 양적속성 무관질문모형

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 민감한 변수 X 와 무관한 변수 Y 를 포함하고 있는 Greenberg et al.의 양적속성 무관질문모형을 사용하도록 하여 민감한 변수 X 에 대하여 층화 추정할 수 있는 층화 직접질문 조건부 양적속성 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 층에서 Greenberg et al.의 양적속성 무관질문모형을 사용하는데 있어서 민감한 변수 X 가 연속인 밀도함수 $g_h(\cdot)$ 를 갖는다고 가정하고, Y 를 밀도함수 $h_h(\cdot)$ 를 갖는 무관한 변수라 하자. 이 때, h 층의 무관한 변수의 모평균 μ_{hy} 를 알고 있다고 가정하며, 응답자가 Z_h 라고 응답한다고 가정하자. 그리고 Z_h 의 모평균과 모분산을 각각 μ_{hz}, σ_{hz}^2 이라 하자.

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, $\hat{\lambda}_{h1} = n_{h1}/n_h$ 이므로 덜 민감한 속성 B_h 에 대한 모비율 π_{h1} 의 추정량 $\hat{\pi}_{h1}$ 는 $\hat{\pi}_{h1} = n_{h1}/n_h$ 이 된다.

h 층의 두 번째 단계에서는 첫 번째 단계에서 “예”라고 응답한 n_{h1} 명의 응답자들만이 다음과 같은 2개의 질문으로 구성된 Greenberg et al.의 양적속성 무관질문모형의 확률장치를 이용하게 된다.

	질문내용	선택확률
질문 1	당신의 민감한 변수 X 에 대한 값은 얼마입니까?	p_h
질문 2	당신의 무관한 변수 Y 에 대한 값은 얼마입니까?	$1-p_h$

그러므로 h 층의 응답자가 첫 번째 단계에서 “예”라고 응답했다는 조건하에서 두 번째 단계에서 Z_h 라고 응답할 때 기대할 수 있는 응답의 평균 즉, Z_h 의 모평균 μ_{hz} 는 다음과 같다.

$$\mu_{hz} = p_h \frac{\mu_{hx}}{\pi_{h1}} + (1-p_h)\mu_{hy}. \quad (2.1)$$

h 층의 첫 번째 단계에서 “예”라고 응답한 n_{h1} 명의 응답자들이 두 번째 단계에서 Z_{hi} ($i = 1, 2, \dots, n_{h1}$)라고 응답했을 때, h 층의 X 에 대한 모평균 μ_{hx} 의 추정량 $\hat{\mu}_{hx}$ 는 다음과 같다.

$$\hat{\mu}_{hx} = \hat{\pi}_{h1} \left[\frac{\hat{\mu}_{hz} - (1-p_h)\mu_{hy}}{p_h} \right] = \frac{1}{n_h p_h} [Z_{h0} - (1-p_h)\mu_{hy} n_{h1}]. \quad (2.2)$$

여기서, $Z_{h0} = n_{h1}\hat{\mu}_{hz}$, $\hat{\mu}_{hz} = \frac{1}{n_{h1}} \sum_{i=1}^{n_{h1}} Z_{hi}$, $E(\hat{\mu}_{hz}|n_{h1}) = \mu_{hz}$, $V(\hat{\mu}_{hz}|n_{h1}) = \frac{\sigma_{hz}^2}{n_{h1}}$ 이다.

따라서 민감한 변수 X 에 대한 모평균 μ_x 의 추정량 $\hat{\mu}_x$ 는 다음과 같다.

$$\hat{\mu}_x = \sum_{h=1}^L W_h \hat{\mu}_{hx} = \sum_{h=1}^L W_h \frac{1}{n_h p_h} [Z_{h0} - (1-p_h) \mu_{hy} n_{h1}], \quad W_h = \frac{N_h}{N}. \quad (2.3)$$

추정량 $\hat{\mu}_x$ 는 모평균 μ_x 의 비편향추정량이며, 그 분산은 다음과 같다.

$$V(\hat{\mu}_x) = \sum_{h=1}^L W_h^2 \left[\frac{\pi_{h1} \sigma_{hz}^2}{n_h p_h^2} + \frac{\mu_{hx}^2 (1-\pi_{h1})}{n_h \pi_{h1}} \right]. \quad (2.4)$$

3. 효율성 비교

이 절에서는 제안한 층화 직접질문 조건부 양적속성 무관질문모형과 층화 양적속성 무관질문모형과의 효율성을 분산을 이용하여 비교해 보고자 한다. 층화 양적속성 무관질문모형에서 민감한 변수 X 의 추정량 $\hat{\mu}_g$ 의 분산은 다음과 같다.

$$V(\hat{\mu}_g) = \sum_{h=1}^L W_h^2 \frac{\sigma_{hz}^2}{n_h p_h^2}. \quad (3.1)$$

여기서, σ_{hz}^2 은 h 층의 응답자들이 Z_h 라고 응답한다고 할 때, Z_h 의 모분산이다.

제안한 층화 직접질문 조건부 양적속성 무관질문모형의 분산 식(2.4)와 층화 양적속성 무관질문모형의 분산 식(3.1)에서 $V(\hat{\mu}_g) - V(\hat{\mu}_x) > 0$ 을 이용하여

$$V(\hat{\mu}_g) - V(\hat{\mu}_x) = \sum_{h=1}^L W_h^2 \frac{(1-\pi_{h1})(\pi_{h1} \sigma_{hz}^2 - p_h^2 \mu_{hx}^2)}{\pi_{h1} n_h p_h^2} > 0$$

으로부터 $V(\hat{\mu}_x) < V(\hat{\mu}_g)$ 를 만족하는 식을 μ_{hx} 의 기준으로 정리하여 나타내 보면 다음과 같다.

$$\mu_{hx} < \sqrt{\pi_{h1}} \sigma_{hz} / p_h. \quad (3.2)$$

식(3.2)는 제안한 층화 직접질문 조건부 양적속성 무관질문모형의 효율성이 기본적으로 μ_{hx} 의 값이 상대적으로 σ_{hz} 의 값에 비하여 작으면 작을수록 더 좋아진다는 것을 나타내고 있다. 또한, p_h 와 π_{h1} 의 값에도 의존하고 있으므로 궁극적으로는 제안한 층화 직접질문 조건부 양적속성 무관질문모형의 효율성이 p_h 와 π_{h1} 그리고 σ_{hz} 의 값에 의존한다는 것을 보여주고 있다.

한편, 식(3.2)에서 σ_{hz} / μ_{hx} 를 C_h 라고 하면, $C_h > p_h / \sqrt{\pi_{h1}}$ 와 같이 달리 표현할 수 있다.

C_h 의 값이 고정되어 있다고 가정하고 p_h 와 π_{h1} 의 크기에 따라 효율성이 어떻게 변하는지를 살펴보기 위하여 식(2.4)와 식(3.1)을 이용하여 제안한 층화 직접질문 조건부 양적속성 무관질문모형과 층화 양적속성 무관질문모형과의 효율성을 수치적으로 비교해 보고자 한다. 모집단이 2개의 층으로 되어 있고, $n = 1,000$ ($n_1 = 700, n_2 = 300$) 이고 $\mu_{hx} = 2, \sigma_{hz} = 4$ ($C_h = 2$) 와 $\mu_{hx} = 2, \sigma_{hz} = 3$ ($C_h = 1.5$) 라고 가정하고, π_{h1} 을 0.3에서 0.9까지, p_h 를 0.1에서 0.9까지 0.1간격으로 변화시켜가면서 두 모형에 대한 상대효율 $RE = V(\hat{\mu}_g) / V(\hat{\mu}_x)$ 를 계산한 결과 다음 Table 1을 얻었다. Table 1에서 1보다 큰 값은 제안한 층화 직접질문 조건부 양적속성 무관질문모형이 층화 양적속성

무관질문모형보다 더 효율적임을 나타낸다. Table 1로부터 층화 직접질문 조건부 양적속성 무관질문모형이 층화 양적속성 무관질문모형보다 전반적으로 효율적이었다. 특히 p_h 값이 0.1이고 π_{h1} 값이 0.3인 경우 층화 직접질문 조건부 양적속성 무관질문모형의 효율성이 가장 좋게 나타났음을 알 수 있었다. 한편, C_h 값이 클 때 층화 직접질문 조건부 양적속성 무관질문모형이 층화 양적속성 무관질문모형보다 효율적임을 알 수 있었다.

