

# 토픽 모델링을 사용한 정보보호 및 보안 관련 국가 R&D 사업 기술 동향 분석

윤선영, 김민영, 김수현\*  
경북대학교 데이터사이언스대학원

## National R&D Technology Trend Analysis using Topic Modeling for Privacy-Preserving and Security

Sun-Young Yoon, Min-Young Kim, Suhyeon Kim\*  
Graduate School of Data Science, Kyungpook National University

**요약** 최근 개인정보 보호 및 보안에 대한 사회적 관심이 비약적으로 증가하면서 이와 관련된 정보 관리 기술의 발전이 중요한 이슈로 대두되고 있다. 기술의 발전과 개인정보를 보호받을 권리를 모두 추구하기 위해서 국가적 차원에서 정보 보호 및 보안 관련 기술에 대한 정책 및 사업적 지원을 확대하고 있다. 이러한 흐름에 발맞춰, 본 논문에서는 텍스트 마이닝과 토픽 모델링 기법을 활용하여 국가연구개발사업에서의 정보보호 및 보안 관련 기술의 동향성을 분석하고, 최신 연구의 방향성에 대해 논의하고자 한다. 본 연구에서는 최근 5개년의 정보보호 및 보안 기술 관련 국가연구개발사업 과제정보 텍스트 데이터에 Latent Dirichlet Allocation 모델을 적용하여 사업 전반의 토픽들을 추출하고, 연도별 토픽 분석을 통해 기술의 시간적 흐름에 따른 변화를 분석하였다. 본 연구의 결과로부터 얻어진 국가연구개발사업에서의 정보 보호 및 보안 기술 동향에 대한 이해를 바탕으로 향후 산업 기술 개발 및 연구 사업의 방향에 대한 새로운 인사이트 및 시사점을 도출할 수 있을 것으로 기대한다.

**Abstract** Recently, there has been a significant increase in social awareness regarding the protection and security of personal information. As a result, the development of information management technologies related to this issue has become an important concern. National policies and business support for information protection and security technologies have expanded at the national level to pursue technological advancement and the right to protect personal information. In line with this trend, this study examined the trends in information protection and security-related technologies in national research and development (R&D) projects using text mining and topic modeling techniques. This study applied Latent Dirichlet Allocation to text data on national R&D projects related to information protection and security technologies over the past five years. This allowed the extraction of topics throughout the projects and an analysis of changes in technology over time through yearly topic analysis. The results provide new insights and implications for future research topics regarding the trends in information protection and security technologies within national R&D projects.

**Keywords** : Topic Modeling, Information Protection, Information Security, Technology Trend Analysis, Text Mining

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00242528) and by the Ministry of Education (No. RS-2023-00245529).

The authors equally contributed: Sun-Young Yoon, Min-Young Kim

\*Corresponding Author : Suhyeon Kim(Kyungpook National University)

email: suhyeonkim@knu.ac.kr

Received March 11, 2024

Revised April 4, 2024

Accepted April 5, 2024

Published April 30, 2024

## 1. 서론

빅데이터, 인공지능과 관련된 정보기술의 빠른 성장은 우리의 사회와 경제를 급격히 변화시키고 있다[1]. 인공지능 기술의 발전이 가속화됨에 따라 다양한 유형의 데이터들이 실시간으로 수집, 저장 및 가공되고 있으며, 국내외 유수의 공공 및 민간 기업들에서 여러 분야에 산재되어 있는 데이터를 함께 분석하여 다방면으로 활용하고자 하는 수요가 폭발적으로 증가해 왔다. 이러한 성장은 사회와 경제에 큰 변화를 가져왔으나, 정보 기술의 발전과 동시에 개인정보 보호, 프라이버시 이슈, 해킹 등의 보안 문제가 크게 나타나고 있다[2]. 빅데이터의 활용은 여러 이해 관계자들에게 새로운 기회를 제공하지만, 대량의 개인정보 유출로 인한 개인의 프라이버시 등을 침해할 여지가 존재한다[3]. 빅데이터의 활용 이면에 도사리는 개인정보 보호 및 사생활 침해 등과 같은 사회적 문제를 해결하는 것은 빅데이터의 활용성을 더욱 높일 수 있는 방향이다[4,5]. 미국과 유럽연합 등은 개인정보 침해 우려에 대한 대응을 신속하게 취하고 있으며, 국내 역시 빅데이터 환경에서의 개인정보 침해 가능성과 대응 방안 등에 대해 데이터 3법 통과 및 마이데이터 활용 등의 본격적인 논의가 진행되고 있다[6].

데이터를 활용하는 과정에서 필요한 개인정보 보호 및 보안 기술들은 국내외에서 다방면으로 연구가 이루어져 왔다[7]. 초기에는 개인정보 비식별화(de-identification)를 위한 가명처리 및 익명처리 기술이 개발 및 적용되어 왔으나[8], 재식별 위험성이 높다는 한계가 제기됨에 따라 전통적인 비식별화를 넘어선 보호 기술이 개발되고 있다. 개인정보 보호의 수준(level)을 정량화하여 데이터의 2차적 활용(예: 데이터 분석, 인공지능 모델링 등)이 가능하도록 보장하는 차분 프라이버시(differential privacy) 방법은 최근 여러 분야에서 주목받고 있다[7]. 또한, 기업 및 기관의 개인정보 수집을 최소화하면서도 직접적인 데이터의 공유없이 활용할 수 있게 하는 분산 학습(distributed learning), 연합 학습(federated learning) 등의 방법에 대한 연구 또한 활발히 이루어지고 있다[9].

정보보호 및 보안에 관한 관심과 연관 연구의 요구에 따라, 지속적으로 관련 산업 및 학계의 기술 동향을 파악하는 것은 매우 시급하고 중요한 과업 중 하나이다. 빅데이터 보안과 관련된 기술의 트렌드를 분석하거나, 산업에서의 데이터 보호 및 보안에 대한 관심을 파악하는 연구 등과 같이 데이터 활용을 넘어 다양한 분야에서 정보 보호 및 보안에 대한 관심 분야를 파악하고자 하는 연구

들이 함께 수행되고 있다[10]. 그러나 정보보호 및 보안에 대해 직간접적인 국가적 전략이나 정책과 연관성이 높은 국가연구개발사업(국가 R&D 사업)의 양상에 대한 연구는 부족한 실정이다. 따라서, 본 논문에서는 정보보호 및 보안 기술의 동향을 파악하고 국가적 차원에서의 적합한 지원 방향을 고려하기 위해 최근 5개년 동안 수행된 정보보호 및 보안 기술 사업 데이터를 분석하고자 한다. 국가 R&D 사업 과제정보 데이터에 토픽 모델링 기법 중 하나인 잠재 디리클레 할당(LDA: Latent Dirichlet Allocation) 모형을 적용하여 핵심 주제어(토픽)들을 추출하고, 연간 변화하는 대표 기술 분야의 흐름을 파악함으로써 해당 분야의 현황을 심도있게 분석하고자 한다.

