

소셜 네트워크 분석과 토픽 모델링을 활용한 설명 가능 인공지능 연구 동향 분석

문건두* · 김경재**

XAI Research Trends Using Social Network Analysis and Topic Modeling

Gun-doo Moon* · Kyoung-jae Kim**

Abstract

Artificial intelligence has become familiar with modern society, not the distant future. As artificial intelligence and machine learning developed more highly and became more complicated, it became difficult for people to grasp its structure and the basis for decision-making. It is because machine learning only shows results, not the whole processes. As artificial intelligence developed and became more common, people wanted the explanation which could provide them the trust on artificial intelligence. This study recognized the necessity and importance of explainable artificial intelligence, XAI, and examined the trends of XAI research by analyzing social networks and analyzing topics with IEEE published from 2004, when the concept of artificial intelligence was defined, to 2022. Through social network analysis, the overall pattern of nodes can be found in a large number of documents and the connection between keywords shows the meaning of the relationship structure, and topic modeling can identify more objective topics by extracting keywords from unstructured data and setting topics. Both analysis methods are suitable for trend analysis. As a result of the analysis, it was found that XAI's application is gradually expanding in various fields as well as machine learning and deep learning.

Keywords : Artificial Intelligence, Explainable AI, XAI, Social Network Analysis, Topic Modeling

Received : 2023. 02. 02. Revised : 2023. 02. 23. Final Acceptance : 2023. 02. 24.

* First Author, Ph.D. candidate, Dongguk University-Seoul e-mail : sidhid@dongguk.edu

** Corresponding Author, Professor, Department of MIS, Dongguk University-Seoul, Pildong-ro 1Gil30, Seoul, 04620, Korea,
Tel : +82-2-2260-3324, e-mail : kjkim@dongguk.edu

1. 서 론

인공지능은 현대인의 삶에 깊숙이 관여하며, 활발한 연구가 진행되고 있는 분야이다. 인공지능에 대한 높아진 관심도와 함께 기계학습(machine learning)도 세간의 주목을 받아왔다. 기계학습은 학습 방식을 입력하면 기계가 스스로 논리를 만들어가도록 제작하는 과정이다. 기계학습 알고리즘 중 대다수는 계산력 부족으로 구현하기에 어려움이 존재했다. 그러나 다층 퍼셉트론 머신러닝 기법의 하나이며, 복잡한 연산도 무리 없이 수행할 수 있고, 예측정확도가 가장 높은 기법인 딥 러닝(Deep Learning)의 등장은 사회적 트렌드로 자리 잡게 되었다. 2012년 ImageNet Challenge에서 최고의 이미지 분류 성능을 보인 이후 이미지 분류 분야에서도 적용되고 있으며, 자연어 처리 및 멀티 모달 데이터 학습 등 다양한 분야에서 활용된다. 또한 피쳐 수와 데이터의 양이 많은 문제를 해결하기에도 적합하다. 딥 러닝은 인공지능이 다양한 분야에서 더욱 뛰어난 성능으로 적용되도록 발전시켰다. 그러나 딥 러닝이 모든 문제를 해결할 수 있는 학습 모델은 아니다. 뛰어난 퍼포먼스를 보여 줘왔음에도, 정보처리 과정이 인간이 이해하기에 너무 복잡하며 입력값이 결과에 끼친 영향력을 알 수 없는 블랙박스 문제가 존재한다(Hwang, 2021). 이는 결과에 대한 명확한 설명을 제공할 수 없는 궁극적인 문제점이 발생하게 되고, 사용자의 결과에 대한 신뢰도를 떨어뜨린다는 단점이 존재하게 된다. 이 문제에 직면하여 통계적 예측에서 해석력을 확보한 기술인 설명 가능한 인공지능(Explainable Artificial Intelligence, XAI)은 설계자와 의사 결정자를 합리적으로 설득한다(Chun et al., 2021).

군사, 의료, 금융 등 신뢰성을 바탕으로 하는 분야에서 단순히 인공지능의 의사 결정을 신뢰하고 따르는 것은 위험도가 높다(Han and Choi, 2017). 이에 미국 방위 고등연구계획국(Defense Advanced Research Project Agency, DARPA)에서 주도하여 XAI 연구를 진행하였다. XAI는 AI 모델을 분석하고 그에 따른 결과를 사람들이 이해할 수 있게 설명하는 기술이다. 최근에는 사회적으로도 인공지능이 내린 결정에 대한 이유제시를 제도화하는 추세이다. 인간의 삶의 영역에 끼치는 인공지능의 영향력이 점차 확대됨에 따라 인공지능의 작동원리 및 의사결정 이유에 대한 중요성이 강조되

고 있다. 유럽연합에서는 2018년부터 General Data Protection Regulation(GDPA)에 의거하여 인공지능에 의해 결정된 사안에 대하여 회사 측에서 설명을 제공하도록 규제하였으며, 미국 또한 주요 금융 관련 결정 사안에 대하여 법적으로 근거를 제시하여야 한다(Han and Choi, 2017). 이처럼 전 세계적으로 인공지능의 발전뿐만이 아닌 의사 결정의 타당성과 신뢰성에 대한 중요도가 높아지며 XAI 연구의 필요성과 다분야에서의 활용도가 증가하고 있다.

본 연구에서는 현대의 XAI 연구의 필요성 및 중요성을 인식하고, ScienceDirect, LEEE Xplore, Springer 등 주요 학술지에서 학술지 논문을 대상으로 논문 내 초록, 키워드를 중심으로 XAI 또는 Explainable AI를 키워드로 하여 논문을 추출하고, 소셜네트워크 분석(Social Network Analysis, SNA)과 토픽모델링을 활용한 XAI 동향 분석을 진행하고자 한다. SNA는 노드와 관계성을 나타내는 링크를 이용하여 네트워크를 시각화하여 분석한 방법론이다. 문서의 키워드 단어 간의 관계성을 이용하여 분석하며 대용량의 데이터를 각각 개체화 후 시각화하기에 관련 연구 진행의 핵심 주제 및 빈도수가 높은 주제를 한눈에 파악할 수 있는 분석법으로 기존의 통계적 연구 동향 분석과는 차별화된 기법이다. 토픽모델링은 비정형 데이터 기반으로 핵심 키워드를 추출하는 기법으로 더욱 객관적인 관점으로 연구 동향 분석이 가능하다. 대용량의 문서에서 예상치 못한 잠재적인 주제를 파악 및 다각도의 트렌드 분석이 가능하다는 장점이 있다. 그 결과를 바탕으로 XAI관련 연구의 발전 방향을 파악하여 이해하는데 도모하고자 하며 후속 XAI 관련 연구에 대한 통찰력과 방향성을 기르기 위한 기초자료로 참고할 수 있도록 제안한다.

2. 이론적 배경

2.1 설명 가능 인공지능(Explainable AI, XAI)

1975년 설명 가능한 의사 결정 체계라는 용어로 등장한 XAI는 2004년이 되어서야 Michel Van Lent, William Fisher, Michael Mancuso에 의해 전문 용어로 자리를 잡았다(Shortliffe et al., 2004; Van Lent et al., 2004). XAI는 해석 가능

한 인공지능(Interpretable AI) 또는 투명한 인공지능(Transparent AI)으로도 불리며 블랙박스(Black Box) 모델의 학습 과정 및 의사결정에 미치는 요인을 분석하여 인간이 이해할 수 있는 수준으로 변환하는 기법이다. 딥 러닝 이전의 모델들은 인간이 설계한 로직으로 작동하였기에 이해하기 복잡하지 않았으나 딥 러닝이 발전하면서 고차원의 계산을 수행하기 시작하면서 이 과정 및 결과에 대한 설명의 필요성이 요구되고 있다.