Table 1. Relative efficiency($V(\hat{\mu}_y)/V(\hat{\mu}_x)$)

C_h	$\pi_{h1} \backslash p_h$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
		2	0.3	3.2697	3.0927	2.8368	2.5423	2.2429	1.9607	1.7069
	0.5	1.9900	1.9607	1.9138	1.8518	1.7777	1.6949	1.6064	1.5151	1.4234
	0.7	1.4263	1.4198	1.4091	1.3944	1.3759	1.3539	1.3289	1.3011	1.2709
	0.9	1.1107	1.1097	1.1080	1.1056	1.1026	1.0989	1.0945	1.0895	1.0840
1.5	0.3	3.2219	2.9284	2.5423	2.1462	1.7880	1.4851	1.2373	1.0376	0.8771
	0.5	1.9823	1.9313	1.8518	1.7509	1.6363	1.5151	1.3931	1.2747	1.1627
	0.7	1.4246	1.4131	1.3944	1.3689	1.3375	1.3011	1.2605	1.2166	1.1705
	0.9	1.1105	1.1086	1.1056	1.1014	1.0960	1.0895	1.0820	1.0734	1.0638

4. 결론

본 논문에서는 모집단이 층으로 구성되어 있고 얻고자 하는 민감한 속성에 대한 정보가 양적속성일 때, 덜 민감한 속성에 대하여 응답자들에게 직접질문을 하여 “예”라고 응답한 사람들에게만 민감한 변수 X 와 무관한 변수 Y 를 포함하고 있는 Greenberg et al.의 양적속성 무관질문모형을 사용하도록 하여 민감한 변수 X 에 대하여 층화 추정할 수 있는 층화 직접질문 조건부 양적속성 무관질문모형을 제안하였고, 제안한 층화 직접질문 조건부 양적속성 무관질문모형과 Greenberg et al.의 양적속성 무관질문모형에 층화추출법을 적용한 층화 양적속성 무관질문모형과의 효율성을 비교한 결과 제안한 층화 직접질문 조건부 양적속성 무관질문모형이 더 효율적이었다.

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., Kubler, R. R., Abernathy, J. R., Horvitz, D. G. (1971). Applications of the RR technique in obtaining quantitative data, *Journal of the American Statistical Association*, 66, 243-250.
- 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.

Adaptive Weighted Total Variation Penalty for Precise Change Point Detection^{*}

Dong-Young Lee¹, Kwan-Young Bak², Jae-Hwan Jhong³

Abstract

Total variation-based methods, such as the fused lasso, are standard for change point detection but are impaired by issues like local monotonicity. To address these limitations, this study comparatively analyzes the fused lasso with two alternative methodologies. These methodologies are the Penalized Regression Spline (PRS) and a novel Adaptive Penalized Regression Spline (APRS) that incorporates data-adaptive weights. Using an efficient coordinate descent algorithm, we evaluated the three methods on both simulated and real-world data. The comparative analysis, conducted under various Signal-to-Noise Ratio (SNR) conditions, revealed that APRS demonstrated superior performance, particularly in minimizing false positives.

Keywords : B-splines; Coordinate Descent Algorithm; Fused Lasso; Weighted Total Variation.

^{*}The work of Kwan-Young Bak was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00342014 and RS-2022-00165581). The work of Jae-Hwan Jhong was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(RS-2024-00342014 and RS-2024-00440787).

¹Master's student, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea. E-mail: dtsy5891@gmail.com

²Assistant professor, School of Mathematics, Statistics and Data Science, Sungshin Women's University, Seoul 02844, Korea. E-mail: kybak@sungshin.ac.kr

³(Corresponding Author) Associate Professor, Department of Statistics, Chung Buk National University, Cheongju 28644, Korea. E-mail: jjh25@chungbuk.ac.kr

국민건강영양조사 가속도계 자료의 수면시간 추정: 미착용 시간 대체 방법과 수면 시간 알고리즘 비교*

이민영^{1†}, 윤나영^{2†}, 최형윤^{3†}, 지희정⁴

요 약

수면 시간은 비만, 심혈관 질환, 정신 건강 등 다양한 건강지표와 연관된 핵심적인 생활습관 요인이다. 특히, 부족하거나 과도한 수면은 우울증 및 사망률 증가와 밀접한 관계가 있다. 이러한 연관성을 평가하기 위해 정확한 수면 시간의 측정은 공중보건학적으로 중요하다. 최근 들어, 웨어러블 기기를 활용한 객관적인 수면 측정이 이루어지고 있으나, 관찰연구에서는 기계적 결함, 장치 미착용 등으로 인하여 데이터 손실이 발생하는 문제가 있다. 이에 본 연구는 2016년 국민 건강영양조사 자료를 활용하여, 낮 시간(09:00~21:00)에 발생하는 미착용 구간을 결측치로 정의하고, 이를 처리하기 위해 다중 대체 방법을 적용했다. 구체적으로 Poisson log-normal 분포 기반의 ZIPLN(Zero-Inflated Poisson log-Normal) 및 ZIPLN+PMM(Predictive Mean Matching)방법, 그리고 다변량 정규분포 기반의 MICE(Multivariate Imputation by Chained Equations) 및 MICE+PMM방법을 적용했다. 적절한 대체 후, 1축 데이터를 활용한 수면 알고리즘인 Sadeh 및 Cole-Kripke 알고리즘을 사용하여 수면 상태(Sleep/Wake)를 구분한 후, Tudor 알고리즘의 파라미터를 다르게 하여 하루 평균 수면 시간을 비교하였다. 그 결과, Cole-Kripke 알고리즘으로 계산한 수면시간이 Sadeh 알고리즘으로 계산한 결과보다 크게 나왔다. 또한 원자료 수면시간이 대체된 데이터보다 더 길게 나타났으며, 이는 미착용 구간을 적절히 처리하지 않을 경우 수면시간이 과대추정될 수 있음을 시사한다.

주요 용어 : 결측 데이터, 대체 방법, 가속도계 데이터, 수면 알고리즘.

*This work was supported by the research grant of the Chungbuk National University in 2023. This research was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00354881)

† These authors contributed equally to this work.

¹28644 충북 청주시 서원구 충대로1, 충북대학교 자연과학대학 통계학과 석사과정,
E-mail : lminy0226@naver.com

²28644 충북 청주시 서원구 충대로1, 충북대학교 자연과학대학 정보통계학과 학부재학생,
E-mail : 3510yny@naver.com

³28644 충북 청주시 서원구 충대로1, 충북대학교 자연과학대학 정보통계학과 학부재학생,
E-mail : 01micchoi@gmail.com

⁴(교신저자) 28644 충북 청주시 서원구 충대로1, 충북대학교 자연과학대학 정보통계학과 조교수,
E-mail : heejung.jee@chungbuk.ac.kr

베이지안 신경망 기반 독성 예측에 대한 불확실성 정량화*

이석진¹, 김정연²

요약

의약 안전평가에서는 “정확한 예측”만큼이나 “예측의 확실성”이 중요하다. 특히 약물 유발 간손상과 같은 순서형 독성 등급은 임상적 의사결정과 직결되므로, 모델이 자신이 아는 것과 모르는 것을 구분해 알려주는 것이 필수적이다. 소표본, 분포의 불일치, 애매한 라벨이 혼한 현실에서 일반 신경망은 과신하기 쉽고, 이러한 과신은 잘못된 후보 선별과 불필요한 후속 실험을 야기한다. 본 연구는 베이지안 신경망을 적용하여 순서형 독성 예측의 불확실성을 해석 가능하게 정량화하는 절차를 제안한다. 입력은 전임상 지표로 구성하고, 출력은 임계값을 갖는 순서형 로지스틱 구조로 설계하였다. 사후샘플을 이용해 후행예측 분포를 계산하고, 엔트로피 기반 분해를 통해 총 불확실성을 모델에 기인한 불확실성(epistemic uncertainty)과 데이터 고유 불확실성(aleatoric uncertainty)으로 나누어 비교·시각화하였다. 두 가지 실험을 수행하였다. 첫째, 학습 자료만으로 모델을 적합하고 별도의 시험 자료를 평가하여, 학습 분포에서 멀리 떨어진 사례에서는 모델에 기인한 불확실성이 커지고, 결정 경계에 가까운 사례에서는 데이터 고유 불확실성이 커지는 경향을 확인하였다. 둘째, 애매하게 분류된 사례를 소프트 라벨 형태로 학습에 포함하여 해당 영역을 평가한 결과, 그 영역의 데이터 고유 불확실성은 증가하고 모델에 기인한 불확실성은 감소하는 변화를 관찰하였다. 제안한 절차는 작은 표본에서도 “모르는 것은 모른다”고 말할 수 있는 근거를 제공하며, 분포 밖 입력 탐지, 후속 실험의 우선순위 결정, 애매 라벨 처리, 위험 소통 등 실무 의사결정에 직접 활용 가능하다.

