

도시 분야 인공지능(AI) 활용 연구 동향 분석: BERTopic 기반 토픽 모델링 분석을 활용하여

한요한*, 정진희**, 이경주***

*한국교통대학교 도시·교통공학과

**한국교통대학교 도시·교통공학과

***한국교통대학교 도시·교통공학전공

e-mail: snrn5008@naver.com

Analysis of Research Trends in the Use of Artificial Intelligence (AI) in Urban Areas: Using BERTopic-based Topic Modeling Analysis

HanYo Han*, Jinhee Jung**, Gyoungju Lee***

*Dept. of Urban & Transportation Engineering, Korea National University of Transportation

**Dept. of Urban & Transportation Engineering, Korea National University of Transportation

***Dept. of Urban & Transportation Engineering, Korea National University of Transportation

요약

최근 인공지능(AI) 기술은 복잡한 도시 현상을 분석하고 문제를 해결하는 데 있어 유효한 수단으로 부상하고 있으며, 이에 따라 도시 연구 분야에서도 AI 기술 활용에 대한 체계적인 분석의 필요성이 점차 증대되고 있다. 이 연구는 국내 도시 분야에서 인공지능 기술이 어떻게 활용되고 있는지를 정량적으로 파악하기 위해, 한국학술지인용색인(KCI)에 등재된 2020년 1월부터 2025년 3월까지의 논문을 대상으로 BERTopic 기반 토픽 모델링 분석을 수행하였다. 분석에는 논문 초록과 저자 키워드를 활용하였으며, Okl 형태소 분석기와 사용자 정의 불용어 리스트를 적용하여 텍스트 데이터를 정제한 후, Sentence-BERT 임베딩, UMAP 차원 축소, HDBSCAN 클러스터링을 결합한 BERTopic 기법을 적용하였다. 그 결과, 총 11개의 주요 토픽이 도출되었으며, 이를 바탕으로 연도별 연구 주제의 변화 추세를 분석하였다. 연구 결과, 도시 분야 인공지능 활용 주제는 특정 영역에 국한되지 않고 다양한 분야로 확산되는 경향을 보였으며, 특히 '차세대 모빌리티'와 '범죄예방형 도시재생'은 최근 급격히 부상한 핵심 주제로 확인되었다. 또한 일부 토픽은 시기별로 뚜렷한 증가 혹은 감소 추세를 나타내어, 도시 분야에서의 AI 활용 주제들이 시간에 따라 동적으로 변화하고 있음을 실증적으로 보여주었다. 이 연구는 국내 도시 분야에서 AI 기술이 적용된 연구의 흐름을 체계적으로 분석하였다는 점에서 의의가 있으며, 향후 도시 문제 해결을 위한 인공지능 활용 가능성과 연구 기획 방향 설정에 기초 자료로 활용될 수 있을 것으로 기대된다.

1. 서론

1.1 연구의 배경 및 목적

최근 빅데이터 기반의 인공지능(AI) 기술은 복잡한 도시 현상을 분석하고 문제를 해결하는 데 있어 강력한 도구로 부상하고 있으며(박중순·김창식, 2019; 제민희·정승희, 2024), 나아가 4차 산업혁명 시대를 이끄는 핵심 기술로 평가되고 있다(김창식 외, 2019). 이러한 기술의 발전과 함께 대규모 텍스트 데이터를 활용한 텍스트마이닝 연구 또한 다양한 관점에서 활발히 진행되고 있다(박중순·김창식, 2019). 텍스트마이닝(Text Mining)은 방대한 텍스트 데이터로부터 의미 있는 정보를 추출하고 잠재적 패턴을 도출하는 기법으로, 컴퓨터 과학 및 자연어 처리(Natural Language Processing, NLP) 기술의 발전에 기반하여 데이터 중심의 의사결정을 지원하는 핵심 분석 도구로 활용되고 있다(Tandel et al., 2019; Gnanavel et al., 2022). 그러나 기존 연구들은 대부분 전통적인 빈도 기반 분석에 의존하고 있어,

주요 연구 주제나 트렌드를 입체적으로 파악하는 데에는 한계가 존재한다(길원제 외, 2024). 특히, 도시 분야에서 인공지능 기술이 실제로 어떤 주제와 연계되어 연구되고 있는지에 대한 구조적이고 시계열적인 분석은 여전히 부족한 실정이다. 이에 이 연구는 대규모 텍스트 데이터를 정량적으로 분석함으로써, 국내 도시 분야에서 인공지능 기술이 어떻게 활용되고 있는지를 심층적으로 파악하는 것을 목적으로 한다. 이를 위해 텍스트마이닝 기법 중 하나인 BERTopic 기반 토픽 모델링 분석을 적용하여, 도시 분야의 인공지능 관련 연구 주제를 유형화하고 그 구조적 특성을 분석하였다. 또한 각 토픽에 대한 핵심 키워드를 정밀하게 도출함으로써, 전반적인 도시 분야의 인공지능 연구 동향을 체계적으로 파악하였다. 이 연구는 기존의 정성적 접근이나 단편적 키워드 분석을 넘어, 시간에 따른 주제 변화를 분석하였다는 점에서 기존 연구와의 차별성을 가진다. 이러한 분석 결과는 향후 도시 문제 해결을 위한 인공지능 기술의 활용 가능성과 연구의 방향성을 제시하는 데 기초 자료로 활용될 수 있을 것으로 기대된다.