기존의 기계학습 모델은 작동 성능을 평가 함수(Evaluation Function)로 진단하여 영향력을 수치로 측정하는 반면, XAI는 피쳐 중요도(Feature Importance), 대리 분석법(Surrogate Analysis), 의사결정트리(Decision Tree), 필터 시각화(Filter Visualization) 등 다양한 방법으로 분석할 수 있다 [Chu et al., 2022]. 의사 결정 트리는 질문과 답을 반복하는 과정을 통해 집단 분류(classification) 및 예측(prediction)하며, 대리 분석법은 본래 모델의 기능을 흉내 내는 대체 모델을 만들어 프로토타입이 동작하는지 판단하는 분석법이다. 이때 학습 데이터를 사용하여 분석 모델을 구축하면 글로벌 대리 분석(Global Surrogate Analysis), 분석 모델이 학습 데이터 하나를 해석하는 과정을 로컬 대리 분석(Local Surrogate Analysis)하고 한다. 이미지 필터 시각화는 학습된 신경망 모델에 이미지가 입력되면 각 층의 반응을 시각적으로 확인하는 방법이다. XAI는 개발자와 사용자 모두에게 인공지능에 대한 이해도를 높여 인간과 인공지능 사이의 상호작용에 대한 합리성을 부여하며, XAI를 통해 인공지능의 출력 결과의 신뢰도와 투명성을 높이고, 머신 러닝 모델의 복잡성을 해소하여 개발자와 사용자 간의 소통 수단으로써의 역할을 수행한다[Lee et al., 2021; Oh et al., 2021]. 일반적인 전통 머신 러닝 모델에 적용하는 XAI가 있는가하면 딥 러닝에 적용할 수 있는 XAI기법(Deep Explanation)은 복잡한 신경망을 분해 및 이해할 수 있어야 한다. XAI의 블랙박스 문제 설명은 모델 설명(Model Explanation), 출력 결과 설명(Outcome Explanation), 모델 검사(Model Inspection) 세 가지로 나뉜다. 출력 결과 설명 방법에는 LRP(Layer-wise Relevance Propagation), LIME(Local Interpretable Model-agnostic Explanation), SHAP(Shapley Additive

Explanations) 등이 있다[Guidotti et al., 2018; Bach et al., 2015; Ribeiro et al., 2016; Lundberg et al., 2017; Lee et al., 2021]. 이러한 해석 모델은 세 가지 측면에서 평가되는데 해석 성능(Interpretable Models)과 해설 성능(Explanations Interpretability) 그리고 최종 사용자의 만족도(XAI System Outcomes)로 측정된다[Mohseni et al., 2018].

고도로 발전된 현대의 인공지능은 전문가들도 어려워하는 복잡한 계산 및 문제에 대한 해결책을 제시하고 있지만, 의사결정에 대한 설명의 부재는 개발자 또한 작동원리에 대한 확신을 가지기 힘들기에 신뢰도를 바탕으로 한 영역에서는 활용에 한계가 존재한다. 그러나 XAI는 이러한 인공지능의 문제점을 해결함과 동시에 블랙박스를 다각도로 볼 수 있는 도구이다. 앞으로 XAI에 관한 연구는 더욱 활발히 진행될 것이며, 인간들이 인공지능을 합리적으로 수용할 수 있는 시발점이 될 것이며 XAI가 해결해야 할 문제를 또한 점점 확대될 것이다. 이에 본 연구에서는 XAI연구의 중요성을 인지하고 XAI관련 연구의 동향을 파악하였다.

2.2 소셜 네트워크 분석(Social Network Analysis, SNA)

소셜 네트워크 분석은 네트워크 분석과 사회 이론을 접목한 분석법으로 광범위한 활용 범위로 다양한 연구 수행에 활용이 가능한 관계 구조의 의미 분석에 초점을 둔 기법이다[Kim et al., 2014]. 네트워크 분석에서 연결된 구조의 특성은 밀도(density), 중심성(centrality), 집중성(centralization) 등의 지표로 파악된다[Hansen et al., 2009]. 소셜네트워크 분석은 설정된 문서 데이터에서 상위 빈도 키워드들을 추출하여 노드(Node)로 인식한 후 각 키워드 사이의 관계인 엣지(Edge)를 선으로 시각화하여 네트워크 구성의 특징을 분석한다[Jaegal, 2019]. 특히 추출된 키워드들은 대표적으로 연결 중심성(Degree Centrality), 근접 중심성(Closeness Centrality), 매개 중심성(Betweenness Centrality)을 기준 지표로 중심성이 평가되어 네트워크에서 특정 노드가 가지는 영향력을 측정할 수 있으며 중심성이 높은 키워드가 핵심 키워드로 나타난다[Kim, 2018]. 연결 중

심성은 키워드들 간의 연결된 관계의 횟수를 측정하며, 횟수가 많을수록 중심 키워드로 나타난다. 매개 중심성은 다른 키워드들 사이에 위치하는 정도를 측정하여 매개적 역할 및 영향력을 나타내며, 근접 중심성은 노드가 다른 노드들과 위치한 거리를 측정하여 전반적으로 짧은 거리일수록 여러 노드와 연결되어 있어 접근성이 좋고 일반적인 영향력이 크다고 본다(Kim et al., 2015). 소셜네트워크 분석은 네트워크의 노드 사이의 연결 관계를 시각화한 분석이기에 전체적인 네트워크 형태로부터 패턴을 파악하여 연구 동향을 한눈에 파악할 수 있다는 장점이 있어 트렌드 분석 연구에 적합하다(Park and Kwak, 2013). 소셜네트워크 분석은 대용량 데이터처리 기술이 발전함에 따라 다양한 연구에 응용되고 있으며 여러 분야에서 각광받는 분석 방법이다.

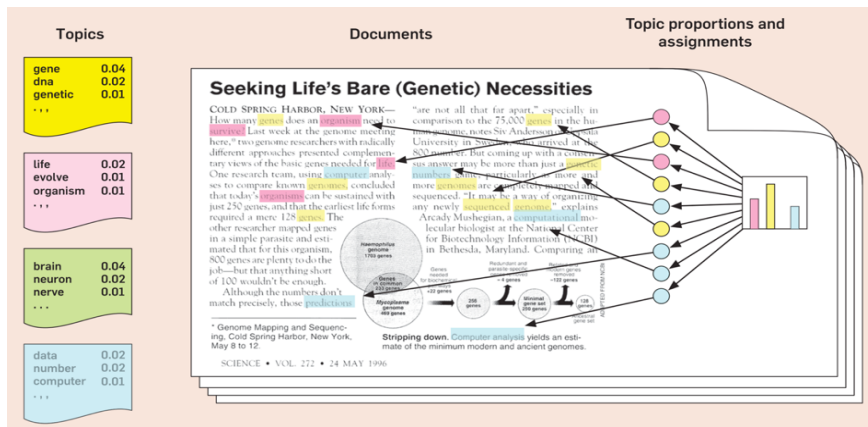
2.3 토픽 모델링(Topic Modeling)

토픽 모델링은 많은 양의 텍스트 문서 데이터로부터 메인 토픽들을 추출하여 주제를 도출하기 위한 분석 방법으로 텍스트 마이닝 방식을 활용해 비정형 데이터를 머신 러닝 알고리즘으로 확률 분포에 기반하여 대량의 텍스트의 주제를 객관적이고 명확하게 파악하는데 유용하다(Han and Kim, 2021). 즉, 기계가 많은 양의 문서 데이터를 분석하여 주제를 추론하여 알려주는 것이다. 이때 적용되는 모델에 따라 데이터를 바라보는 관점이 달라지게 되는데, 토픽 모델링에서

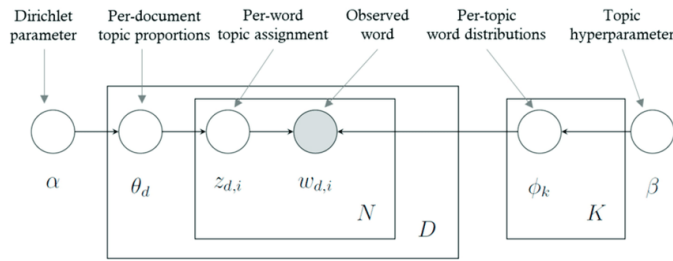
사용되는 대표적인 알고리즘은 Deerwester(1990)이 제안한 LSI(Latent Semantic Indexing), Hofmann(1999)이 제안한 pLSA(Probabilistic Latent Semantic Analysis), Blei(2003)가 고안한 LDA(Latent Dirichlet Allocation)이다(Lee et al., 2021). 이 중 LDA방법이 결과 해석에 용이하며, 과적합 문제를 해결하기에 여러 토픽을 도출하는데 적합하여 학술 동향 분석 연구에서 가장 많이 사용된 알고리즘이다.