주요 용어: 베이지안 신경망, 불확실성 정량화, 화학물질 독성 예측, 약물 유발 간손상(DILI).

*본 연구는 환경부 주관 2025년도 「화학물질 안전관리 특성화대학원」 사업을 통한 성과물임을 밝힙니다.

¹02504 서울특별시 동대문구 서울시립대로 163, 시립대학교 인공지능학과 석사과정.

²(교신저자) 02504 서울특별시 동대문구 서울시립대로 163, 시립대학교 인공지능학과 부교수.

OECD 국가의 성평등 태도 비교: 세계가치조사(WVS) 자료 기반 분석 *

이아름¹

요약

본 연구는 세계가치조사(World Values Survey, Wave 7) 마이크로데이터를 활용하여 OECD 16 개국의 성평등 태도 구조와 개인 수준 요인을 분석하였다. 성별, 교육수준, 가족가치 인식 세 변수를 핵심 예측 요인으로 설정하고, 이들이 성평등 관련 응답 패턴과 어떠한 연관성을 갖는지를 검토하였다. 성평등 태도는 고용 우선권(Q118), 정치적 리더십(Q119), 대학교육 선호(Q120), 취업 모에 대한 관점(Q121) 네 문항을 기반으로 했으며, 각 문항의 척도 차이를 조정하기 위해 역코딩과 0-1 범위 변환을 실시한 후 평균값을 산출하여 GE(Gender Equality) 지수를 구성하였다. 자료 처리와 분석은 R을 활용해 수행하였다. 전체 분석 대상은 28,377명이며, 교육수준은 국가별 교육 체계 차이를 고려해 각 국가의 내부 분포에 따라 세 구간(저·중·고)으로 재분류하였다. 기초 분석 결과, OECD 국가 간 성평등 태도의 수준과 분포에는 뚜렷한 차이가 나타났다. 전반적으로 여성은 남성보다 더 평등한 태도를 보였지만, 일부 국가에서는 남성이 더 높은 평등 성향을 보이는 예외도 확인되었다. 교육수준이 높을수록 평등 태도가 강하게 나타났으며, 이는 고학력 집단에서 전통적 성역할에 대한 지지가 약하다는 기존 연구 결과와 일치한다. 가족 가치는 대부분의 국가에서 전통적 성규범과 정적인 관련성을 보였고, 가족의 중요성을 높게 평가하는 집단일수록 평등 태도가 낮은 경향이 관찰되었다. 다만 성별과 교육 수준에 따라 이러한 연관성의 크기와 방향이 달라, 가족 가치가 성평등 태도에 미치는 영향이 사회집단별로 비대칭적으로 작동함을 시사한다. 본 연구는 WVS 국제비교 자료를 기반으로 OECD 국가의 성평등 태도 구조와 개인 수준 요인의 전반적 경향 및 국가 간 차이를 제시했다는 점에서 의의가 있다. 향후에는 가중치 적용, 국가수준 변수를 고려한 다층모형 분석 등을 통해 보다 정교한 비교 연구가 이루어질 필요가 있다.

¹창원대학교 의류학과

느타리 수확후배지의 재활용 가치 평가: 축사갈개 활용을 중심으로

이정명¹, 이원석², 김연진³, 이수연⁴

요 약

버섯 수확후배지는 전국 연간 약 88만 톤 발생하여 다양한 활용 가능성이 있음에도 불구하고 현행 폐기물관리법상 동·식물잔재물로 분류되어 농가의 자체 처리 및 재활용에 제약이 있다. 그러나 수확후배지는 버섯 재배 과정에서 배지 내 영양원의 약 15~25% 사용되고 상당량의 영양원이 잔존하기 때문에 사료나 비료 등으로의 재활용이 가능하다. 특히 느타리버섯 수확후배지는 톱밥 비율이 높아 건조 후 축사 갈개로 재활용할 수 있으며, 이는 농가가 폐기물 처리비용을 절감하는 동시에 추가 수익을 얻을 수 있는 경제적·환경적으로 우수한 자원순환 모델이다. 본 연구에서는 느타리 수확후배지의 폐기물 처리비용과 재활용 비용을 비교하여 사적·사회적 편익을 평가하고, 정책 제안할 수 있는 자료로 사용하고자 하였다. 분석에는 비용·편익(B/C) 분석법을 적용하였으며, 주요 비용 항목으로는 수집비용, 소각 및 매립비용, 재활용 공정비용을, 주요 편익 항목으로는 재활용 제품 판매수입, 재생원료 사용에 따른 절약액, 수집 및 소각·매립 과정에서 발생하는 환경비용을 설정하였다. 사적 비용은 재활용 수거비와 재생공정상 발생한 재활용 공정비용의 합으로 산출하였으며, 사회적 비용은 환경부 폐기물처분부담금을 적용하여 계산하였다. 사적 편익은 재활용 제품으로 판매한 수입으로 산출하였으며, 사회적 편익은 현재 우사, 계사에서 사용하는 갈개 비용을 재생원료 사용에 따른 에너지 절약액으로 산정한 것에 재활용되지 않았을 경우 발생하는 폐기물 처리비용을 합산하여 산출하였다. 느타리 수확후배지를 축사갈개로 재활용할 경우의 순 편익은 428,080원/톤으로 분석되었다. 느타리 수확후배지의 전국 연간 발생량 58,870톤을 기준으로 할 때 연간 약 252억 원의 사회적 순편익이 발생하는 것으로 나타났다. 향후 30년간 동일한 재활용 수준이 지속된다고 가정하고 연 5.5%의 할인율을 적용할 경우, 총 사회적 편익의 현재가치(NPV)는 약 3,528억 원으로 추정되었다. 연구 결과 느타리 수확후배지의 축사갈개 재활용은 경제적·환경적 편익이 매우 큰 것으로 나타났으며 향후 수확후배지의 재활용 및 자원화 촉진을 위한 정부와 사회의 지원이 필요함을 시사한다.

주요 용어 : 느타리 수확후배지, 농산부산물 재활용, 비용·편익 분석.