1.2 연구의 범위

이 연구에서는 국내 논문의 연구 동향을 분석하기 위해 한국학술지인용색인(Korea Citation Index, 이하 KCI)을 활용하여 데이터를 수집하였다. KCI는 국내 전 학문 분야의 학술지 및 논문 정보를 제공하는 대표적인 학술 플랫폼이다. 분석 대상의 시간적 범위는 2020년 1월 1일부터 2025년 3월 31일까지로 설정하였으며, 최근 5년간 KCI에 등재된 논문 중 ‘도시’와 ‘AI’를 키워드로 검색하여 해당 논문의 제목과 초록을 수집하였다. 이후 전체 단어 수가 5개 이하인 논문, 초록이 누락된 문헌, 예술 및 문학 관련 세부 분야에 해당하는 논문은 분석에서 제외하였다. 최종적으로 565건의 논문이 분석 대상에 포함되었다.

2. 분석 방법

2.1 분석방법

이 연구의 분석 절차는 다음과 같다.

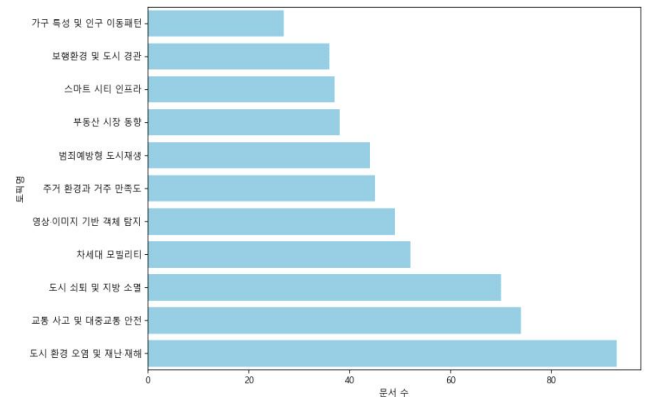
첫째, 앞서 수집한 논문의 제목, 초록, 저자 키워드를 결합하여 분석에 활용할 텍스트 데이터를 구성하였다. 둘째, 한국어 형태소 분석 도구인 Okt(Open Korean Text) 라이브러리를 활용하여 문장을 형태소 단위로 분해하였다. 형태소는 문장을 구성하는 가장 작은 의미 단위로, 분석의 기초 자료로 활용된다(김동현 외, 2023). 셋째, 텍스트 내에서 의미적 가치가 낮거나 분석에 불필요한 불용어를 제거하였으며, 이러한 전처리 과정을 통해 텍스트의 언어적 정확성과 분석의 신뢰도를 제고하였다. 넷째, 전처리된 데이터를 기반으로 BERTopic 알고리즘을 적용하여 토픽 모델링을 수행하였다. BERTopic은 문장의 의미를 벡터로 변환하는 문장 임베딩 기법과 의미적으로 유사한 문장을 군집화하는 클러스터링 기법을 결합한 주제(토픽) 분석 방법이다(길완제 외, 2024; 박재혁, 2024). 이 알고리즘은 기존의 확률 기반 모델보다 문맥 정보를 정교하게 반영함으로써 분석의 정밀도를 높일 수 있다(길완제 외, 2024). 다섯째, 도출된 세부 토픽 간 의미적 유사도를 계산하고, 계층적 클러스터링(hierarchical clustering) 기법을 적용하여 유사한 토픽들을 상위 토픽으로 통합하였다. 이후 각 상위 토픽에 대해 c-TF-IDF(class-based term frequency-inverse document frequency) 기반의 대표 키워드를 추출하여 주제의 성격을 명확히 규정하였다(남우민 외, 2024). 여섯째, 2020년부터 2025년까지의 연도별 상위 토픽 출현 비율을 분석하고, 전년 대비 증감률을 산출하였다. 산출된 변화량의 분포를 기준으로 상위 10%는 Hot, 하위 10%는 Cold, 나머지는 Neutral로 분류하였다. 이러한 일련의 분석을 통해 국내 도시 분야에서의 인공지능(AI) 연구 동향을 정량적으로 규명하였다.