LDA는 문서 데이터 내의 키워드들을 디리클레(dirichlet) 분포를 통하여 설정한 개수만큼의 키워드를 생성하는 알고리즘으로 일반적인 주제 분석과 달리 잠재된 주제를 파악하는데 유용한 확률론적 추론 모델이다(Kim, 2020). 생성된 키워드들을 기반으로 유형화한 주제로 전반적인 데이터를 파악할 수 있다. <Figure 1>은 LDA의 원리를 보여준다. 모든 문서는 여러 개의 토픽을 가지고 있으며 각각의 토픽은 단어들의 분포라고 가정한다. 그리고 문서 집합이 동일하다면 개별 문서의 토픽의 비중은 변하지 않고, 각 단어들은 토픽에서 샘플링 되었다고 가정한다. <Figure 2>는 LDA의 과정을 보여주는데, 먼저 매개변수에 의해 특정 문서의 토픽 분포도를 추정하고, 매개변수에 의해 각 토픽들에 대한 단어의 비중을 추정한다. 추정된 토픽 분포도에 의해 개별 문서에서의 특정 단어가 추출된 토픽을 그리고 마지막으로 단어의 비중과 단어가 추출된 토픽으로 단어를 추출할 수 있다.

LDA는 알고리즘이 단순하며 단어들 사이의 관계



<Figure 1> The intuition behind Latent Dirichlet Allocation [Blei, 2012]



(Figure 2) Document generating process using Latent Dirichlet Allocation [Blei, 2003]

에 의미를 형성한다고 가정하고 한 문서에서 동시 출현한 단어들을 중요하게 포착한다. 또한 문맥에서 다른 의미를 가진 단어들을 효과적으로 분류해준다 [Moon, 2020]. 여러 토픽을 도출함으로써 다양한 관점에서의 주제 분석이 가능하여 더욱 객관적인 연구 진행이 가능하며 [Park et al., 2017], 단순히 주제만 보여주는 것이 아니라 키워드들도 보여주기 때문에 더 세부적인 분류가 가능하다. 본 연구에서도 연구 동향 분석에 주로 사용되는 LDA 기법을 사용하였으며 연구 결과의 정확도를 높이기 위하여 불필요한 수식어와 용어를 제거 및 표준 단어로 변환하는 데이터 전처리 과정을 진행하였다.

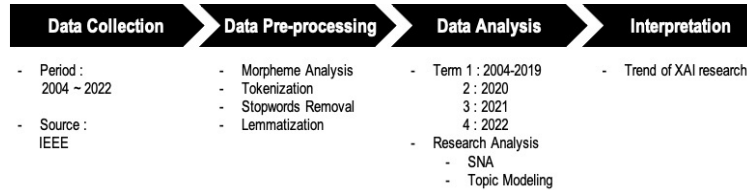
2.4 선행 연구

XAI 연구는 인공지능의 판단에 대한 설명이 필수적인 신뢰도 기반인 재무 및 금융 분야에서 활발히 진행되었다. Kim [2021]은 중소기업 전용 심층신경망 기반의 신용평가 모델을 구축한 후, XAI를 활용하여 인공지능이 판단한 신용 평가 결과에 대한 원인을 파악하여 의미를 해석하는 연구를 진행하였고, Chun et al. [2021]은 다양한 신용 정보 데이터를 활용하여 인공지능 기반의 설명가능한 신용평가 알고리즘을 제안하였다. Biecek [2021]은 신용 평가 예측 모델에 XAI를 이용하여 복잡한 예측 모델을 명확히 이해하고자 하였고, Jung et al. [2021]은 회계 부정 탐지 모형의 개발에 있어 XAI 기법 중 하나인 SHAP을 사용하여 내부회계관리제도 운영보고서 감성분석의 유용성을 검증하였다. Kim and Hong [2022]는 딥러닝 기반 개인신평가모형의 신뢰성 확보를 위한 XAI의 탐색적 적용에 대한 연구를 진행하였다. XAI 연구는 추천을 기반으로 한 시스템 연구에서도 활용되었다.

Hwang [2021]은 지식그래프를 시각화한 설명 인터페이스(KGXI)를 구축하여 알고리즘에 기반한 추천 서비스에 설명 정보가 더해졌을 때 대중들의 선호도에 관한 연구를 진행하였고, Chen [2022]은 기존 추천 시스템에 딥 러닝 기반의 텍스트 요약 방법을 결합한 모델로 영화 데이터베이스에서 추출한 리뷰를 수집하고 요약하여 추천 영화에 대한 설명을 제공하는 연구를 진행하였다. Kim [2022]은 디자인 사이언스를 방법론을 기반으로 XAI를 이용하여 사용자에게 추천 이유를 설명하는 방법론에 관한 연구를 제시하였다. Hwang [2022]은 사용자 리뷰를 입력 변수로 사용한 감성분석에 준지도학습과 XAI를 이용하여 감성라벨에 대한 정보를 얻고자 하였다. 뿐만 아니라 XAI는 악성코드 탐색과 같은 보안 분야에도 활발히 이용되었다. Kim and Hong [2022]는 악성 사이트 탐지를 위한 XAI 기반 기계학습 특징 선에 대한 연구를 진행하였고, Shin et al. [2022]은 악성코드 대응을 위한 XAI 프레임워크 연구를 진행하였으며, Son et al. [2022]은 XAI를 사용한 이상 탐지 결과의 위양성 감소에 대한 연구를 진행하였다. 이처럼 인공지능에 익숙해진 현시대의 사람들은 단지 인공지능의 판단 자체보다 판단에 대한 이유를 요구한다는 것을 알 수 있다. 그리고 그것은 의료, 재무 등 주요 분야뿐만이 아니라 점차 설명성이 부여되는 모든 분야에 적용되고 있으며 블랙박스를 이해하고 시스템을 신뢰하기 위한 XAI의 중요도가 더욱 커지고 있다.

3. 연구 방법

본 연구는 일반적인 텍스트 마이닝 분석 절차로 (Figure 3)과 같이 데이터 수집, 데이터 전처리, 데이터 분석, 결론 순으로 진행하였다.



〈Figure 3〉 Research Process

3.1 데이터 수집

본 연구는 XAI가 등장한 2004년부터 2022년까지 약 19년간 축적된 'XAI'와 'EXP' 키워드를 토대로 해외 학술지 전기전자공학회(IEEE, Institute of Electrical and Electronics Engineers)로부터 수집하였다. 'XAI' 키워드를 기준으로 470편, 'EXP' 키워드를 기준으로 865편, 총 1,342편을 추출하였다. 기간 1은 2004년부터 2019년까지 게재된 124편, 기간 2는 2020년에 게재된 207편, 기간 3은 2021년에 게재된 463편, 기간 4는 2022년에 게재된 541편으로 나누어 기간별 분석을 진행하였다. 이때 분석 기간을 4개의 기간으로 나눈 이유는 다음과 같다. 먼저 XAI의 개념이 태동되었으나 개념적으로만 소개된 시기를 기간 1으로 설정하였으며, 2020년 이후부터는 본격적으로 XAI를 활용한 분석 결과를 다루는 연구들이 나타났기에 이후부터는 1년 단위로 나누어 분석을 수행하였다. 특히 2020년에는 연간 500여 편의 연구가 진행되어 유사한 편수의 논문이 진행되었으므로 1년 단위로 연구 동향을 분석하는 것이 의미가 있을 것으로 판단하였다. 수집된 학술지 분포는 〈Table 1〉과 같다.