본 연구의 구성은 다음과 같다. 제2장에서는 데이터 정보보호 및 보안 관련 동향 분석과 토픽 모델링 기반 동향 분석에 관한 선행 연구를 분석한다. 제3장에서는 분석 데이터, 분석 방법 등을 제시하고, 제4장에서는 데이터 분석 기반 결과를 보여준다. 제5장과 6장에서는 연구결과에 대한 토론, 시사점, 한계 및 향후 계획을 설명한다.

## 2. 관련 연구

### 2.1 데이터 정보보호 및 보안 관련 연구 동향 분석

4차 산업 시대의 신기술이 급속하게 발전함에 따라 다양한 분야에서의 정보보호 및 보안의 중요성이 강조되고 있으며, 대량의 데이터가 축적되고 있는 현시점에서 각 분야에서의 정보보호 및 보안 조치의 필요성이 커지고 있다. 이에 따라 다양한 분야에서 정보보호 및 보안 관련 연구가 꾸준히 진행되고 있다[11-13].

여러 정보보호 및 보안 관련 선행 연구 중, 특히 빅데이터 보안과 관련된 동향 분석 연구가 진행되었다. 학술지와 학술대회 발표집에 게재된 총 62편의 국내외 논문을 분석하여 빅데이터 산업 보안 분야의 연구 동향을 파악하였으며, 이를 통해 향후 연구 방향을 모색하여 정보를 제공하였다[11]. 또한, 산업보안과 관련된 동향 분석 연구 또한 수행되었다. 산업보안 관련 연구논문들을 중점적으로 분석함으로써 산업보안 연구의 동향을 파악하고 향후 산업보안 분야에서 연구해야 할 방향성을 제시하였다[12]. 그뿐만 아니라, 토픽 모델링 기반의 국내 산업보안 동향을 분석하고, 미래 트렌드를 예측할 수 있는 정보를 제공하는 연구 또한 수행되었다[13].

정보보호 및 보안 연구 동향 관련 선행연구를 조사한 결과 다양한 분야에서의 정보 보안이 중요함을 파악하였으나, 국내의 전 범위 적인 국가연구개발사업 단계에서의 데이터 보호 및 보안과 관련된 연구 동향의 파악은 미미하다.

## 2.2 토픽 모델링 기반 기술 동향 분석

기술에 대한 연구 동향을 분석하는 것은 여러 방법으로 진행되어 왔다. 초기 기술 동향 분석 연구에서는 순차적 데이터를 활용한 추세분석을 진행하거나 기본적인 서지 분석을 수행했으나, 이러한 선행연구들은 문헌 자료에 존재하는 내용적 측면을 고려하기 어려웠다[14]. 이를 극복하기 위해 대량의 텍스트 데이터로부터 패턴을 추출할 수 있는 텍스트 마이닝 기법 중 하나인 토픽 모델링을 활용하여 다양한 분야에서 기술 동향을 분석하고 있다[15].

국내외의 토픽 모델링 기반 기술 동향 분석 연구들에서는 논문, 특허, 보고서, 뉴스 기사 등의 다양한 서지형 데이터를 활용한다. Web of Science 등의 논문 데이터베이스를 활용하여 블록체인 분야의 논문 초록 텍스트를 대상으로 토픽 모델링 분석을 진행하거나[16], DBpia로부터 수집한 헬스케어 분야의 약 5,000건의 논문 문서에 LDA 토픽 모델링을 활용하여 해당 분야의 전반적인 연구 동향을 분석한 바 있다[17]. 특허 데이터의 경우, 핀테크 기술의 최신 트렌드를 파악하기 위해, 미국 특허청에 등록된 특허를 수집 후 핀테크 하부 기술에 대한 토픽 분석을 진행한 바 있다[18]. 또한, 증강현실 기술 분야에서의 특허들에 대해 토픽 모델링을 적용 후, 주요 키워드를 추출함으로써 기업별 잠재적인 기술 전략을 파악하기도 하였다[19].

언급된 선행 연구 이외에도 다양한 분야에서 여러 종류의 텍스트 데이터를 활용하여 토픽 모델링 기반의 연구 동향을 분석하는 연구들은 활발히 수행되고 있다. 이는 최근의 연구 동향 분석 연구들에서는 대량의 문서 데이터를 전부 읽고 분석하는 인적, 시간적 비용을 줄여나가는 것이 중요하며, 문서의 핵심 키워드와 시간에 따른 변화를 파악하는 것에 초점을 맞추고 있음을 시사한다. 본 연구에서는 선행 연구들에서 사용하는 토픽 분석 기법의 강점을 살리면서도, 특정 분야의 연구 및 산업의 동향을 한눈에 파악할 수 있는 데이터를 활용하여 정보보호 및 보안 연구의 동향을 분석하고자 한다.

## 3. 연구 방법

### 3.1 분석 데이터

본 연구에서는 국가과학기술지식서비스(NTIS: National Science & Technology Information Service)에서 제공하는 국가연구개발사업(i.e., 국가 R&D)의 과제정보, 특허, 논문 등을 포함한 방대한 양의 국가 R&D 정보를 활용하였다. R&D 데이터 신청을 통해 최근 5개년 국가 R&D 사업과제정보 약 33만건의 데이터를 수집하였으며, 이 중 정보보호 및 보안과 관련된 데이터 샘플을 필터링하여 활용하였다. 본 연구에서 사용한 데이터 변수 및 대표 샘플 예시는 Table 1에 나타나 있다.