¹18388 경기도 화성시 병점중앙로 283-33, 경기도농업기술원 농업연구사. E-mail: jmlee@gg.go.kr

²18388 경기도 화성시 병점중앙로 283-33, 경기도농업기술원 농업연구관. E-mail: born815@gg.go.kr

³18388 경기도 화성시 병점중앙로 283-33, 경기도농업기술원 농업연구사. E-mail: boong0323@gg.go.kr

⁴18388 경기도 화성시 병점중앙로 283-33, 경기도농업기술원 농업연구관. E-mail: onpickle@gg.go.kr

청소년기 사회적 관계 다중궤적 분석: Group-based multi-trajectory model의 적용

장은아¹, 정혜원²

요약

본 연구는 청소년기의 부모·교사·친구·다문화 이웃 및 친구와의 관계를 동시에 포함한 사회적 관계 발달이 시간의 흐름에 따라 어떠한 이질적 경로를 보이는지 확인하기 위하여 group-based multi-trajectory model(GBMTM)을 적용하였다. GBMTM은 단일한 종속변인의 변화만을 분석하는 group-based trajectory model의 확장으로, 여러 지표변인의 변화궤적을 하나의 모형에서 동시에 추정함으로써 지표변인 각각의 발달궤적뿐만 아니라 지표변인 간의 공통 변화 패턴을 파악할 수 있는 기법이다(Nagin et al., 2018). 분석에는 한국교육개발원에서 수집한 한국교육종단연구2013 3~8차년도(2015~2020, 중1~고3) 자료(n=5,317)를 활용하였으며, 종속변인은 부모관계, 교사관계, 친구관계, 다문화 이웃·친구관계의 네 가지 사회적 관계 지표로 구성하였다. 또한, 청소년과 관련한 개인, 가정, 학교 관련 변인을 투입하여 사회적 관계 다중궤적 분류에 이들 변인이 어떠한 영향을 미치는지 분석하였으며, 다중궤적에 따른 성인기 결과를 확인하기 위하여 시민의식, 도덕성 및 생애목표의식을 원격 결과변인으로 설정하였다. 분석 결과, 청소년의 사회적 관계 변화에는 6집단 모형이 가장 적합한 것으로 판단하였으며, 구체적으로 전반적 저수준 유지형(17.4%), 부모관계 취약·다문화관계 친화형(13.1%), 전반적 중수준 유지형(35.3%), 전반적 상승형(12.0%), 전반적 감소형(14.4%), 전반적 고수준 유지형(7.8%)의 6개 집단 양상을 확인할 수 있었다. 성별, 자아개념, 공동체의식, 정신건강, 부모지원 및 교사열의 등의 변인이 다중궤적 분류에 유의한 영향을 미치는 것으로 나타났다. 아울러, 저수준·중수준·고수준 집단의 경우 결과변인 역시 전반적으로 상·중·하 수준의 경향을 보였으며, 특히 부모관계 취약·다문화관계 친화형의 경우 시민의식, 도덕성, 자기성장 등의 변인이 상대적으로 높게 나타났다. 본 연구는 집단중심 다중궤적모형을 적용하여 청소년기의 사회적 관계 발달이 단일하고 선형적인 경로가 아니라, 다차원적이고 비선형적인 궤적을 통해 변화함을 실증적으로 제시하였다는 점에 의의가 있다.

주요 용어 : 청소년, 사회적 관계, group-based multi-trajectory model, 한국교육종단연구2013.

¹34134 대전시 유성구 대학로 99, 충남대학교 교육학과 박사수료. E-mail: rhcsla456@naver.com

²(교신저자) 34134 대전시 유성구 대학로 99, 충남대학교 교육학과 교수. E-mail: chw7@cnu.ac.kr

고혈압이 있는 중년여성의 주관적 건강상태에 미치는 영향요인*

전은미¹

요 약

본 연구는 고혈압이 있는 중년여성의 주관적 건강상태와 주관적 건강의 영향요인을 확인하고 그에 따른 폐경 후 중년여성의 주관적 건강관리를 위한 기초자료를 제공하고자 한다. 본 연구대상자는 2022~2023년 국민건강영양조사 자료에서 중년여성(40~59세)은 2,262명, 그 중 고혈압 중년 여성 308명을 대상으로 분석하였다. 수집된 자료는 SPSS/WIN 25.0을 이용하여 복합표본에서의 분석방법을 적용하였다. 고혈압이 있는 중년여성의 주관적 건강은 복합표본 빈도분석과 기술통계를 이용하였고 각 변수에 따른 주관적 건강과 영향요인은 복합표본 일반선형모형을 이용하여 분석하였다. 연구결과 국내 고혈압이 있는 중년여성의 주관적 건강은 2.83점(1~5점 척도)이었다. 주관적 건강에 통계적 차이를 보이고 있는 요인으로는 교육수준, 배우자유무, 근력운동, 비만, 우울로 나타났다. 고혈압이 있는 중년여성의 주관적 건강에 영향을 미치는 요인을 확인하기 위해서 다중회귀분석을 한 결과, 주관적 건강에 통계적으로 유의한 영향요인은 교육수준, 근력운동, 비만유무, 우울로 나타났다. 교육수준에서는 초졸이 다른 범주에 비해 주관적 건강인식이 높았으며, 근력운동은 일주일에 5일 이상 하는 것에 비해 1~2일 운동이 주관적 건강이 낮게 나타나고, 비만인 경우 주관적 건강이 낮았고 우울감이 한 단위 오르면 주관적 건강은 -0.05만큼 내려감을 볼 수 있다. 이상의 결과를 바탕으로 다각적이고 개별적인 간호중재프로그램 개발이 필요하다.

주요 용어 : 고혈압, 주관적 건강, 여성, 일반선형모형.

*이 논문은 2025학년도 배재대학교 교내학술연구비 지원에 의하여 수행됨.

¹35345 대전광역시 서구 배재로 155-40(도마동), 배재대학교 간호학과 교수. E-mail: charminggold@pcu.ac.kr

Wavelets Scattering Meet Transformers: Detecting Anomalies in Robotic Trajectories^{*}

Yunchae Jung¹, Im Jiwon², Haeun Ryu³, Minsu Park⁴

Abstract

Modern robots generate a wealth of multivariate sensor signals, yet detecting when these trajectories go wrong remains challenging. In this work, we propose an unsupervised framework for abnormal trajectory detection using wavelet scattering and Transformer networks. Unlike vision-based approaches, our method focuses purely on trajectory-level sensor dynamics. Using the DROID (Dataset for Robot Instruction and Demonstration), we model normal motion from joint positions and velocities without supervision. The Wavelet Scattering Transform (WST) provides stable, physics-informed representations of temporal-frequency structure, which are then fed into a transformer-based time-series encoder to capture long-range dependencies across sensors and time. The model is trained exclusively on normal trajectories and evaluated through controlled injections of drifts, spikes, and temporal delays. Results indicate that the WST-transformer combination effectively identifies subtle deviations in robotic motion, outperforming baseline unsupervised detectors. The proposed framework offers a promising step toward robust, sensor-level anomaly monitoring and predictive maintenance in autonomous robotic systems.

Keywords : statistical signal processing, wavelet scattering transform, transformer time-series models, unsupervised anomaly detection, robotic systems.

^{*}This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (RS-2022-NR068754 and RS-2024-00337752).

¹Master's student, Department of Statistics and Data Science, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea. E-mail: bella55560@gmail.com

²Master's student, Department of Statistics and Data Science, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea. E-mail: jiwonlim0492@gmail.com

³Undergraduate Student, Department of Information and Statistics, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea. E-mail: haeunryu104@gmail.com

⁴Assistant professor, Department of Statistics and Data Science, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea. E-mail: minsu.park51@gmail.com

Sequence Kernel Association Test (SKAT) for Genetic Variant Identification Based on MMSE Scores

Jingyeong Jeong¹, Eunjee Lee¹

Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline and memory impairment, representing the leading cause of dementia in older adults. Both environmental and genetic factors contribute to AD onset, yet conventional genome-wide association studies (GWAS) that test single nucleotide polymorphisms (SNPs) individually often fail to capture rare variant effects or gene-gene interactions. These limitations contribute to the issue of "missing heritability," underscoring the need for gene-set based analytical methods that can enhance genetic explanatory power. In this study, we analyzed genomic and clinical data from the Alzheimer's Disease Neuroimaging Initiative 2 (ADNI-2) cohort to explore genetic associations with cognitive function, as measured by Mini-Mental State Examination (MMSE) scores. After rigorous quality control (QC) and genotype imputation, we constructed linkage disequilibrium (LD) blocks to capture local genetic structure and applied the Sequence Kernel Association Test (SKAT) to perform SNP-set level association analysis. Significant association signals were identified on chromosomes 18 and 19, with the most notable block on chromosome 19 showing genome-wide significance ($p=1.93\times 10^{-6}$). This region was located near the GNA11 gene, which encodes a G-protein subunit involved in intracellular signaling. Prior research has implicated GNA11 in vascular and neuronal regulation, suggesting its potential role in the pathophysiology of AD. Overall, this study highlights the utility of SKAT in detecting gene-set level associations that may be overlooked by single-variant analyses and provides insights into novel genetic factors underlying cognitive decline. Future integration of multi-phenotype and functional genomic analyses could further clarify the biological mechanisms linking these loci to AD progression.

¹Chungnam National University, department of information and statistics, Daejeon, Korea.