〈Table 1〉 Data

Term	Period	Journals
1	2004 ~ 2019	124
2	2020	207
3	2021	463
4	2022	541
Total		1335

3.2 데이터 전처리

대표적인 텍스트 마이닝 기법인 소셜네트워크 분석과 토픽 모델링은 대량의 비정형 텍스트 데이터 속에서 의미

있는 패턴을 발견하여 유용한 정보를 추출한다(Berry et al., 1997). 비정형 텍스트를 자연어로 처리할 때, 토큰화하여 형태소를 분석하고, 불용어 처리, 표제어를 추출하여 진행하였다. 토큰화는 말뭉치에서 일정한 의미가 있는 가장 작은 말의 단위로 나누는 과정이며, 자연어 처리 과정에서는 보통 형태소를 토큰으로 이용한다(Kang and Park, 2019). 불용어 처리는 연구 결과를 왜곡할 수 있는 불필요한 요소와 의미가 없는 단어의 제거 작업이며, 표제어 추출은 말뭉치의 복잡성을 줄이며, 텍스트를 정규화하는 작업이다.

3.3 데이터 분석

3.3.1 SNA 분석

소셜 네트워크 분석은 선정된 문서 데이터 내에서 상위 빈도 키워드를 추출하여 노드로 인식한다. 그리고 인식된 노드들의 연결성을 링크로 나타내는 기법이다. 각 각의 노드에 의미를 부여하기 보다는 전체적인 형태를 파악함으로써 시각화된 소셜 네트워크 분석은 전반적인 문서 데이터의 내용 및 트렌드를 한눈에 파악하기에 용이하다. 본 연구에서는 해외 학술지 IEEE에 게재된 저널을 대상으로 하였으므로 영문 키워드로 분석을 진행하였다. 본 연구의 SNA분석 결과는 〈Table 2〉와 같다.

기간 1에서는 'Task Analysis', 'Training', 'Machine Learning'이 연결 빈도가 높은 상위 키워드로 기계 학습과 태스크 검증과 관련한 XAI 연구가 활발했다고 볼 수 있으며, 'Predictive Models'는 상위 키워드들과 연결되어 있고, 'Artificial Intelligence'는 'Training', 'Task Analysis'와 연결되어 있는 것으로 보아 구체적으로 예측 기반의 머신 러닝이 추세였으며 인공지능의 성능의 검증에 대한 연구가 진행되었다는 것을 알 수 있다. 연결 정도는 5이상으로 나타났다. 기간 2의 학술지에서는 'Task Analysis', 'Machine

<Table 2> Result of Social Network Analysis

TERM	
1	
2	
3	
4	

Learning', 'Predictive Models', 'Computational Modeling'이 상위 키워드이며, 이는 기간1과 같이 기계 학습과 태스크 분석과 같은 기본적인 성능 분석과 함께 'Predictive Models', 'Computational Modeling', 즉 예측 모델링과 계산 모델링에 대한 XAI 연구가 진행되었다는 것을 알 수 있다. 'Analytical Models', 'Artificial Intelligence'와 같은 키워드들이 주요 키워드들과 링크로 연결되어있으며 이는 인공지능 시스템의 수학적 분석이 이루어졌다고 볼 수 있다. 이 기간에서는 'Fuzzy Logic' 키워드도 등장을 하였는데 이는 인공지능 시스템에서의 불확실성과 애매함을 다루는 연구가 진행되었다고 해석할 수 있다. 연결 정도는 5 이상으로 나타났다. 기간 3에서는 'Predictive Model', 'Visualization', 'Computational Modeling'이 상위키워드로, 'Deep Learning', 'Data Models', 'Artificial Intelligence', 'Conferences'가 상위키워드들과 연결되어있으며, 'Decision Making'과 'Machine Learning'의 연결을 보아, 의사결정 예측 모델 머신 러닝에 대한 XAI 연구가 실행되었다는 것을 알 수 있으며, XAI에서 머신 러닝 모델의 과정을 시각화하는 방법을 사용하는 연구가 진행되었다고 해석할 수 있다. 연결 정도는 10 이상으로 나타났다. 기간 4에서는 'Deep Learning', 'Predictive Models', 'Artificial Intelligence', 'Training', 'Neural Networks'가 상위키워드로 나타났으며, 'Deep Learning'과 연결된 'Prediction Algorithms', 'Intrusion Detection', 'Transfer Learning', 'Decision Making', 'Convolution Neural Network'노드들을 보면 다양한 딥러닝의 복잡한 신경망에 XAI를 적용하여 인공지능에 적용하는 연구가 활발했다고 해석할 수 있다. 또한 'Medical service', 'Feature extraction', 'Computer vision'이 주요 키워드들과 연결되어있으며, 이는 의료분야, 객체 인식 분야 뿐만 아니라 지능형 에이전트 구현을 위한 컴퓨터 시각에도 XAI 연구가 접목되었음을 알 수 있다. 연결 정도는 10 이상으로 나타났다.

3.3.2 토픽 모델링

많은 양의 문서 데이터의 내용을 한눈에 파악할 수 없기에 토픽 모델링을 활용하여 주어진 문서에서 주요 키워드를 도출한 후 각 키워드에 맞는 주제를 정하여 전체적인 내용을 파악한다. 본 연구에서는 XAI의 연

구 동향을 파악하기 위하여 주요 학술지 IEEE의 논문을 대상으로 토픽 모델링을 진행하였다. 토픽모델링으로는 Biel et al.(2015)가 고안한 LDA알고리즘을 이용하여 문서 데이터에서 잠재된 주제를 도출해내 고자 하였으며, 연구 주제에서의 최적의 토픽 수를 10개로 설정하였다. 키워드 추출 시 10개로 설정하였으나, 같은 랭킹에 속한 키워드가 있어 최대 14개까지 추출되었다.

〈Table 3〉을 통해 도출된 기간별 토픽은 다음과 같다. 기간 1의 토픽은 'Law and technology', 'Data analysis', 'Machine learning', 'Data analysis', 'E-health', 'Security AI', 'Disease detection', 'Security software', 'AI based text analysis'이고, 'Law and technology'와 'Data analysis'는 정보화 시대가 도래하면서 디지털 데이터와 관련된 법률적 문제와 이를 해결하기 위한 기술적 측면에서 연구가 이루어졌음을 보여준다. 'Machine learning'과 'Data analysis'는 인공지능 분야에서 가장 중요한 주제 중 하나인 머신 러닝과 데이터 분석에 대한 연구를 나타내고, 'E-health'와 'Disease detection'은 의료 분야에서 디지털 기술이 적용된 사례를 다루고 있으며, 'Security AI'와 'Security software'는 보안 분야에서 인공지능 기술의 활용에 대한 연구를 나타낸다. 마지막으로 'AI based text analysis'는 자연어 처리와 텍스트 마이닝에 대한 연구를 보여줍니다. 이는 다양한 분야에서 인공지능 기술의 활용이 이루어지고 있음을 보여주는 것으로 해석할 수 있다. 특히, 법률과 기술에 대한 연구는 인공지능이 사회에 미치는 영향에 대한 관심이 높아지면서 중요성이 부각되었다고 볼 수 있다.