Table 1. An example of NTIS data

Variable	Example
Research Title	Developing technology to automatically analyze IoT software for security vulnerabilities
Research Objective	Developing a software platform for automatic analysis of IoT system software and network vulnerabilities
Research Contents	Integrating codeclone vulnerability validation platform for known vulnerabilities with codeclone vulnerabilities detection technology ....
Expectations	Checking for security vulnerabilities to create a more secure IoT environment ....
Keywords	IoT, black box, software security, vulnerabilities, white box
Government Research Grants (KRW)	₩ 1,000,000,000

Table 1과 같이 과제명, 연구목표, 연구내용, 기대효과, 한글 키워드 등의 변수를 추출하여 분석에 사용하였으며, 이 중에서 '정보보호 및 보안'와 관련된 데이터 분석을 위해 "정보 보호(Information protection)", "개인 정보 보호(Privacy-preserving)", "보안(Security)", "프라이버시(Privacy)" 등과 같은 키워드를 기반으로 데이터 샘플을 추출하여 데이터를 분석하였다.

### 3.2 분석 방법

본 연구에서는 토픽 모델링 기법을 활용하여 기술 동향 분석을 진행하고자 한다. 제안하는 방법의 전반적인 절차는 Fig. 1에 나타나 있다.

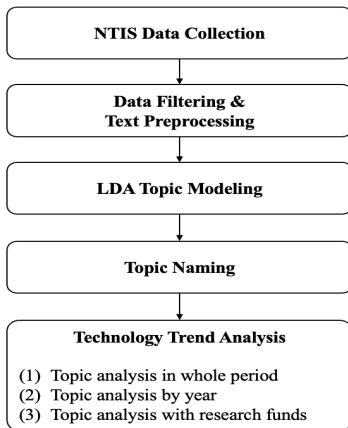


Fig. 1. Topic modeling-based trend analysis process

먼저, 추출된 '정보보호 및 보안'과 관련된 약 9만 개의 국가 R&D 사업정보 데이터에 대한 전처리를 진행하였다. 데이터의 효과적인 분석을 위해 초기 텍스트 데이터 전처리 과정을 실시하였다. '연구', '기반', '보안', '사업', '개인정보관리', '기술' 등과 같이 빈번하게 등장하거나 의미적으로 중요하지 않은 키워드인 stop words를 제거하였다. 또한, '블록 체인', '머신 러닝'과 같이 띄어쓰기로 구분되어 있는 복합명사를 '블록체인', '머신러닝'으로 매핑 처리하여 일관성 있는 단어 형태로 통합하였다. 토큰화 작업 시에는 1글자 불용어인 '나', '너', '저', '그' 등을 제거하였다. 이와 같은 전처리 과정을 통해 데이터를 정제하여 분석의 정확도를 높이고 의미 있는 인사이트를 도출하기 위한 분석 기반을 마련하였다.

전처리 된 텍스트 데이터에 토픽 모델링 기법을 적용하여 정보보호 및 보안 분야 국가연구개발사업의 주요 토픽들을 추출하였다. 토픽 모델링 기법은 비정형 텍스트 데이터에서 중심이 되는 주제어(i.e., topic)를 찾아내어 새로운 정보를 추출하고, 문서 내의 의미 구조를 파악하는 방법으로, 텍스트 데이터 기반의 트렌드를 파악하기 위한 효과적인 수단 중 하나이다[20]. 본 연구에서는 가장 대표적인 토픽 모델링 기법 중 하나인 LDA를 활용하여 토픽 추출을 진행하였다. LDA 알고리즘은 확률 기반의 토픽 모델링 기법으로, 잠재 토픽의 확률적 분포를 학습하는 방법이다[21]. LDA의 목적은 전체 문서의 토픽, 각 문서별 토픽의 비율 및 토픽별 단어들의 확률적 분포를 추정하고자 하는 것이다[22]. 이를 위해 LDA 알고리즘에서는 디리클레(Dirichlet) 분포를 이용하여 문서 집합 내의 단어들에 대한 분포를 추정하고, 깃스 샘플링(Gibbs sampling)을 이용하여 파라미터 및 확률 추정

을 용이하게 한다. 다양한 분야에서 방대한 양의 텍스트 데이터를 분석할 때, 문서 내 토픽과 단어의 확률 패턴을 파악할 수 있어 데이터로부터 인사이트를 도출하는 데 효과적이다.

본 연구에서는 이러한 LDA 토픽 모델링을 활용하여 추출된 데이터로부터 주요 토픽들을 도출하였다. 그 후, 토픽 내 단어의 확률 분포를 기반으로 각 토픽을 대표할 수 있는 상위 확률값의 단어들을 지정하고, 결과 분석에 용이하도록 토픽 내 단어들을 포괄할 수 있는 명칭을 정하는 토픽 네이밍을 진행하였다. 이러한 토픽 결과를 바탕으로 전체 년도에 걸친 연구 동향을 파악하였고, 연도별로 주요 토픽들을 파악하여 연도별로 변화하는 연구 동향을 파악하였다. 추가적으로 각 토픽에 할당된 연구비를 함께 분석하여 연구의 규모와 중요도를 파악하고자 하였다.

## 4. 분석 결과

### 4.1 토픽 개수 선정

일반적으로 LDA와 같은 방법의 토픽 모델링 기법을 활용할 시, 도출할 토픽의 개수를 사전에 정해야 한다. 본 연구에서는 기술 동향 분석을 진행하기 전, 토픽 개수의 선정을 위해 토픽 내의 단어 간의 응집도를 확인하는 신뢰성 있는 지표 중 하나인 topic coherence를 활용하였다[16]. Fig. 2는 토픽 수를 1부터 20까지 변화시켜가면서 LDA 토픽 모델링으로부터 추출된 토픽들의 topic coherence의 값을 계산한 결과를 나타낸 것이다. Fig. 2를 살펴보면, 토픽 수가 10일 때, topic coherence 값이 0.7로 가장 높게 나타난 것을 확인할 수 있다. 이를 바탕으로 본 연구에서는 토픽 수를 10으로 지정 후 최종적인 토픽들을 추출하였으며, 이후 각 토픽을 중심으로 정보보안 및 보안과 관련된 연구주제 동향을 분석하였다.