UN 지속가능발전목표(SDG) 달성을 위한 국가연구개발사업의 데이터 기반 측정 프레임워크*

최봉기¹, 노경란²

요 약

지속가능발전목표(SDGs, Sustainable Development Goals)는 유엔이 제시한 전 지구적 과제로서, 우리나라 또한 국가 중장기 정책의 목표로 설정하고 있다. 본 연구는 국가연구개발(R&D) 투자 관점에서 SDG 실현을 위한 정부의 연구개발투자활동을 정량적으로 측정하는 방법론을 제시한다. 한국과학기술정보연구원의 국가과학기술지식정보서비스(NTIS)에서 수집한 연구개발 과제의 데이터를 대상으로, 대규모 언어모델(LLM)을 적용하여 과제별 SDG 적합성을 분류하였다. 분류의 신뢰도를 확보하기 위하여, LLM 결과를 텍스트 기반의 전통적 주제 분류 기법과 비교·보완하는 혼합적 접근을 사용하였다. 분석 결과 한국의 연구개발 과제는 G9(사회기반시설, 산업화 및 혁신)에 전체 투자액의 25% 이상 집중되어 있으며, 이 뒤를 이어 G8(경제성장 및 양질의 일자리), G12(지속가능한 소비와 생산)이 각각 10% 수준을 차지 하였다. 과제수 기준으로 G9, G3(건강과 웰빙 증진), G12, G8의 합계가 전체의 과반을 점유하였다. 도출된 17개 SDG의 분포를 주제별 유사성에 근거하여, 성장(Growth), 전이(Transition), 제3의 길(ThirdWay) 세 축으로 구분한 거시 경향 분석 결과에서 우리나라의 연구개발 체제는 성장 중심의 담론이 구축 되어 있음을 확인하였다. 추가적으로 새로운 과학정책의 가능성을 타진하기 위하여, 연구개발단계, 세부과제지원유형, 연구수행주체별로 SDG 분포의 패턴을 분석 하였다. 이러한 데이터 기반 측정 프레임워크 구축 방법과 결과는 부처별 연구개발 사업 기획에서 지속가능성 지표 설계와 달성을 위한 전략적 투자 방향 수립에 기여할 것으로 기대된다.

주요 용어: 지속가능발전목표, 대규모 언어모델, 과학기술정책, 데이터 분석.

*이 논문은 2025년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다
(과제번호: K25L4M2C3-01).

¹02456 서울특별시 동대문구 회기로66 한국과학기술정보연구원 글로벌R&D분석센터 과학계량분석연구팀 선임 연구원. E-mail: boongkee@kisti.re.kr

²2456 서울특별시 동대문구 회기로66 한국과학기술정보연구원 글로벌R&D분석센터.

국가거점국립대 신입생 중도탈락 예측 모형 개발 및 활용 방안*

최원기¹, 김민수²

요 약

본 연구에서는 국가거점국립대학교인 C 대학교를 대상으로 신입생 중도탈락에 선제적으로 대응하기 위한 예측 모형을 개발하였다. 설명 변수는 입학 정보, 단과대학, 입시전형 및 주소지, 학사정보의 네 범주로 구성하였으며, 학사정보의 유무에 따라 1학년 1학기 모형(학사정보 미포함)과 1학년 2학기 모형(학사정보 포함)으로 구분하여 분석을 수행하였다. 단변량 로지스틱 회귀분석을 통해 각 변수와 중도탈락 간의 관계를 파악하고 회귀계수를 도출한 후, 변수 간 상호작용에 따른 회귀계수 부호 왜곡을 방지하기 위해 단변량 분석 결과와 일관된 방향으로 계수를 조정하였다. 또한 각 범주별 회귀모형의 Area Under Curve(AUC)를 기반으로 가중치를 부여하여 최종 중도탈락 예측 모형을 구축하였다. 본 모형은 도출된 중도탈락 확률을 바탕으로 대학에서 원하는 기준에 따라 신입생을 위험군과 비위험군으로 분류할 수 있어, 대학 차원의 선제적 개입 및 관리 전략 수립에 활용 가능하며 본 연구에서는 각 범주별 활용 가능 정책을 제시하고 있다. 아울러 회귀계수 해석을 통해 중도탈락에 영향을 미치는 주요 요인을 정량적으로 설명하였으며, 이러한 연구 결과는 C 대학교 실무 정책에 적용 중이며 향후 타 국가거점국립대학교로의 확장되어진다면 지역 소멸과 수도권 집중 현상의 위기를 해결할 수 있는 방안이 될 것으로 기대한다.

주요 용어 : 국가거점국립대학교, 로지스틱 모형, 선제적 대응, 신입생 중도탈락.

¹61186 광주광역시 북구 용봉로 77, 전남대학교 수학과/통계학과 석사과정. E-mail: wongi8994@gmail.com

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

다중 코딩 에이전트 협업 스펙트럼 분석: 논문 필터링을 위한 다중 코딩 에이전트 구성 비교 연구*

한혁¹

요약

본 연구는 LLM 기반 코딩 에이전트의 협업 구성 수준이 계획 수립과 실행 결과에 미치는 영향을 체계적으로 비교하기 위해 동일한 데이터셋과 동일한 코드 자동화 파이프라인에서 (1) 단독(Claude Code), (2) 이중 협업(Claude Code + Codex CLI), (3) 삼중 협업(Claude Code + Codex CLI + Gemini CLI)의 세 가지 시나리오를 설정하여 실험을 진행하였다. 에이전트 간 상호작용은 Model Context Protocol (MCP)을 기반으로 구현하여, 역할 분화(구현 중심·패턴 인식·비판적 검증)와 에이전트 간 합의를 위한 채널 구조를 표준화하였다. 주요 평가지표는 선별 논문 수, 선별 임계값 등으로 구성하였다. 실험 결과, 이중 협업 시나리오는 확장 용어집·다단계 퍼널(고재현→정밀) 등을 적용함으로써 128/500 편(25.6%)의 최고 효율을 달성하였으며, 단독 수행 대비 12.8배 향상되었다. 반면, 삼중 협업 시나리오는 독립 평가 → 가중 합의 → 심의 구조와 보수적 veto 기제를 통해 정밀도와 신뢰성을 극대화했으나, 최종 선별은 14/500 편(2.8%)으로 효율성이 낮게 나타났다. 단독 구성 시나리오는 해석 가능성과 속도, 재현성 면에서 강점을 보였으나 표현 변이 및 문맥 신호 포착에 한계가 있었다(10/500 편, 2.0%). 종합적으로, 탐색·확장 국면에서는 이중 협업이, 품질·검증이 핵심인 국면에서는 삼중 협업이, 빠른 프로토타입 단계에서는 단독 구성이 각각 적합한 접근으로 분석되었다. 이를 통해 본 연구는 인공지능 협업 환경에서 계획 수립의 정교화와 역할 기반 합의 프로토콜 설계를 위한 실증적 근거를 제시하며, 향후 다중 에이전트 협업의 효율-정밀성 균형을 규명하는 연구 방향을 제안할 수 있을 것으로 기대된다.

주요 용어: 에이전트, 에이전트협업, 멀티에이전트, 코딩에이전트.

*이 논문은 2025년도 한국과학기술정보연구원(KISTI)의 기본사업으로 수행된 연구입니다
(과제번호: K25L4M2C3-01).

¹02456 서울특별시 동대문구 회기로66 한국과학기술정보연구원 글로벌R&D분석센터 과학계량분석연구팀 책임 연구원. E-mail: hyhahn@kisti.re.kr

AOP-Wiki 그래프 데이터 기반 Key Event 경로 자동 탐색을 위한 데이터의 경로 구조 분석과 확장 가능성 탐색*

황신정¹, 강민설², 최호식³

요 약

본 연구는 독성학 분야에서 문헌 리뷰 근거를 기반으로 구축되어있는 독성발현경로 (AOP, Adverse Outcome Pathway) 공개 데이터인 AOP-wiki 를 그래프화 하여 기존에 밝혀져 있지 않은 주요사건 (KE, Key Event) 간의 잠재적 경로를 탐색하는 방법을 제안한다. 기존에는 도메인 전문가가 특정 사례에 해당하는 AOP 를 선별하고 공통적으로 사용되는 KE를 중심으로 연결하는 단순화된 경로를 제시하는 방식이 주로 사용되었다. 하지만 이러한 접근은 주관적 판단에 의존하며 잠재적 연결 관계를 체계적으로 확장하기 어렵다는 한계가 있다. 본 연구에서는 동적 계획법 (DP, Dynamic Programming)을 적용하여 엣지 가중치를 기반으로 증거를 최대화 하는 경로를 탐색하였다. 엣지 가중치는 각 관계의 인과성을 반영하기 위해 AOP-wiki 내부의 Weight of Evidence (WoE), LLM 을 활용한 KE 관계 파악 점수를 종합적으로 사용하여 산출하였다. 최종적으로 기존의 밝혀진 AOP 경로와 새롭게 탐색 된 AOP 의 엣지 수준을 기반으로 증거를 최대화 하는 경로를 비교 후 순위화하여 새롭게 탐색 된 경로의 순위가 기존의 밝혀진 가중치보다 높은 탐색 된 경로를 확인한다.