기간 2의 토픽은 'Financial applications', 'Security research', 'Autonomous driving application', 'Blockchain', 'Speech recognition', 'Improved interpretability of deep learning models', 'Vulnerability analysis for security', 'Human-robot interaction', 'Artificial intelligence systems'이 등장하였다. 2020년은 COVID-19 대유행이 전 세계적으로 발생하면서 많은 사람들이 건강과 안전에 대한 우려를 가지게 되었고, 이에 따라 의료분야와 보안 분야를 주축으로 많은 분야에서 인공지능 기술을 활용한 연구와 적용이 활발하게 이루어졌다. 'Financial Applications'과 'Blockchain'과 같은

토픽에서 나타나듯 금융 분야에서 인공지능을 활용하여 금융 거래 및 투자에 대한 예측과 분석을 수행하는 연구가 진행되었으며 금융 데이터의 안전한 저장 및 전송을 위해 블록체인 기술과의 결합을 통한 보안성 강화 연구가 진행되었다. ‘Speech Recognition’과 ‘Improved Interpretability of Deep Learning Models’는 음성 인식과 딥러닝 모델 해석성 개선에 대한 연구를 나타내며, ‘Vulnerability Analysis for Security’는 코로나의 대유행으로 주목받은 보안 분야를 나타낸다. 코로나19로 인해 많은 기업과 조직이 원격근무를 시행하게 되면서, 사이버 공격자들이 이를 이용해 보안 취약점을 찾아내고 공격을 시도하는 경우가 증가했다. 이뿐만 아니라 코로나19로 인해 정보의 중요성이 더욱 부각되면서, 개인정보 유출, 악성 코드 등의 사이버 공격 또한 증가하게 되었다. 이러한 상황에서 보안 분야는 더욱 중요한 역할을 수행하게 되었고, 이에 따라 보안 분야에서의 인공지능 기술 연구도 더욱 활발해졌다. ‘Security research’는 보안 분야에서 인공지능 기술을 활용하여 보안 위협을 예측하고 탐지하는 연구들이 활발하게 이루어졌다고 해석할 수 있는 토픽이다. ‘Autonomous Driving Application’ 코로나바이러스로 인한 팬데믹으로 사람들이 대중교통을 이용하는 것을 불안하게 느끼면서 차량의 수요가 급증했고, 이에 자율주행 차량의 개발과 연구가 활발해졌으며 미래에 대중교통이 제한적으로 제공될 때 차량 이동 수요를 충족시킬 수 있는 대안적인 수단으로 주목받았다고 볼 수 있다. 마지막으로 ‘Human-Robot Interaction’과 ‘Artificial Intelligence Systems’는 인간과 로봇 간 상호작용 및 인공지능 시스템 개발에 대한 연구를 보여주는데, 코로나바이러스를 겪으며 바이러스 발생 시, 대응에 필요한 예방 및 대처 전략을 개발하는 연구가 많은 관심을 받았다고 해석할 수 있다. 또한 인공지능 기술을 사용하여 바이러스의 전파 패턴을 분석 및 예측 모델을 개발과 이를 기반으로 국가 및 지역별 적절한 대처 전략 수립에 대한 연구가 진행되며, 기업 및 산업 현장에서 인공지능 기술의 적용 범위가 더욱 확대되었다는 것을 보여준다.

기간 3의 토픽은 ‘Data analysis’, ‘AI security’, ‘Security’, ‘Image processing’, ‘Robotics’, ‘Medical artificial intelligence’, ‘Pattern recognition’, ‘Medical Technology’, ‘Anomaly detection’, ‘Business application’이 주요 토픽이었으며, 2021

년은 코로나19의 여파가 여전히 이어지면서 디지털화와 인공지능 기술의 발전이 더욱 가속화된 시기였다. 먼저, ‘data analysis’와 관련된 XAI 연구는 데이터의 양이 증가함에 따라 데이터 분석 기술의 중요성이 커지면서 다양한 연구가 이루어졌다. 특히, 딥러닝 모델을 이용한 데이터 분석 기술의 발전과 함께 XAI 기술을 활용한 모델 해석성(interpretability) 연구가 주목 받았고, ‘AI security’와 ‘Security’ 토픽에서 악성코드 탐지, 침입 탐지, 보안 위협 대응 등의 보안 분야에서 XAI 기술이 활용되는 연구가 많이 이루어졌다고 해석할 수 있다. 특히, XAI 기술을 활용한 보안 취약점 분석 및 대응 기술 개발에 대한 관심이 높아졌다. ‘Image processing’ 분야에서는 인공지능을 이용한 의료 영상 분석, 자율주행차량 영상 인식 등의 이미지를 사용한 딥 러닝 분야에서 XAI 기술이 접목되는 연구가 진행되었다고 볼 수 있다. 또한, ‘Robotics’ 분야에서는 인공지능을 이용한 로봇 제어 기술 및 로봇과 인간의 상호작용 분야에서 XAI 기술의 활용이 더욱 확대되었다고 해석할 수 있으며, ‘Medical artificial intelligence’, ‘Pattern recognition’, ‘Medical technology’, ‘Anomaly detection’과 관련된 의료 XAI 연구는 코로나19의 영향으로 인해 더욱 중요한 주제가 되었다. XAI 기술을 활용한 의료 영상 분석 및 진단 지원, 증상 예측, 감염병 예방 등 다양한 목적으로 XAI가 활용되었으며 의료인공지능과 관련된 XAI 연구는 기존의 의료 진단 방식을 개선하고 진료 결과를 예측하는 모델의 정확도를 향상시키는 데 초점을 맞추는 연구가 진행되었다. 패턴 인식 분야에서는 이전에 비해 더욱 세밀하고 정확한 패턴 인식 기술이 개발되었다. 이를 통해 보안 및 감시 시스템에서 비정상적인 행동 패턴 등을 빠르게 탐지하고 예방할 수 있으며, 이상치 탐지 분야에서는 기업이나 조직에서 발생하는 이상 현상을 탐지하여 조기에 대처할 수 있도록 하는 연구가 각광받았다. 마지막으로, ‘Business application’에서는 XAI 기술을 이용하여 비즈니스 프로세스 자동화 및 최적화에 대한 연구가 진행되었으며 특히 XAI 기술을 활용하여 비즈니스 관련 유추 및 추론에 대한 연구가 활발했다. 인공지능 기반의 비즈니스 규칙을 자동화하여 효율적인 업무 처리를 지원하는 연구가 이루어지고 있으며 이를 통해 업무 프로세스의 자동화와 효율성 향상, 비즈니스 분석 등에 활용하고자 하는 연구가 진행되었다.