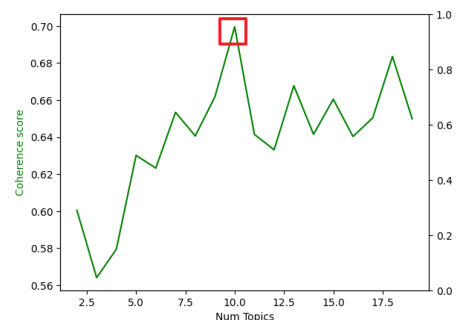


Fig. 2. Coherence analysis for topic count selection

## 4.2 토픽 모델링 기반 연구과제 동향 분석

### 4.2.1 전 기간 토픽 분석

토픽 모델링을 활용하여 정보보안 및 보호 분야의 분석을 수행한 후 토픽별 주요 키워드를 기반으로 전체 기간에 걸친 토픽의 이름을 정하였다.

Table 2. Topics by project name case

Rank	Topic Name	Keywords within Topics
1	Autonomous Driving and Robotics	sensor, image, robot, module, driving, algorithm, autonomous, vehicle, signal, real-time, monitoring, camera, process, part, drone, automotive, safety, prototype, facility, test
2	Genetics and Biology	cell, gene, regulatory, protein, genomic, bio, therapeutic, mechanism, disease, immune, molecular, marker, tissue, drug, animal, neural, genetic, metabolic, target, virus
3	Digital Technologies	smart, platform, network, communication, energy, test, algorithm, power, cloud, module, distributed, digital, server, device, wireless, blockchain, software, realtime, computing
4	Video and Image Processing	artificial intelligence, online, content, algorithm, image, neural network, behavior, image, design, visual, face, voice, education, digital, cognition, human, tools
5	Culture/Society/Nation Policy and Education	education, international, national, policy, science, overseas, innovation, global, competitiveness, Korea, university, leading, safety, navigation, field, domestic, infrastructure, future, industry, industry-academia, network
6	Chemistry/Electronics/Semiconductors	quantum, nano, device, electronic, chemical, energy, particle, electrical, synthesis, metal, ion, semiconductor, optical, molecular, spin, electrode, organic, catalysis
7	Medical and Diagnostics	patient, medical, diagnosis, clinical, treatment, imaging, health, disease, healthcare, biomedical, surgery, disability, hospital, exercise, radiation, dementia, early, artificial intelligence, illness, elderly, care

8	Agriculture	food, crops, microorganisms, varieties, soil, plant, agriculture, harmful, cultivation, process, chemical, farm, farming, toxic, genetic resources, plastic, production, seed, strain, medicine, raw material
9	Safety Monitoring	safety, city, facility, disaster, site, construction, climate, satellite, weather, atmosphere, fine dust, transportation, building, earthquake, structures, monitoring, damage, roads, disaster, architecture, architecture
10	Ecosystem	oceans, creatures, ecology, evolution, communities, conservation, plants, forests, fisheries, birds, cages, synapses, basal, carbon, pain, water quality, element, drought, river, science, stream, science

2017년도에서 2021년도까지 최근 5개년간의 전체 토픽 결과를 살펴본 결과, Table 2와 같이 '자율주행 및 로봇', '유전학 및 생물학', '디지털 기술' 등이 높은 순위로 나타났다. 이를 통해 최근 5개년 동안 '자율주행', '유전학 및 생물학', '디지털 기술' 연구 등에서 정보보호 및 보안 관련 연구가 활발히 진행되어 왔다는 것을 파악할 수 있다.

Table 3. Topics by research project name case

Research Project Name	Topic Name
Nuclear nonproliferation, nuclear security policy studies	Culture/Society/Nation Policy and Education
Developing an iot device for real-time room occupancy analysis using OpenCV algorithms	Chemistry/Electronics/Semiconductors
Developing a smart healthcare service design for managing medication guidance and medication history for severe or special care patients	Medical and Diagnostics
Developing next-generation microbial products using soil microbiome transplantation technology	Agriculture
Established a traffic video big data platform and developed sudden death detection technology	Safety Monitoring

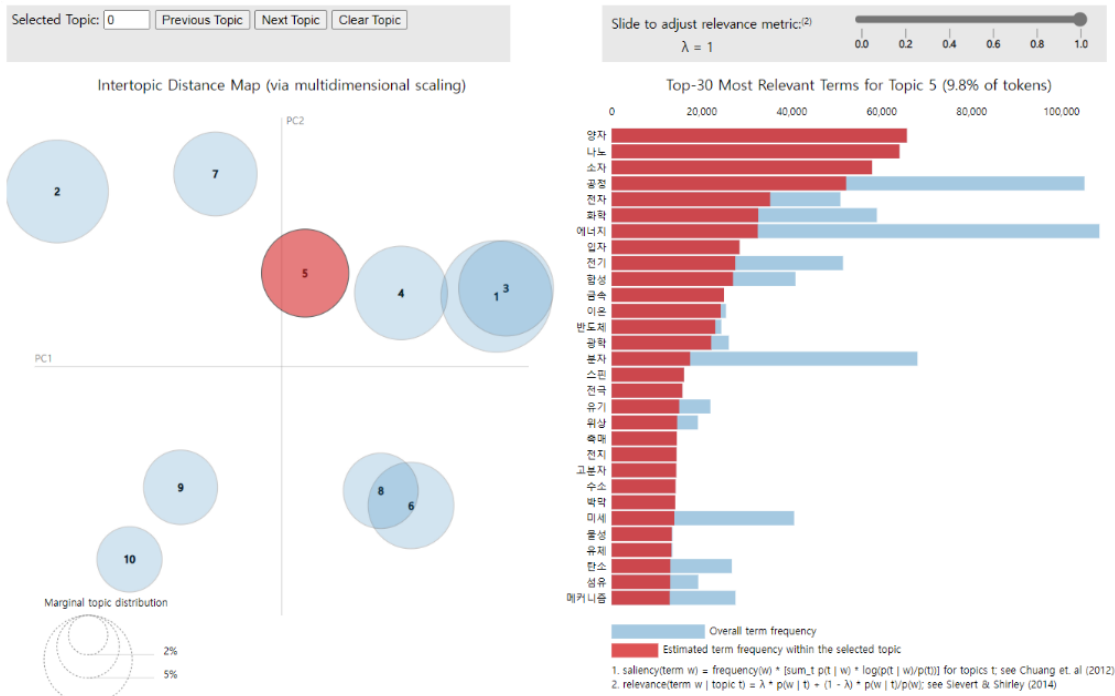


Fig. 3. Intertopic Distance Map

Table 3은 case by case로 LDA 분석을 통해 토픽이 제대로 할당되었는지 확인해 보았다. 이를 통해 과제명이 토픽과 연관되어 있음을 확인할 수 있었다. ‘핵비확산, 핵안보 정책 연구’ 과제명은 문화/사회/국가 정책 및 교육 토픽으로 분류하고 있었다. ‘OpenCV 알고리즘을 활용한 실시간 실내 재실률 분석용 IoT Device 개발’ 과제명은 화학/전자/반도체 토픽으로 분류하고 있었고, ‘중증 또는 특별 케어 환자용 복약지도 및 복용이력 관리를 위한 스마트 헬스케어 서비스 디자인 개발’과 같은 과제명은 의료 및 진단과 같은 토픽으로 분류하고 있다는 것을 알 수 있었다. 이러한 결과를 통해 LDA 분석이 효과적으로 토픽을 분류할 수 있음을 확인할 수 있었다.