주요 용어 : AOP, 독성발현경로, Dynamic Programming, 그래프 탐색, 엣지 가중치, 순위화.

*[생활화학제품 안전관리 기술개발 사업 과제번호 : RS-2023-00215309]에 의해 지원받았음.

¹02504 서울시 동대문구 시립대로 163, 서울시립대학교 일반대학원 도시빅데이터융합학과, 석사과정.

²02504 서울시 동대문구 시립대로 163, 서울시립대학교 일반대학원 도시빅데이터융합학과, 박사과정.

³02504 (교신저자) 서울시 동대문구 시립대로 163, 서울시립대학교 일반대학원 도시빅데이터융합학과.

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

Profiling the User of Real Estate Tax Tech Market^{*}

*Dongjae Kim¹, Yunhui Kim², Goeun Kang³,
Junho Shin⁴, Donghyeon Kim⁵, HyungBin Moon⁶*

Abstract

The recent acceleration of digital transformation and the National Tax Service's introduction of its 'One-Click Service' are driving the expansion of the domestic tax tech market. Nevertheless, empirical research analyzing actual service users based on data remains limited. Therefore, this study compared nationwide real estate tax filing data (capital gains tax) with the filing data of tax solution service users. It then converted service users' tax filing history into RFM (recency, frequency, monetary) variables and applied the K-means clustering method to segment them into subgroups. The goodness-of-fit test revealed significant differences in the distribution of service user filings compared to all filers in terms of age and gender. Notably, filings from users in their 40s and 50s were significantly higher than expected, while those from users 60 and older were lower. Cluster analysis based on filing patterns (RFM) confirmed that the market consists of 5 distinct user clusters. The findings are significant as they provide crucial empirical evidence for tax tech companies to develop data-driven, customized services optimized for each segment's needs and establish effective marketing strategies.

Keywords : Consumer Segmentation, User Profiling, Tax Tech Market, Online Real Estate Tax Solutions; K-Means Clustering.

^{*}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 2025(2024-0-00018).

¹Master's Student, Department of Data Engineering, Pukyong National University, 45, Yongso-ro, Nam-gu, Busan 48513, Republic of Korea. Email: dongjaekim@pukyong.ac.kr

²Undergraduate Student, Department of Computer Engineering, Pukyong National University, 45, Yongso-ro, Nam-gu, Busan 48513, Republic of Korea. Email: seizethedayunhui@gmail.com

³Undergraduate Student, Department of Computer Engineering, Pukyong National University, 45, Yongso-ro, Nam-gu, Busan 48513, Republic of Korea. Email: kke02093@gmail.com

⁴Master's Student, Department of Data Engineering, Pukyong National University, 45, Yongso-ro, Nam-gu, Busan 48513, Republic of Korea. Email: 202455077@pukyong.ac.kr

⁵President, NEWEYE Corp. Co., Ltd. 365, Sinseon-ro, Nam-gu, Busan 48513, Republic of Korea. Email: kdh@taxai.co.kr

⁶(Corresponding Author) Associate Professor, Department of Data Engineering, Pukyong National University, 45, Yongso-ro, Nam-gu, Busan 48513, Republic of Korea. Email: hbmoon@pknu.ac.kr

Streamlit-based automated analysis platform for real-time calcium imaging in freely moving mice

Hyeokjun Kwon¹, Jaesuk Yum²

Abstract

High-resolution, real-time visualization of deep brain activity in freely moving animals remains a considerable challenge in neuroscience. To address this, we employed fiber-optic confocal (FOC) fluorescence endomicroscopy to monitor hippocampal Ca²⁺ dynamics in freely moving GCaMP6s transgenic mice and developed a Streamlit-based automated analysis platform for intuitive visualization. The Cell-vizio system provided a ~20 μm field of view at 12 frames/s, and hippocampal Ca²⁺ signals were recorded under saline, 1 mg/kg, and 2 mg/kg bicuculline administration (i.p.). The work-flow integrates motion correction via jNoRMCorre and ROI segmentation refined by perimeter-guided watershed methods. The Streamlit interface enables complete preprocessing, clustering, and visualization without coding, allowing users to generate Ca²⁺ intensity plots (mean \pm SEM) and spatial cluster maps with minimal manual input. This platform revealed clear dose-dependent hippocampal dynamics: 1 mg/kg bicuculline induced ~50-fold signal increases during myoclonic jerks, whereas 2 mg/kg elicited a distinct increase within two each other spatial ensemble in the dentate gyrus. This Streamlit-based system enhances accessibility and reproducibility in Ca²⁺ imaging analysis and may serve as a scalable framework for data-driven neuropharmacological and medical research.

Keywords : Bicuculline, Cell-vizio, Freely moving, jNoRMCorre, Streamlit.

¹Graduate student, College of Pharmacy, Chungbuk National University, 194-31 Osongsaengmyeong 1-ro, Heungdeok-gu, Cheongju-si, Chungbuk 28160, Republic of Korea. E-mail: HyeokJunK@chungbuk.ac.kr
²(Corresponding Author)Associate Professor, College of Pharmacy, Chungbuk National University, 194-31 Osongsaengmyeong 1-ro, Heungdeok-gu, Cheongju-si, Chungbuk 28160, Republic of Korea.
E-mail: jyun@chungbuk.ac.kr

Music Therapy and Gardening Enhances Post-Discharge Adherence in Psychiatric*

*Wonsik Jung^{1,2}, Insu Jan^{3,4}, Hyeonjeong Lim⁵, Subin Jeong⁶, Hyun Ju Lee²,
Gahui Han⁷, Shin-Young Park², Hoyeong Jeong², Yaebyeon Kim²,
Soeun Lee², Eunjee Lee⁸, Kwang-Yeon Choi^{1,2}*

Abstract

Maintenance pharmacotherapy after acute hospitalization for schizophrenia, bipolar disorder, and major depression is essential for relapse prevention and functional recovery, yet medication adherence—measured by the proportion of days covered (PDC) and medication possession ratio (MPR)—averages only about 50%, ranging from 20% to 80%. While interventions such as medication education and motivational interviewing can improve adherence, their effects are often short-lived without continuous support. Horticultural and music therapy have been reported to reduce depression and anxiety, but their long-term effects on adherence remain unclear. This study evaluated the impact of these therapies on medication adherence and outpatient treatment retention among psychiatric inpatients at Chungnam National University Hospital. From July 2022 to July 2023, 139 patients who received weekly horticultural and music therapy were compared with 278 matched controls hospitalized between July 2020 and June 2022, matched 1:2 by gender, age, and diagnosis. After excluding those without post-discharge visits, 387 patients were analyzed. Medication adherence, assessed by MPR and PDC over 12 months, was analyzed using repeated-measures ANOVA, and outpatient retention was evaluated with a Cox proportional hazards model. The treatment group demonstrated significantly higher MPR at 4, 6, and 8 months after discharge, although the difference diminished at 10 and 12 months. Age and music therapy participation were significant predictors of adherence and retention, with young and middle-aged patients and music therapy participants showing greater adherence and longer outpatient follow-up. These findings suggest that music therapy during hospitalization enhances post-discharge medication adherence and continuity of care, particularly among younger and middle-aged psychiatric patients.

*This material was based on work partially supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant(RS-2022-NR068754). This research was supported by the National Research Foundation of Korea(NRF) grant funded by the Korean government (MSIT) (No. 2019M3E5D1A02068548;2020R1C1C1011378).

¹Chungnam National University, Department of psychiatry, College of medicine, Daejeon, Korea.

²Chungnam National University Hospital, Department of psychiatry, Daejeon, Korea.