〈Table 3〉 Result of Topic Analysis(Continued)

TERM	TOPIC	Reinforcement learning	Alarm	Anomaly detection	Safety assurance.	AI reliability	Security and protection	Artificial intelligence	Medical application	Signal processing	Recognition and inference
4	1	Reinforcement learn	Code	IoT	Counterfactual	Malware	Knowledge	Student	Disease	Signal	Face
	2	Reinforcement	Attribute	Anomaly	Counterfactual explanation	Malware detection	Cluster	Tree	Object	Recognition	Knowledge
	3	Face	Student	Internet	Safety	Mortality	UNET	Smart	Recognition	Activity	Cluster
	4	Vehicle	Point	Thing	Generate counterfactual	Disease	Adversarial	Software	Deployment	Emotion	Activity
	5	Game	Feature attribution	Internet thing	Assurance	Question	Brain	Home	Plant	Word	Fuzzy
	6	Agent	Traffic	Intrusion	Knowledge	Production	Attack	Smart home	Pneumonia	EMG	Rule
	7	Trajectory	Encode	Sensitivity	Robot	Cancer	Deep-fake	Virtual	Claim	Segmentation	Graph
	8	Policy	Alarm	Intrusion detection	Slice	x-ray	Water	Decision tree	Crop	Lesion	Bias
	9	Attribute	Match	Anomaly detection	AMR	Answer	Defense	Head	Background	Facial	Mixture
	10	Employee	Signal	series	Safety assurance	failure	Cyber	Plan	Grad-cam	activation	Fuzzy system
	11						Glaucoma				

기간 4에서는 ‘Reinforcement learning’, ‘Alarm’, ‘Anomaly detection’, ‘Safety assurance’, ‘AI reliability’, ‘Security and protection’, ‘Artificial intelligence’, ‘Medical application’, ‘Signal processing’, ‘Recognition and inference’ 토픽이 등장하였다. ‘Reinforcement Learning’, 강화 학습은 인공지능 분야에서 핵심적인 기술 중 하나로, 2021년 인공지능 기술의 신뢰성 증진 연구가 주요 연구 대상이었지만, 2022년에는 안전 보장 연구가 강조됨으로써 AI 시스템의 안전성과 보안성 강화에 더욱 초점이 맞춰졌기에, 이에 따라 강화 학습을 활용하여 안전성과 신뢰성 모두를 고려한 모델을 생성하는 연구가 주목받았다. 또한 ‘Alarm’, ‘Anomaly detection’, ‘Safety Assurance’, ‘AI reliability’, ‘Security and protection’과 같은 주제들은 인공지능의 보안 및 보호와 안정성을 보장하기 위한 연구를 나타내며, 고도로 발전한 인공 지능 기술로 새로운 보안 위협에 취약해질 위험 또한 존재하며, 인공지능의 활용 분야가 넓어지면서 보안성 강화의 필요성이 더욱더 대두되고 있다고 볼 수 있다. 또한, ‘Medical application’, ‘Signal processing’, ‘Recognition and inference’과 같은 주제들은 의료 분야에서 XAI 기술의 적용 기술 및 가능성을 나타내며 인공 지능이 의료 분야에서 활용 영역이 더욱 커지고 있음을 보여준다.

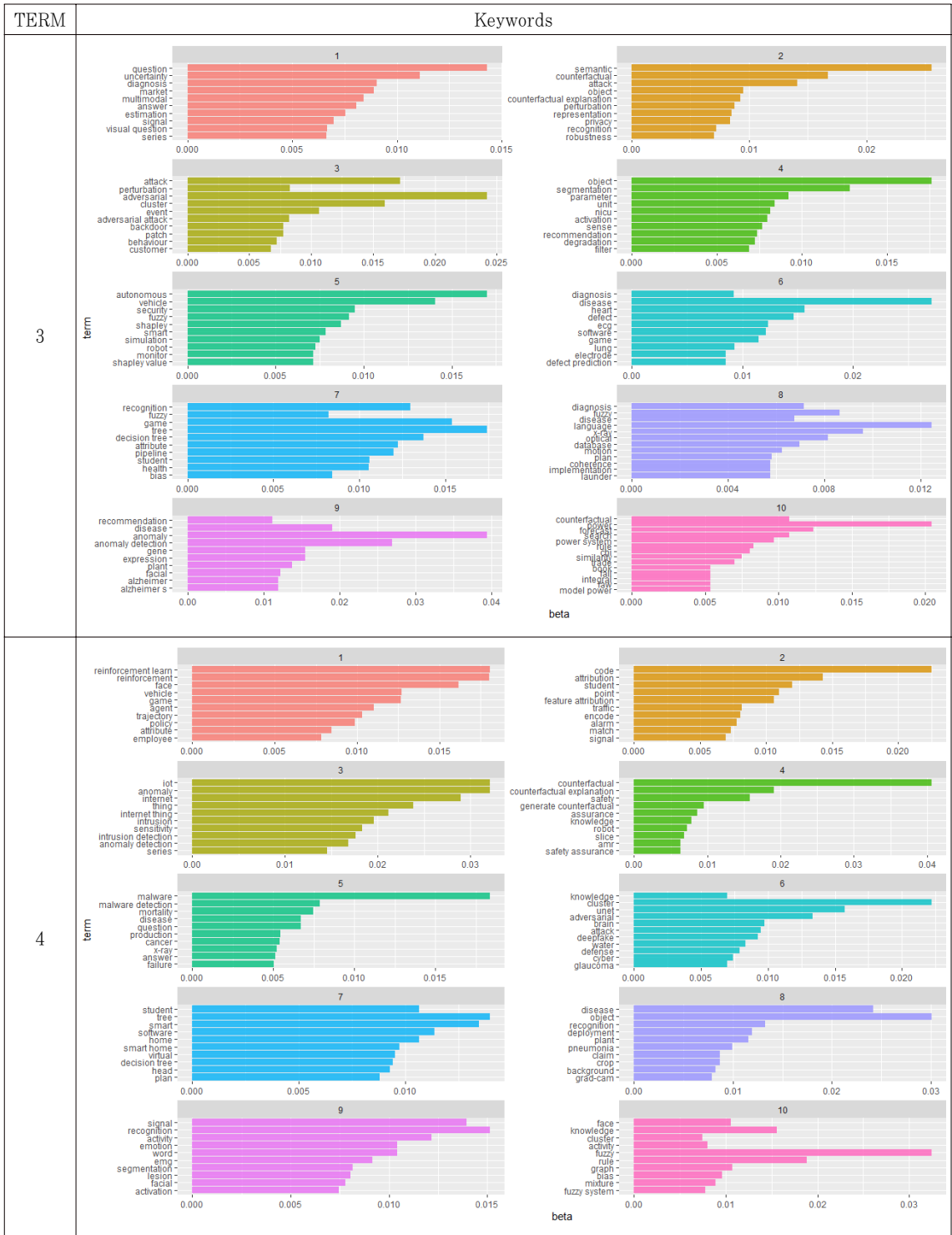
한편, 도출된 키워드의 분포도는 〈Table 4〉와 같다. 〈Table 4〉는 각 토픽에 해당하는 키워드들의 중요도를 보여주며, 토픽에 대한 세부적 설명이 된다. 기간 1의 토픽 ‘Disease detection’의 상위 키워드 ‘Cancer’,

‘Anomaly detection’, ‘Breast’, ‘Breast Cancer’은 XAI가 건강 이상 진단 분야에서 특히 유방암 진단 연구에 응용되었다고 볼 수 있다. ‘Medical technology’의 상위 키워드 ‘Document’, ‘Video’, ‘Invasive’, ‘Minimally’는 문서나 비디오를 보고 최소한의 외과적 진단을 할 수 있는 의료 XAI연구가 진행되었다고 해석할 수 있다. 기간 2에서 토픽 ‘Financial Application’에서 상위 키워드 ‘Forecast’을 보면, 2020년 코로나로 인한 힘든 경제로 인한 사회 상황은 금융 AI의 예측 성능에 대한 관심도를 볼 수 있다. XAI의 금융 분야 응용은 사용자들에게 신뢰성, 투명성, 정확성을 부여할 수 있기 때문이다. 또한 하위 키워드 ‘Anomaly Detection’에서 해석할 수 있듯, 금융 이상 거래 감지와 같은 기본적인 금융 인공지능에도 XAI가 도입되었다고 볼 수 있다. 2020년에 화두였던 또 하나의 인공지능 기술은 딥 페이크였다. 토픽 ‘Deep learning models’의 상위 키워드 ‘Deep-fake’, ‘Retrieve’는 보편화된 딥 페이크 기술의 악용과 남용으로 인한 사회적 피해와 통제 불가능한 파괴력으로 인한 위협성을 인지하고, 이를 해결하기 위한 딥 페이크 탐지 및 딥러닝 기반의 XAI연구가 활발히 진행되었다. 기간 3에서 ‘Medical technology’ 토픽에서 ‘Disease’, ‘Heart’, ‘Defect’, ‘ECG’와 같은 상위 키워드는 실제 2021년 다양한 웨어러블 기기에 심전도 기반 건강관리 인공지능이 다수 도입되었고, 이에 설명 가능 인공지능을 응용하여 실사용자들의 이해를 돕고 신뢰성을 부여하고자 하는 연구가 진행된 것으로 해석할 수 있다. 또한 토픽 ‘Anomaly Detection’의 키워드