Fig. 3은 토픽 모델링의 결과를 통한 여러 토픽 간의 상대적 거리를 시각화한 것이다. 각 토픽은 전체적으로 균등한 분포를 나타내고 있는 것을 확인할 수 있다. 또한 토픽 간의 거리도 존재하는 것을 보아 각 토픽은 상대적으로 독립적인 주제를 나타내는 것을 알 수 있다. 이는 토픽 모델링 과정에서 각각의 주제가 잘 분리되어 있음을 의미하며, 정보보안 및 보호 분야에서 다양한 주제의 연구가 수행되고 있음을 알 수 있었다.

#### 4.2.2 연도별 토픽 분석

본 연구에서는 정보보호 및 보안 관련 연구사업이 연도별로 어떻게 변화하였는지를 살펴보기 위해, 2017년도부터 2021년도까지 연도별 토픽 분석을 진행하였다. Table 4는 연도별로 추출된 주요 토픽을 나타낸 것으로, 정보보호 및 보안 연구에서 핵심 분야의 변화를 살펴볼 수 있다.

Table 4의 결과를 살펴보면 2017년도에는 ‘통신 및 네트워크’, ‘문화/사회/국가정책 및 교육’, ‘디지털 기술’이 토픽이 상위 3개의 토픽으로 나타났으나, 2018년도에는 ‘디지털 기술’과 함께 ‘의학 연구’, ‘영상 처리 및 통신 기술’ 토픽이 상위 3개의 토픽으로 나타났다. 특히, 2018년 이후부터 ‘의학 연구’ 토픽이 상위권에 오르기 시작한 것으로 보아, 정보보호 및 보안 연구가 의학 연구 분야에서 활발히 진행되고 있음을 알 수 있다.

2019년도 ‘의학 연구’, ‘문화/사회/국가정책 및 교육/과학기술’, ‘모바일 기술’ 토픽이 상위 3개의 토픽으로 파악되었다. 특히, ‘의학 연구’ 토픽이 가장 높은 순위로 나타나며, 2018년에 상위권에 포함되지 않았던 ‘문화/사회/국가정책 및 교육’ 토픽이 2019년에 다시 상위권으로 올라오는 추세를 확인할 수 있었다. 2020년도에는 ‘의학

Table 4. Topics for each of the years from 2017 to 2021

Rank	2017	2018	2019	2020	2021
1	Communications and Networks	Digital Technologies	Medical Research	Medical Research	Digital Technologies
2	Culture/Society/ Nation Policy and Education	Medical Research	Culture/Society/ Nation Policy, Education, and Food	Communications and Networks	Medical Research
3	Digital Technologies	Image Processing and Communication Technology	Mobile Technology	Digital Technologies	Culture/Society/ Nation Policy, Education, and Food
4	Genetics and biology	Medical and Diagnostic	Chemistry/Electronics/S emiconductors	Culture/Society/ Nation Policy and education and Technology	Medical and Diagnostic
5	Image Processing and communication technology	Safety Monitoring	Image Processing and Image Technology	Image Processing and Communication Technology	Chemistry/Electronics/S emiconductors
6	Medical Research	Chemistry/Electronics/S emiconductors	Automatic Driving and Robots	Chemistry/Electronics/S emiconductors	IT Technology
7	Safety Monitoring	Agriculture	Digital Technologies	Medical and Diagnostic	Safety Monitoring
8	Medical and Diagnostic	Communications and Networks	Energy and Power	Automatic Driving	Agriculture
9	Chemistry/Electronics/S emiconductors	Culture/Society/ Nation Policy and education	Safety Monitoring	Safety Monitoring	Cognitive Science and Physics
10	Renewable Energy and Urban Infrastructure	Marine Biology	Medical and Diagnostic	Health and Welfare	Smart Maring Technology

연구’, ‘통신 및 네트워크’, ‘디지털 기술’ 토픽이 상위 3 개로 등장한 것을 알 수 있었다. ‘의학 연구’ 분야에서 정보보호 및 보안 연구의 지속적인이고 활발한 연구의 진행을 확인할 수 있었으며, 2020년도에 코로나 팬데믹의 영향을 반영하여 ‘의료 및 진단’ 토픽에서 코로나 바이러스와 관련된 키워드가 등장하는 것을 함께 파악할 수 있었다. 이를 통해 사회적 상황 변화에 따른 정보보호 및 보안 연구의 적극적인 대응으로 활발한 연구 사업 지원이 진행되고 있다는 것을 파악할 수 있다.

2021년도에는, ‘디지털 기술’, ‘의학 연구’, ‘문화/사회/국가정책 및 교육’ 토픽이 상위 3개의 토픽으로 나타났다. 이러한 토픽들은 2017년도부터 2021년도까지 일반적으로 상위권에 위치하는 것을 확인할 수 있으며, 4차 산업 혁명 시대의 흐름에 따라 2021년도에 ‘IT 신기술’에 대한 토픽이 부각되는 것을 확인할 수 있다. 이는 정보보호 및 보안 연구가 IT 신기술 분야에서도 활발하게 연구가 이루어지고 있음을 시사하며, ‘인지과학 및 물리학’, ‘스마트 해양 기술’과 같은 새로운 토픽들의 등장으로 연관 지원 사업 분야의 범위가 다양화되고 있음을 확인할 수 있었다.