³Chungnam National University, Department of statistics and data science, Daejeon, Korea.

⁴Korea Research Institute of Bioscience and Biotechnology, Korea Bioinformatics Center, Daejeon, Korea.

⁵Chungnam National University, Institute of Natural Sciences, Daejeon, Korea.

⁶Seoul National University Hospital, Biomedical Research Institute, Seoul, Korea.

⁷Chungnam National University, Department of mathematics, Daejeon, Korea.

⁸Chungnam National University, Department of information and statistics, Daejeon, Korea.

Predictive Modeling of Maritime Disasters Using Automated Machine Learning

Xin Zan¹, Jai Woo Lee¹

Abstract

Maritime disasters often lead to significant human and economic losses. Predicting the survival probability of a passenger using automated machine learning (AutoML) can enhance rescue planning and resource allocation efficiency. This study applies several machine learning algorithms – including Logistic Regression, Random Forest, and Support Vector Machine – within an automated modeling framework to the Estonian maritime disaster dataset to model and compare survival prediction performance. The analysis results show that XGBoost (optimized via AutoML and SMOTE) achieved the highest performance in terms of AUC and prediction stability. Considering the limited sample size of the original dataset, a larger dataset of the same type (the Titanic disaster dataset) was further employed to validate the best-performing automated model and evaluate its generalization capability. The experimental results demonstrate that the AutoML-based model maintains high accuracy and stability when applied to large-scale datasets. These findings reveal key factors influencing survival probability and confirm the robustness of the optimal automated model across datasets. This study can provide valuable insights for improving data-driven and automated rescue decision-making and contributes to the broader understanding of how AutoML techniques can be effectively applied to real-world survival prediction tasks.

Keywords : Artificial Intelligence, Big Data, Automated Machine Learning, Survival Prediction, Maritime Accidents

¹고려대학교 빅데이터사이언스학과

멜론 스마트팜 자립경영을 위한 목표소득 달성 시나리오 분석*

김효중¹, 신길호¹, 이승현², 하지영²

요 약

한국의 농가소득은 1995년 21,803천원에서 2022년 46,153천원으로 211.7% 증가한 반면, 경영비는 5,543천원에서 25,119천원으로 453.2%가 증가로 두배 차이를 보이고 있어[1] 문제가 심화되어 가고 있다. 최근 연구는 농업소득 연구([2], [3], [4], [5] 등)가 진행되고 있다. 특히, 스마트팜은 장기간 추진되고 있는 농업혁신 정책으로[6] 첨단 기술을 활용해 적은 노동력으로도 고품질 농산물을 생산할 수 있는 효율적인 시스템으로 현대 농업의 필수 요소로 자리 잡고 있다[7]. 이러한 시점에서 스마트팜을 도입한 경영체의 농가소득에 대한 연구가 활발히 진행되고 있다. 때문에 목표소득을 수립하고, 목표소득에 맞는 적정 재배면적과 그에 따른 경영비 투입과 재배 및 경영 노하우에 대한 연구가 필요한 시점이다. 한국은행은 2024년 1인당 국민총소득이 49,955천원이라고 밝혔는데[8], 목표소득으로 10,000만원을 달성해야 평균이상의 소득을 창출하여 지속가능한 농업으로 유지될 수 있을 것으로 판단되어 진다. 이에 멜론을 대상으로 목표소득 달성을 위한 경영성과표를 제시하고자 한다. 본 연구에서는 규범적 시나리오 기반에 정량적 시나리오 기법과 사례기반 시나리오를 활용하였다. 이를 위해 6단계의 절차를 수립하였다. 1단계는 구조화된 설문조사, 2단계는 구조화된 설문조사의 결과를 바탕으로 목표소득을 추출, 3단계는 구조화된 설문조사 결과를 바탕으로 농업소득(경영성과)을 분석 및 우수 경영체 추출, 4단계는 우수 경영체의 농업소득을 분석, 5단계는 우수 경영체의 초기 투자금을 추출, 6단계는 1-5단계 자료를 기반으로 한 10,000만원 목표소득의 달성 기준을 제시하였다. 설문지는 2종으로 스마트팜을 도입한 전남 31호를 대상으로 한 설문조사는 인구통계 9개항목, 목표소득 2개항목, 소득조사는 농촌진흥청 소득조사표 활용하여 구성하였다. 1단계로 설문조사의 조사시기는 2024년(23년~24년) 작기는 2024년 8월, 2025년(24년~25년) 작기는 2025년 8월에 2년(각 1개월)에 걸쳐서 조사하였다. 소득조사를 위한 조사기법에 대한 교육은 매년 2회를 실시하고, 1회차당 6시간 이상의 교육을 받은 스마트팜 생육조사 전문인력이 스마트팜 경영체를 직접 방문하여 조사하였다. 연구 결과를 달성하기 위해 2단계에서는 연간 10,000만원 목표소득을 달성하기 위해 노동력 집중기에는 고용노동을 활용하고, 그 외의 작업은 부부(2인) 노동 중심(전업농)으로 진행할 때에 이상적으로 판단하고 있는 적정 재배면적을 추출하였다. 그 결과, 전체 면적은 약 2,300평($m=2,306$, $sd=636.1$)이고, 그중에서 실제 재배면적은 약 2,100평($m=2,103$, $sd=581.1$) 정도를 인식하고 있는 것으로 분석되었다. 3단계에서는 농업소득 분석 및 우수 경영체를 추출하였다. 그 결과, 2024년대비 2025년도에도 총매출액 증감, 경영비 감소로 소득이 증가하는 사례도 있다. 이는 조사대상 경영체별로 재포기량, 생산량, 판매단가, 경영비, 생산 및 경영 노하우 등의 차이 폭이 크기 때문에 10a당으로 환산하여 목표소득을 제시에는 한계가 발생할 수 있다는 점에서 목표소득을 유지하고 있는 우수 경영

*본 논문은 농촌진흥청 지원을 받아서 작성됨.(RS-2024-00400125).

¹58213 대한민국 전라남도 나주시 산포면 세남로 1508, 전라남도농업기술원 자원경영과.

E-mail: plantmaniac@korea.kr

체 추출이 필요하다고 할 수 있다. 4단계에서는 2024년과 2025년의 경영성과에서 평균 2,100평 내외이면서 소득이 5,000만원 내외인 우수 경영체를 추출하였다. 멜론 재배작기는 특성상 2기작 또는 2모작(조사 작물비율 50%), 2기작+1모작 또는 3모작(조사 작물비율 33%)이므로 소득 10,000만원을 목표로 할 경우에는 최대 50%를 기준으로 하여 5,000만원을 적용하였다. 그 결과, 2년 연속으로 목표를 달성한 3개 경영체(시나리오 2)를 추출하였다. 5단계는 2년 연속으로 목표를 달성한 3개 경영체의 투자금 분석을 보면, 초기 투자금의 평균은 251,667,933원(토지 가격은 제외한 금액이고, 조사작물 비율을 100% 적용한 금액임)으로 나타났다. 비닐 등 기타재료비는 최근 투자금을 대입하였고, 나머지 소모성 기타재료비는 제외하였다. 경영체에서도 대농구 및 영농시설은 도입 시기가 순차적으로 차이가 있음을 고려하여 대입하였다. 초기 투자금과 앞서 분석한 경영분석에서 감가상각비를 계산해 보면, 평균 감가상각년수가 14.8년으로 나타나 유사성을 보이고 있다. 6단계에서는 전체면적 2,300평이면서 재배면적이 평균 2,100평을 기준으로 목표 소득이 10,000만원인 기준을 수립하였다. 총수확량은 최고치의 97% 수준, 판매단가는 약 90% 수준, 경영비는 105% 수준일 때에 목표 소득이 10,000만원이 추정된다(단, 감가상각비, 토지임차료, 자가노동비 및 용역비는 현 수준을 유지, 판매단가의 차이가 10%인 이유는 멜론의 작기별로 가격차이가 심하기 때문임) 따라서, 목표 소득(2기작)을 10,000만원을 추정하기 위해서는 총매출액 202,109,283원 이상의 수입을 달성해야 하며, 경영비 102,072,992원 이하로 지출되어야만 한다.

주요 용어 : 멜론 스마트팜, 목표소득, 경제적 자립, 적정 재배면적.