<Table 4> Key Words Distribution



<Table 4> Key Words Distribution(Continued)



'Expression', 'Plant', 'Facial', 'Alzheimer'는 알츠하이머성 치매 진단의 설명가능성에 대한 연구가 진행되었다는 것을 보여준다. 'Diagnosis' 토픽에서 등장한 'Optical' 키워드는 녹내장 진단과 해석할 수 있도록 설명해주는 XAI가 개발되었고, 이러한 기술을 통하여 점점 다양한 의료 분야에서 XAI의 응용도가 높아지고 있는 것을 볼 수 있다. 기간 4의 토픽 'AI reliability'의 상위 키워드 'Malware', 'Malware Detection', 'Mortality'은 악성코드 탐지 분야에서의 활용도를 나타낸다. 사실 악성코드 탐지 인공지능은 많은 연구가 진행되어왔지만, 탐지 성능에 대한 실용성 및 판단의 타당성은 검증하기 어려움이 존재했다. 또한 코로나19로 인하여 디지털 소통의 급증으로 인하여 그동안 드러나지 않았던 보안 이슈들이 등장하였으며, 개개인의 사이버 피해 또한 속출하였다. XAI의 악성코드 탐지 분야 도입은 실용성 및 타당성의 한계를 극복할 뿐만 아니라 보안 분야의 신뢰성을 부여한다. 토픽 'Artificial Intelligence'의 상위 키워드 'Student', 'Smart', 'Tree'는 의사결정트리리를 이용한 XAI가 스마트 홈 시스템에 도입되었다고 해석할 수 있으며, 2022년은 보안, 금융, 의료와 같은 전문적인 분야뿐만 아니라 사람들의 실생활에도 XAI가 응용되는 것을 볼 수 있다.

4. 결 론

인공지능 사용이 익숙해진 현대 사회에서 개발자와 사용자의 간극을 좁혀주고, 인공지능에 대한 사용자의 신뢰도를 상승시키는 XAI는 인공지능이 보편화될수록 더욱더 필요성이 증가한다. 이에 XAI의 중요성을 인식하고, 본 연구에서 해외 학술지 IEEE로부터 XAI의 개념이 공식적으로 정립된 2004년부터 2022년까지 XAI, EXP 키워드와 관련한 저널 총 1,335편을 대상으로 네트워크 분석과 토픽 모델링을 진행하였다. 본 연구는 보편적인 텍스트 마이닝 절차인 데이터 수집, 데이터 전처리, 데이터 분석 및 해석 순으로 진행되었다. 2004년부터 2019년까지 XAI의 개념이 태동하였으나 개념적으로만 소개된 시기를 기간 1로 설정하였고, 2020년부터 본격적으로 XAI를 활용한 분석 결과를 다루는 연구가 나타났기에 1년 단위로 나머지 기간을 설정하여 분석을 진행하였다. SNA 결과를 시각화하여 주요 키워드를 한눈에 해석하고, 기간별 토픽을 도출함으로써 시간의 흐름에 따른

XAI연구의 동향을 다양하게 살펴보았다.

SNA 결과로는 기간 1에서 기계학습 및 테스트 검증과 관련한 XAI 연구가 진행되었으며, 기간 2에서는 예측과 계산 모델링에 대한 XAI 연구가 활발했다고 해석할 수 있다. 기간 3에서는 XAI 머신러닝 모델과정의 시각화 연구가 주목받았으며, 기간 4에서는 딥러닝의 복잡한 신경망에 XAI를 적용한 인공지능 및 다양한 분야에 적용하는 연구가 활발하였다고 해석된다.

토픽모델링 결과로는, 기간 1에는 기계학습에 XAI를 접목하는 기술적 측면에서의 연구가 주로 진행되었다고 볼 수 있으며, 기간 2에는 전 세계적으로 코로나의 영향을 받아 예측 및 의료와 보안 분야 적용 XAI연구가 활발하였다. 또한 기간 3에서는 코로나의 여파로 인하여 데이터 관련 분야의 연구와 의료 및 다양한 산업의 적용 연구가 이루어졌고, 기간 4에서는 신뢰성뿐만 아닌 안정성을 바탕으로 XAI연구를 적용하여 활용도를 더욱더 넓히고자 하는 연구가 진행되었다.

본 연구는 대량의 문서 데이터의 전체적인 패턴을 파악하기에 좋은 SNA와 문서 데이터의 내용을 파악하기에 적합한 토픽 모델링을 사용하여 XAI 연구에 기간별 학술지 내용 파악을 통한 추세를 분석하고자 하였다. 분석 결과, XAI는 기존의 머신 러닝과 블랙박스 문제를 해결하는 방법으로 등장하였으나, 필요성이 점차 증가하며 신뢰도 및 보안이 바탕으로 있는 금융 및 의료 분야 등 여러 분야에서 주목받고 있는 것을 확인하였고, 점차 실생활에 적용되고자 하는 연구가 활발하다는 것을 알 수 있었다. 본 연구의 결과를 통하여 XAI연구의 추세뿐만 아니라 향후 XAI활용에 대한 연구의 보완점에 대해서도 논의할 수 있다. 또한 후속 XAI 연구의 통찰력과 방향성을 기르기 위한 기초자료로 참고 될 수 있다.

본 연구는 IEEE 학술지만을 대상으로 XAI연구 동향 분석을 진행하였기에, XAI연구에 대한 다각적인 추세를 파악하기에는 한계가 있었다. 또한 XAI 개념이 최근 출현하였기에 향후 다양한 학술지에서 장기적인 추세 분석이 이루어진다면 더욱 확실하고 구분되는 추세 변화를 분석할 수 있을 것이다.