#### 4.2.3 토픽별 연구비 규모 분석

본 연구에서는 추출된 주요 토픽에 대한 국가적 관심을 연구비 규모 측면에서도 살펴보고자 하였다. 먼저, 최근 5개년 동안 정보보호 및 보안과 관련된 직간접적인 연구 과제들에 대한 전체적인 연구비의 정보가 Table 5에 나타나 있다. Table 5를 살펴보면, 해당 분야의 연구비 규모는 지속적으로 증가하거나 일정 수준을 유지하고 있는 것으로 나타났다. 과제별 연구비는 각 과제의 규모나 수준에 따라 다르게 설정되기 때문에 연구비만으로는 구체적인 연구 관심 및 지원의 방향성을 완전히 파악하기는 어려우나, 연구비의 약 상승 추세와 지속적인 연구비 지원이 유지되고 있는 것을 보아 국가적인 관심이 큼을 파악할 수 있다. 이를 통해서 정보보호 및 보안 분야에서 국가 정책적으로 연구개발사업에 지속적인 투자를 하고 있음을 확인할 수 있다. 이러한 연구비 투자는 다양한 분야에서의 정보보안의 중요성을 인식하고, 지속적인 기술 발전과 안전성 강화에 기여하고 있는 것으로 사료된다.

Table 5. Total National R&D Funding for Each Year

Year	Research Funds(Total,(Unit:KRW 1,000))
2017	304,629,431.8
2018	318,565,674.9
2019	341,350,478.8
2020	339,686,273.4
2021	342,593,364.7

Fig. 4는 5개년 간 매년 나타났던 토픽들의 연도별 연구비 평균을 그래프로 표현한 것이다. ‘안전 모니터링’ 토픽은 2017, 2021년도에 가장 높은 연구비 분포를 나타내고 있다. 해당 연도에는 안전 모니터링 분야에서 정보보안 및 보호 연구가 중요했다는 것을 시사한다. ‘문화 및 사회, 국가정책’ 토픽과 ‘화학/전자/반도체’, ‘디지털 기술’ 토픽은 5년간 비슷한 연구비 분포를 보이고 있다. 이를 통해 이러한 분야에서는 꾸준히 정보보안 및 보호 관련 연구의 중요성이 대두되고 있음을 시사한다. ‘의료 및 진단’, ‘의학 연구’ 토픽에서는 2017년 이후 연구비가 크게 증가한 후, 지속적으로 높은 수준을 유지하고 있다. 2017년 이후부터 의료 및 의학 연구 분야에서 정보보안의 중요성이 인식되고 지속적인 투자가 이루어지고 있음을 나타낸다.

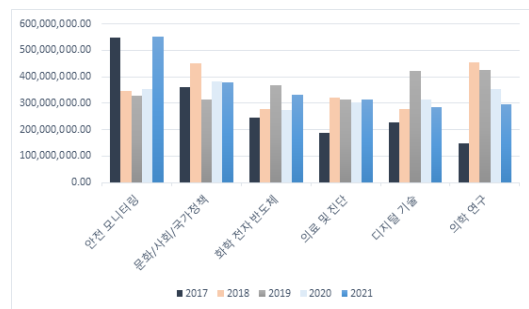


Fig. 4. Distribution of research funds by topic and year

## 5. 토론

본 연구에서 지난 5개년의 연구 기간 동안 대표적으로 주목받은 토픽 중, ‘디지털 기술’, ‘의료 및 진단’, ‘문화/사회/국가정책 및 교육’, ‘안전 모니터링’ 등에 대하여 정보보호 및 보안 분야의 국가연구개발사업 지원, 학계 및 산업계에서의 연구 패러다임 등을 함께 고려하여 분석하

였으며, 각 토픽에 대한 토론 및 시사점은 다음과 같다.

첫째, ‘디지털 기술’ 토픽과 관련하여, 현대 사회에서는 다양한 분야의 디지털 기술 적용이 확대됨에 따라 이와 관련된 정보 보안의 중요성이 더욱 대두되고 있다. 특히, 금융, 게임, 마케팅 등 다양한 분야에서 클라우드, 알고리즘, 인공지능과 같은 신기술을 활용한 개인정보 보호 연구가 진행되고 있음을 확인할 수 있었다[10,11]. 다양한 분야에서의 신기술 적용이 확대되는 만큼 해킹, 사이버 공격, 데이터 유출 등의 위협에 대응하기 위한 연구의 필요성이 증가하고 있어 이에 대한 지속적이고 활발한 연구가 더욱 중요해질 것으로 보인다.

둘째, ‘의료 및 진단’ 토픽과 관련하여 분석을 진행하였다. Fig. 4를 살펴보면, 2018년 이후 해당 토픽 관련 연구비는 크게 증가하였으며, 이후에도 매년 높은 수준을 유지하고 있는 것을 보아 정보보호 및 보안 분야에서 지속적으로 중요한 연구 영역으로 인식되고 있음을 나타낸다. 환자의 의료 기록, 유전자 정보, 수술 정보 등의 의료 데이터는 민감한 개인정보를 대량으로 포함하고 있어, 해당 분야에서의 연구를 수행할 시 안전하고 엄격한 데이터 관리 및 정보 보호 시스템이 필수적이다[23-25]. 특히, 2020년부터 코로나 바이러스와 관련된 연구가 주목받기 시작함에 따라, 해당 연도의 연관 단어들에서 코로나 바이러스에 대한 정보를 찾아볼 수 있으며, 이는 전염병 관련 연구에서도 정보보호 및 보안 분야 연구와 밀접한 관계가 있음을 시사한다.

셋째, ‘문화/사회/국가정책 및 교육’ 토픽과 관련하여, 교육 및 국가 정책 관점에서 보안은 매우 중요한 요소임을 인지할 수 있다. Fig. 4를 살펴보면, 매년 이 주제와 관련된 연구 및 개발에 상당한 국가연구개발사업 연구비가 할당되고 있는 것을 알 수 있으며, 이는 정부와 사회적 차원에서 문화산업 및 정책 분야에서의 정보보호 및 보안에 대한 인식이 높아짐을 시사한다[12]. 국가적으로 사이버 공격, 국가 차원의 사이버 전쟁, 사이버 테러리즘, 해킹 등에 대응하기 위한 전략적 연구개발에 집중하고 있으며, 향후 관련 분야의 연구비에 지속적인 높은 투자가 이루어질 것으로 예상된다.