Improving Session-Based User Intent Prediction via Denoising-Enabled Hierarchical Multi-Task Learning

Peter Bin Sun¹, Young-jun Chor¹

Abstract

Detecting and utilizing user intent in recommender systems has been the important method for improving the quality of recommendations for increased user satisfaction. In this presentation, we propose a multi-task learning framework enhanced with denoising techniques to improve session-based user intent and item prediction. Our approach integrates a denoising module designed to mitigate noise and inconsistencies inherent in real-world user interaction data, with the hierarchical neural network architecture that estimates a user's latent intent and uses the intent prediction to predict the next item user is likely to be interested in (building upon IntentRec). By jointly optimizing intent and item prediction tasks with denoising process, our framework enhances the robustness and accuracy of latent intent inference and subsequent recommendation quality.

Keywords : Recommender Systems, User Intent Prediction, Denoising, Hierarchical Learning, Multi-Task Learning.

일사누적속도에 따른 급액제어를 통한 완전 순환식 양액공급 시스템 개발*

김효중¹, 신길호¹, 홍우진¹, 김주혜¹, 김동관¹

요약

시설원예에서 널리 사용되는 비순환식 수경재배는 급액이 작물의 실제 생육 요구와 관계없이 고정적으로 공급되는 구조로 인해 과도한 배액이 발생하고, 이로 인해 고농도의 폐양액이 외부 환경으로 유출되는 문제가 지속적으로 제기되어 왔다. 특히 전남 지역 주요 작목인 딸기와 토마토의 연간 폐양액 배출량은 각각 47,977톤과 33,625톤으로 매우 높은 수준으로 추정되었으며, 이는 수질오염과 양분 손실을 유발하는 주요 원인으로 지적되고 있다. 이러한 구조적 한계를 해결하기 위해서는 생육·환경 조건을 실시간 반영하여 급액량과 농도를 정밀하게 제어하고, 배액을 회수·재사용할 수 있는 완전순환식 수경재배 기술의 개발이 필요하다. 본 연구는 이러한 문제를 해결하고자 일사누적속도에 기반하여 급액량을 자동 조절하는 알고리즘을 설계하고, 이를 토마토와 딸기 재배에 적용하여 배액량과 생육에 미치는 영향을 검증하였다. 시험구에서는 일사구간(60~180 J/cm²/30min)에 따라 급액량을 관행 대비 10~45% 범위에서 감소하도록 조절하였으며, 급액·배액 특성, 생육 지표, 품질, 양분 조성 등을 종합적으로 평가하였다. 또한 비순환식 수경재배 농가의 급액·배액 분석을 병행하여 순환식 시스템 도입 필요성을 정량적으로 제시하였다. 연구 결과, 일사누적속도 기반 급액 조절기술은 급액량이 감소했음에도 급액 EC와 pH가 안정적으로 유지되었으며, 배액율은 관행의 46.2%에서 처리구 33.9~38.1% 수준으로 크게 감소하였다. 이는 광환경에 기반한 급액량 조절이 과도한 공급을 효과적으로 억제하고 배액 발생을 줄이는 데 높은 효율성을 지님을 의미한다. 생육 및 품질 또한 대부분의 지표에서 관행과 유사한 수준을 유지하여 급액량 감소가 생산성 저하로 이어지지 않는 것으로 나타났다. 토마토의 초장·엽수·엽장 등 모든 주요 생육 지표는 처리 간 큰 차이를 보이지 않았으며, 과중·당도·산도 등 품질 지표 또한 안정적으로 유지되었다. 한편, 비순환식 재배 농가의 배액 분석 결과에서는 급액 대비 배액의 K, Ca, Mg, SO₄ 등 주요 이온 농도가 1.5~2배 이상 증가하는 양상이 나타났으며, 이는 다량의 양분이 그대로 손실되고 환경으로 유출되고 있음을 의미한다. 이러한 분석은 비순환식 시스템에서 양분 축적과 폐양액 유출 문제가 필연적으로 발생한다는 사실을 정량적으로 입증한 것으로, 완전순환식 양액공급 기술 개발의 필요성을 강하게 뒷받침한다. 종합하면, 일사누적속도 기반의 급액 조절 기술은 생육 안정성을 유지하면서 배액 발생을 크게 줄일 수 있는 기술로 확인되었으며, 이를 기반으로 한 완전순환식 수경재배 시스템은 환경적·경제적·생산적 측면에서 높은 가능성을 가진 기술로 평가된다. 본 연구는 향후 스마트온실 기반의 지속가능한 양액관리와 양분 재활용 기술 개발을 위한 핵심적 근거를 제시하며, 순환식 수경재배 시스템의 실용화에 중요한 기여를 할 것으로 기대된다.

주요 용어 : 토마토, 일사누적속도, 완전 순환식 양액공급시스템.

*본 논문은 농림축산식품부 지원을 받아서 작성됨.(RS-2025-02302990).

¹58213 대한민국 전라남도 나주시 산포면 세남로 1508, 전라남도농업기술원 자원경영과.

E-mail: plantmaniac@korea.kr

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

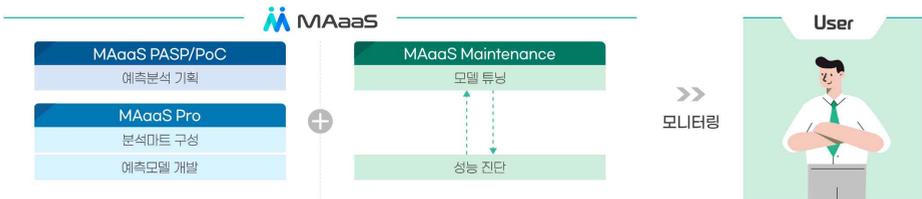
수요 예측 제조 수요예측 분석 시스템 구축 유통 판매량 예측 모델 개발 물류 물동량 예측 시스템 구축 공공 열 수요 예측 모델 개발	텍스트 분석 통신 VoC 분석 금융 텍스트 분석을 활용한 평판 분석 유통 텍스트 분석을 통한 마케팅 증권 챗봇 메시지 분석
상품 추천 유통 AI 상품 추천 프로그램 구축 유통 온라인 쇼핑을 개인화 상품 추천 모델 개발 유통 개인화 추천 서비스 모델 구축 금융 고객관리 시스템을 통한 상품 추천	품질 분석 및 예지보전 제조 취약 부위 예측 모델 개발 제조 특정 공정이상 예측 분석 제조 공정 최적화 및 설비 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.

회사명 ㈜지디에스컨설팅그룹
설립일 2015년 2월
대표이사 김은석
업종 데이터베이스, 소프트웨어 자문, 개발 및 공급, 전자상거래업
주소 서울 성동구 광나루로 228, 9층 (성수동2가, 렉스모터스)
연락처 Tel: 02-2135-8895(대표)
 Fax: 02-2135-8896



- 2018.07 고려대 '데이터 사이언스 센터' 업무 제휴 (교육 및 인턴십)
- 2017.07 KB국민카드 '빅데이터 센터' 업무 제휴 (카드 데이터 상품개발)
- 2016.09 KT '빅데이터 지원단' 업무 제휴 (빅데이터 가공, 중계, 서비스 분야)
- 2016.08 데이터 거래 중개 사업 개시 (미래창조과학부 ICT 기금사업) 부설연구소 '한국공공정보개발원' 설립 (산기협 2016113799호)
- 2016.07 빅데이터 분석 사업 개시 (지속가능한 빅데이터 사업, 경기도)
- 2015.04 국가 DB 구축 사업 개시 (유동인구 실사 DB 구축, NIA)
- 2015.02 ㈜지디에스컨설팅그룹 설립 (㈜지디에스케이 內 컨설팅 사업본부의 법인 분할)
- 2004.10 벤처기업 등록 (041134221-1-01033호)
- 2003.02 ㈜지디에스케이 설립 (대표이사 김은석 차)

Business : 사업 영역

데이터 활용 및 유통의 일관된 서비스 제공!!

사업영역	빅데이터 분석 컨설팅	데이터 가공	빅데이터 플랫폼 구축	데이터 판매/유통
사업영역				
제품 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"> · 국방기술제품지원, 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, 사학스 연구자를 위한 데이터 (16) · 코웨이, 고객정보관리를 위한 데이터 (16)외다수