References

- [1] Bach, S., Binder, A., Montavon, G., Klauschen,

- F., Müller, K. R., and Samek, W., "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation", *PloS one*, Vol. 10, No. 7, 2015, e0130140.
- [2] Berry, M. and Linoff, "Data Mining Techniques: For Marketing, Sales, and Customer Support", Wiley, 1997, ISBN: 0471179809, 9780471179801.
- [3] Biecek, P., Chlebus, M., Gajda, J., Gosiewska, A., Kozak, A., Ogonowski, D. and Wojewnik, P., "Enabling Machine Learning Algorithms for Credit Scoring: Explainable Artificial Intelligence (XAI) methods for clear understanding complex predictive models", *arXiv preprint arXiv:2104.06735*, 2021.
- [4] Blei, D. M., "Probabilistic topic models", *Communications of the ACM*, Vol. 55, No. 4, pp. 77-84, 2012.
- [5] Blei, D. M., Ng, A. Y., and Jordan, M. I., "Latent dirichlet allocation", *The Journal of Machine Learning Research*, Vol. 3, 2003, pp. 993-1022.
- [6] Chen, B., "Deep Learning-Based Natural Language Processing Model for Explainable Personalization Recommendation Service", *Kyunghee University*, 2022.
- [7] Chu, H., Shin, H., Choi, S., Yoo, Y., and Park, C., "Sensitivity Analysis Using Explainable AI for Building Energy Use", *Journal of the Architectural Institute of Korea*, Vol. 38, No. 11, 2022, pp. 279-287.
- [8] Chun, Y., Kim, S., Lee, J., and Woo, J., "Study on credit rating model using explainable AI", *Journal of the Korean Data And Information Science Society*, Vol. 32, No. 2, 2021, pp. 283-295.
- [9] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R., "Indexing by latent semantic analysis", *Journal of the American Society for Information Science*, Vol. 41, No. 6, 1990, pp. 391-407.
- [10] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D., "A survey of methods for explaining black box models", *ACM Computing Surveys (CSUR)*, Vol. 51, No. 5, 2018, pp. 1-42.
- [11] Han, J. and Choi, J., "Explainable Artificial Intelligence", *Journal of KSNVE*, Vol. 27, No. 6, pp. 2017, 8-13.
- [12] Han, S. and Kim, T., "News Big Data Analysis of 'Metaverse' Using Topic Modeling Analysis", *Journal of Digital Contents Society*, Vol. 22, No. 7, 2021, pp. 1091-1099.
- [13] Hansen, D., Shneiderman, B., and Smith, M. A., "Analyzing social media networks with NodeXL: Insights from a connected world", *Morgan Kaufmann*, 2010.
- [14] Hofmann, T., "Probabilistic latent semantic indexing", In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1999, August, pp. 50-57.
- [15] Hwang, H., "Sentiment Analysis based on Semi-Supervised Learning for using XAI", *Seoul National University of Science and Technology*, 2022.
- [16] Hwang, S., "Knowledge graph-based UI for explainable recommendation", *Yonsei University*, 2021.
- [17] Jaegal, D., "The Applicability of Social Network Analysis for Program Evaluation in Public Sector", *Journal of Research Methodology*, Vol. 4, No. 3, 2019, pp. 1-32.
- [18] Jung, W., Yoon, J., and Kim, K., "Accounting Fraud Detection Using Forensic

- Techniques Based on Sentiment Analysis and Interpretable Machine Learning: Focused on Internal Control over Financial Reporting”, *Korean Accounting Review*, Vol.46, No.6, 2021, pp. 181-218.
- [19] Kang, K. and Park, S., “Keyword Analysis of KCI Journals on Business Administration using Web Crawling and Machine Learning”, *Korean Journal of Business Administration*, Vol.32, No.4, 2019, pp. 597-615.
- [20] Kim, B., Jeong, M., Jeon, S., and Shin, D., “Global research trends on geospatial information by keyword network analysis”, *Journal of Korea Spatial Information Society*, Vol.23, No.1, 2015, pp. 69-77
- [21] Kim, H. and Shin, S., “A Study on Explainable Artificial Intelligence (XAI)- based Machine Learning Feature Screening for Malicious Web Site Detection”, *Proceedings of Symposium of the Korean Institute of Communications and Information Sciences*, 2020, pp. 411-412.
- [22] Kim, M., Choi, J., and Kim, H., “The Framework of Research Network and Performance Evaluation on Personal Information Security: Social Network Analysis Perspective”, *Journal of Intelligence and Information System*, Vol. 20, No. 1, 2014, pp. 177-193.
- [23] Kim, M. and Hong, T., “Exploratory application of explainable artificial intelligence techniques to secure the reliability of deep learning-based personal credit evaluation model”, *Proceedings of the Korea Intelligent Information System Society Conference*, 2022, pp. 194-194.
- [24] Kim, S., Kim, D., and Kang, J., “A method suggesting the reason for the user recommendation: Design Sci Methodology methodology”, *Proceedings of Journal of Korea Service Management Society*, 2022, pp. 130-130.
- [25] Kim, S., Kim, W., Jang, Y., and Kim, H., “Development of Explainable AI-Based Learning Support System”, *The Journal of Korean Association of Computer Education*, Vol. 24, No. 1, 2021, pp. 107-115.
- [26] Kim, S. and Lee, G., “A Study on the Explainability and Hyperparameter Characteristics of Deep Neural Networks: A Case Study of SMEs Credit Scoring System”, *Journal of SME Finance*, Vol.42, No.1, 2022, pp. 3-37.
- [27] Kim, T., “COVID-19 News Analysis Using News Big Data: Focusing on Topic Modeling Analysis”, *The Journal of the Korea Contents Association*, Vol. 20, No. 5, 2020, pp. 457-466.
- [28] Lee, J., Park, J., Yoon, J., Lee, S., Yeom, S., Lee, Y., and Yoon, H., “Analyzing Precedents for Sports Match Fixing by Applying Topic Modeling”, *The Korean Journal of Measurement and Evaluation in Physical Education and Sports Science*, Vol. 23, No. 2, 2021, pp. 51-65.
- [29] Lee, S., Son, C., Shin, S., and Lee, K., “Airfoil Inverse Design using XAI(eXplainable Artificial Intelligence)”, *The Korean Society for Aeronautical and Space Sciences*, 2021, pp. 50-51.
- [30] Lundberg, S. and Lee, S., “A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*”, 2017, pp. 4765-4774.
- [31] Mohseni, S., Zarei, N., & Ragan, E. D. “A multidisciplinary survey and framework for design and evaluation of explainable AI systems”, *ACM Transactions on Interactive Intelligent Systems*

- (TiiS), 11(3-4), 2021, 1-45.
- [32] Moon, Y., "Analysis of Trends in Havruta Research based on Topic Modeling", *The Journal of Learner-Centered Curriculum and Instruction*, Vol. 20, No. 4, 2020, pp. 1149-1175.
- [33] Oh, H., Son, A., and Lee, Z., "Occupational accident prediction modeling and analysis using SHAP", *Journal of Digital Contents Society*, Vol. 22, No. 7, 2021, pp. 1115-1123.
- [34] Park, J. and Kwak, K., "The Effect of Patent Citation Relationship on Business Performance : A Social Network Analysis Perspective", *Journal of Intelligence and Information Systems*, Vol. 19, No. 3, 2013, pp. 127-139.
- [35] Park, S., Jeon, J., Kim, S., and Kim, S., "A Big-Data Analysis of Presidential Issue Ownership of two Prestigious Korean Newspapers: Focusing on LDA (latent Dirichlet allocation) Topic Modeling", *The Journal of Political Science & Communication*, Vol. 20, No. 3, 2017, pp. 25-55.
- [36] Ribeiro, M. T., Singh, S., and Guestrin, C., "Why should I trust you?" Explaining the predictions of any classifier", In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. pp. 1135-1144.
- [37] Shin, K., Lee, Y., Bae, B., Lee, S., Hong, H., Choi, Y., and Lee, S., "Trustworthy AI Framework for Malware Response", *Journal of the Korea Institute of Information Security & Cryptology*, Vol. 32, No. 5, 2022, pp. 1019-1034.
- [38] Shortliffe, Edward H., and Bruce, G., "A Model of inexact reasoning in medicine", *Mathematical Biosciences*, Vol. 23, No. 3-4, 1975, pp. 351-379.
- [39] Son, J., Woo, S., Paek, H., Hwang, B., and Choi, S., "False positive reduction in anomaly detection using XAI", *Proceedings of the Korean Information Science Society Conference*, 2022, pp. 609-611.
- [40] Van Lent, M., Fisher, W., and Mancuso, M. "An explainable artificial intelligence system for small-unit tactical behavior", *IAAI'04: Proc. 16th Conf. on Innovative Applications of Artificial Intelligence* (p./pp. 900-907), San Jose, CA, USA: AAAI Press, ISBN: 978-0-262-51183-4, 2004.

■ 저자소개

**문 건 두**

동국대학교 경영정보학과 박사과정을 수료하였으며 텍스트 마이닝을 이용한 트렌드 분석에 관심을 가지고 있다.

**김 경 재**

현재 동국대학교 경영대학 경영정보학과 교수로 재직 중이다. KAIST에서 경영정보시스템을 전공으로 박사학위를 취득하였으며, 연구 관심 분야는 비즈니스 애널리틱스, CRM, 추천기술, 빅데이터 등이다.