마지막으로 ‘안전 모니터링’ 토픽과 관련하여 정보보호 및 보안 연구가 중요하게 진행되는 것은 다양한 관리 시스템에서 개인정보 노출에 대한 위험성이 증가하고 있는 것과 관련성이 높다. 교통 관리 시스템, 재난 대응 시스템, CCTV와 같은 감시 시스템들은 기본적으로 사용자의 개인정보 노출 위험을 내포하고 있으며, 해당 시스템에서의 데이터를 활용하여 안전 모니터링을 진행할 경우



외부로부터의 사이버 공격으로 인한 개인정보 노출을 방어하는 기술이 필수적이다[26]. 이처럼 개인정보 보호 기술이 안전 모니터링 분야에 필요해짐에 따라, 정보보호 자동화 기술 및 보안 결합 모니터링 기술 등에 대한 연구들이 함께 진행되어야 할 것으로 보인다.

## 6. 결론

본 연구에서는 NTIS에서 제공하는 국가연구개발사업 정보 데이터를 활용하여 최근 5개년의 정보보호 및 보안 분야의 연구 동향을 분석하였다. 토픽 모델링 방법을 활용하여 과제 정보 텍스트 데이터로부터 주요 키워드를 추출하였으며, 분석 과정에서 빈번성과 영향력을 기준으로 5개년 간의 주요 토픽을 선정하였다. 본 연구에서 최근 5개년간 주요 토픽을 분석한 결과, ‘디지털 기술’, ‘의료 및 진단’, ‘문화/사회/국가정책 및 교육’, ‘안전 모니터링’, ‘화학/전자/반도체’, ‘의학 연구’ 토픽에서 보안 관련 연구가 지속적으로 이루어지고 있으며, 이들 토픽에는 높은 연구비가 할당 및 유지되고 있음을 알 수 있었다. 이는 정보보호 및 보안 연구가 주요 토픽들의 분야에서 중요한 역할을 하고 있으며, 국가적으로 지속적인 관심과 투자를 받고 있음을 나타낸다.

본 연구의 기여점, 기대효과 및 향후 계획은 다음과 같다. 최근 데이터의 급증에 따른 개인정보보호 및 보안 이슈로 관련 연구 및 기술 개발이 각광받고 있다. 본 연구에서는 이러한 특정 분야의 국가적 관심과 지원에 대한 정보에 기반하여 연구 동향을 파악하고 이를 실제 사회 현상들과 연결시켜 분석을 진행하였다. 본 연구의 분석 결과는 정보보호 및 보안 관련 기술의 동향을 파악하여 연구를 진행하고자 할 때, 새로운 연관 분야를 발굴하거나 차별화된 방향성을 탐색하는 데에 도움을 줄 수 있다. 또한, 정보보호 및 보안 분야의 기술 개발에 대한 국가적 차원의 정책 결정과 연구 방향성에 근거를 제공하여 특정 기술 영역에 대한 투자 필요성을 식별하고 효율적인 자원 배분을 할 수 있을 것으로 사료된다.

본 연구에서 활용한 국가연구개발 과제 정보 텍스트 데이터는 국가적 사업 기술 동향을 확인하기 위한 중요한 데이터에 해당하지만, 데이터 내 일부 과제들의 정보가 누락되어 반영할 수 없는 문제점이 존재한다. 또한, 전범위적인 국내 연구 동향을 분석하는 데에 정보량이 한정적이라는 한계가 존재한다. 따라서, 본 연구에서 획득한 결과를 바탕으로 향후 국가연구개발 과제 정보 텍

스트 데이터뿐만 아니라 논문, 특허 등의 데이터를 지속적으로 분석해 나갈 것이며, 소셜 미디어 데이터를 활용하여 정보보호 및 보안에 대한 국민의 인식을 분석하는 연구를 수행하고자 한다. 이러한 연구들은 다양한 분야에 걸쳐 정보보호 및 보안에 대한 효과적인 대응 전략을 마련할 수 있는 기틀을 제공할 것이며, 이는 새로운 보안 위협에 신속하게 대응할 수 있는 기술의 발전으로 이어질 수 있다.

## References

- [1] W. Y. Yoo, "An Analysis of Research Trends in IoT Security" *Journal of Convergence Security*, Vol.18, No.1, pp.61-67, 2018.
- [2] M. Gheisari, G. Wang, M. Z. A. Bhuiyan, "A survey on deep learning in big data", *In 2017 IEEE international conference on computational science and engineering (CSE) and IEEE international conference on embedded and ubiquitous computing (EUC)*, IEEE, Guangzhou, China, pp.173-180, July. 2017.  
DOI: <https://doi.org/10.1109/CSE-EUC.2017.215>
- [3] W. S. Choi, J. Y. Lee, J. Shin, "A Study on the Protection and Utilization of Personal Information for the Operation of Artificial Intelligence and Big Data in the Fourth Industrial Revolution", *Journal of Convergence Security*, Vol.19, No.5, pp.63-73, Dec. 2019.  
DOI: <https://doi.org/10.33778/kcsa.2019.19.5.063>
- [4] Y. G. Shin, "The Development and importance analysis of compliance for personal information protection in the AI service process for AI service providers", *Journal of Korea Regional Information Society*, Vol.25, No.2, pp.1-29, June. 2022.  
DOI: <https://doi.org/10.22896/karis.2022.25.2.001>
- [5] W. T. Lee, J. M. Kang, "A study on Model of Personal Information Protection based on Artificial Intelligence Technology or Service", *Journal of Korean Internet Broadcasting and Telecommunications Association*, Vol.16, No.4, pp.1-6, Aug. 2016.  
DOI: <https://doi.org/10.7236/JIIBC.2016.16.4.1>
- [6] Y. J. Shin, "A Study on the Application of Responsibility and Principles for Personal Information Protection in AI Services", *Korean Journal of Crime Information Research*, Vol.7, No.1, pp.45-74, June. 2021.  
DOI: <https://doi.org/10.33563/KSCIA.2021.7.1.3>
- [7] K. Y. Lee, "Trends in Personal Information Protection and Management Technologies for Big Data Utilization", *Journal of the Korea Institute of Communication Sciences (Information and Communication)*, Vol.37, No.1, pp.32-39, 2019.
- [8] H. J. Lee, S. Y. Lee, B. H. Jeon, "Research on the development of automated tools de-identify personal information

- of data for AI learning - Based on video data -", *Journal of Platform Technology*, Vol.11, No.3, pp.56-67, June. 2023.  
DOI: <https://doi.org/10.23023/JPT.2023.11.3.056>
- [9] B. P. Kim, "Privacy-Enhancing Technologies for AI and the Challenges of Legal Frameworks", *Journal of Law and Economics Studies*, Vol.20, No.1, pp.113-151, 2023.
- [10] S. P. Cho, "The Study on Threats of Information Security and Their Solutions in the Fourth Industrial Revolution", *Journal of Korea Security Science Association*, No.51, pp.9-36, June. 2017.  
DOI: <https://doi.org/10.36623/KSSA.2017.51.1>
- [11] S. K. Park, K. T. Hwang, "A Review of Research on Big Data Security", *Informatization policy*, Vol.23 No.1, pp.3-19, Mar. 2016.  
DOI: <https://doi.org/10.22693/NIAIP.2016.23.1.003>
- [12] D. K. Lee, J. S. Park, "Impact of Security Education on Security Performance by Security Officers: Focusing on the Double Mediation Effects of Protection Motivation and Physical Security Activities", *Korean Journal of Industrial Security*, Vol.12 No.3, pp.111-135, Dec. 2022.  
DOI: <https://doi.org/10.33388/kais.2022.12.3.111>
- [13] S. T. Yoo, K. S. Park, Y. S. Lee, S. J. Hwang, K. S. Kim, "Analysis of Domestic Industrial Security Trends using LDA Topic Modeling", *Journal of The Korean Association for Industrial Security*, Vol.10, No.2, pp.79-103, Sep. 2020.  
DOI: <https://doi.org/10.33388/kais.2020.10.2.079>
- [14] S. Kivikunnas, "Overview of process trend analysis methods and applications", *In ERUDIT Workshop on Applications in Pulp and Paper Industry*, pp.395-408, Sep. 1998.
- [15] N. G. Kim, D. H. Lee, H. C. Choi, W. X. S. Wong, "Investigations on Techniques and Applications of Text Analytics", *Journal of Korean Institute of Communications and Information Sciences*, Vol.42, No.2, pp.471-492, Feb. 2017.  
DOI: <https://doi.org/10.7840/kics.2017.42.2.471>
- [16] S. H. Kim, H. C. Park, J. H. Lee, "Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: A study on blockchain technology trend analysis", *Expert Systems with Applications*, Vol.152, Aug. 2020.  
DOI: <https://doi.org/10.1016/j.eswa.2020.113401>
- [17] C. S. Kim, S. J. Choi, K. Y. Kwahk, "Investigation of research trends in information systems domain using topic modeling and time series regression analysis", *Journal of Digital Contents Society*, Vol.18, No.6, pp.1143-1150, Oct. 2017.  
DOI: <https://doi.org/10.9728/dcs.2017.18.6.1143>
- [18] S. H. Seo, H. Y. Lee, "Fintech trend analysis using topic modeling of BM patents", *Proceedings of KIIE 2015 Fall Conference*. pp. 471-480, 2015.
- [19] B. G. Jeong, J. W. Kim, J. H. Yun, "Competitive Intelligence Analysis of Augmented Reality Technology Using Patent Information", *Sustainability*, Vol.9, No.4, Mar. 2017.  
DOI: <https://doi.org/10.3390/su9040497>
- [20] G. S. Lee, I. J. Lee, Y. K. Lee, "A LDA Topic Modelling Analysis of Virtual Reality Research in Tourism", *The Korea Academic Society of Tourism and Leisure*, Vol.33, No.11, Oct. 2021.  
DOI: <https://doi.org/10.31336/JTLR.2021.10.33.10.85>
- [21] D. M. Blei, A. Y. Ng, M. I. Jordan, "Latent dirichlet allocation", *Journal of machine Learning research*, Vol.3, pp.993-1022, Jan. 2003.  
DOI: <https://dl.acm.org/doi/epdf/10.5555/944919.944937>
- [22] J. D. Park, "A Study on Issue Tracking in Multicultural Research Using Topic Modeling", *Journal of the Korean Society for Library and Information Science*, Vol.53, No.3, pp.273-289, Aug. 2019.  
DOI: <https://doi.org/10.4275/KSLIS.2019.53.3.273>
- [23] A. Guerra-Manzanares, L. J. L. Lopez, M. Maniatakos, F. E. Shamout, "Privacy-preserving machine learning for healthcare: open challenges and future perspectives", *In International Workshop on Trustworthy Machine Learning for Healthcare*, Vol.13932, pp.25-40, July. 2023.  
DOI: [https://doi.org/10.1007/978-3-031-39539-0\\_3](https://doi.org/10.1007/978-3-031-39539-0_3)
- [24] T. H. Lee, J. H. Lee, C. H. Jun, "Bilingual autoencoder-based efficient harmonization of multi-source private data for accurate predictive modeling", *Information Sciences*, Vol.568, pp.403-426, May. 2021.  
DOI: <https://doi.org/10.1016/j.ins.2021.03.064>
- [25] T. H. Lee, S. H. Kim, J. H. Lee, C. H. Jun, "HarmoSATE: Harmonized embedding-based self-attentive encoder to improve accuracy of privacy-preserving federated predictive analysis", *Information Sciences*, Vol.662, Mar. 2024.  
DOI: <https://doi.org/10.1016/j.ins.2024.120265>
- [26] S. Balachandar, R. Chinnaiyan, "Secure solutions for smart city command control centre using AIOT", *ArXiv*, Vol.abs/2108.00003, Jul. 2021.  
DOI: <https://doi.org/10.48550/arXiv.2108.00003>

윤 선 영(Sunyoung Yoon)

[준회원]



- 2022년 2월 : 인제대학교 통계학과 (통계학사)
- 2023년 3월 ~ 현재 : 경북대학교 데이터사이언스학과 석사과정

<관심분야>

텍스트 마이닝, 인공지능

김민영(Minyoung Kim)

[준회원]



- 2023년 2월 : 영남대학교 정보통신공학과 (공학사)
- 2023년 3월 ~ 현재 : 경북대학교 데이터사이언스학과 석사과정

<관심분야>

자연어처리, 그래프 분석

---

김수현(Suhyeon Kim)

[정회원]



- 2020년 2월 : 울산과학기술원 융합경영대학원 비즈니스분석 (이학 석사)
- 2023년 2월 : 울산과학기술원 산업공학과 (공학박사)
- 2023년 3월 ~ 2023년 8월 : 서울대학교 공학연구원 박사후연구원
- 2023년 9월 ~ 현재 : 경북대학교 데이터사이언스대학원 조교수

<관심분야>

데이터 마이닝, 인공지능 응용