3. 분석 결과

3.1 토픽 모델링 분석 결과

분석 결과, 총 37개의 세부 토픽이 도출되었으며, 계층적 클러스터링(hierarchical clustering) 분석을 통해 의미적으로 유사한 토픽들을 상위 범주로 통합하였다. 이 과정을 통해 최종적으로 11개의 상위 토픽으로 재구성되었으며, 각 토픽을 대표하는 상위 고빈도 단어 10개를 추출하였다. 또한, 각 토픽에 포함된 핵심 키워드를 종합적으로 분석하여, 해당 주제의 내용을 함축적으로 나타낼 수 있는 토픽 명칭(topic label)을 도출하였다([표 1] 참조). 도출된 토픽을 바탕으로 상위 토픽별 문헌 분포를 분석한 결과, ‘도시 환경 오염 및 재난·재해’(토픽 9, 16.5%), ‘교통사고 및 대중교통 안전’(토픽 11, 13.1%), ‘도시 쇠퇴 및 지방 소멸’(토픽 2, 12.4%)의 비중이 상대적으로 높게 나타났다. 반면, ‘가구 특성 및 인구 이동패턴’(토픽 5)은 4.8%로 비교적 낮은 연구 비중을 보였다([그림 1] 참조).

[그림 1] 상위 토픽별 문헌 분포 분석 결과



3.2 토픽별 추세 분석

3.2.1 전체 토픽 변화 개요

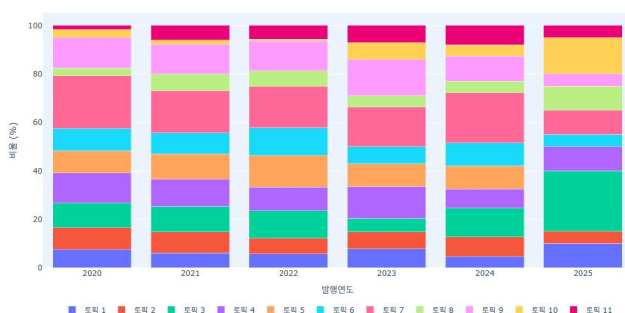
연도별 상위 토픽의 상대적 비율을 비교한 결과, 전반적으로 연구 주제가 점차 다양화되는 경향이 나타났다([그림 2] 참조). 특히 ‘스마트 시티 인프라’(토픽 7), ‘도시 환경 오염 및 재난·재해’(토픽 9), ‘부동산 시장 동향’(토픽 4)이 다른 토픽에 비해 상대적으로 높은 비중을 차지하였다. 토픽 분포의 다양성을 정량적으로 평가하기 위해 Shannon Entropy를 적용하였으며, 연도별 변화 추이를 분석한 결과 2020년(3.21)에 비해 2021년(3.32), 2023년(3.34)에서 명확한 증가세가 관찰되었다. Shannon Entropy는 정보이론에 기반한 지표로, 특정 분포 내 항목들의 불확실성 혹은 다양성을 정량화하는 데 사용된다.

[표 1] 토픽 모델링 결과

topic 1 주거 환경과 거주 만족도	topic 2 도시 쇠퇴 및 지방 소멸	topic 3 범죄예방형 도시재생	topic 4 부동산 시장 동향	topic 5 가구 특성 및 인구 이동패턴	topic 6 보행환경 및 도시 경관
주거	인구	범죄예방	가격	가구	보행
요인	지방	도시재생	주택	군집	거리
가구	쇠퇴	정비	아파트	유입	경관
만족도	빈집	범죄	매매	이동	인지
세대	축소	감시	거래	패턴	점수
리뷰	활성화	안전	감성지수	유입	영상
형태	젠트리피케이션	지속가능성	상권	이동	가로
거주	감소	공공시설	부동산	통근	녹색
일상생활	소멸	재건축	물가	주거지역	이미지
민원	주거	재개발	지수	관계성	정성
topic 7 스마트 시티 인프라	topic 8 영상·이미지 기반 객체 탐지	topic 9 도시 환경 오염 및 재난·재해	topic 10 차세대 모빌리티	topic 11 교통 사고 및 대중교통 안전	
스마트	영상	침수	포트	교통	
시티	위성	미세먼지	항공	통행	
계획	건물	홍수	드론	사고	
네트워크	해상도	대기오염	모빌리티	도로	
그린	피복	재난	무인	철도	
건축물	분할	피해	수단	다발	
중심성	도로	강우	이착륙	위험	
활성화	제작	오염	자율	운행	
경쟁력	객체	농도	도심	장치	
입지	이미지	재해	시설	버스	

이는 해당 값이 높을수록 각 항목 간 분포가 보다 균등하고 다양하게 이루어졌음을 의미한다. 이와 함께, 2022년 이후 일부 토픽의 비중이 두드러지게 증가하면서, 기존 토픽들과 공존하는 구조가 형성되었고, 이는 도시 분야에서 인공지능(AI)을 활용한 연구 주제의 스펙트럼이 점차 확장되고 있음을 시사한다.

[그림 3] 연도별 상위 토픽 비율



3.2.2 연도별 토픽 추세 분류

앞서 3.2.1절에서 확인한 연도별 상위 토픽의 비율을 보다 심층적으로 분석하기 위해, 각 연도별로 전년 대비 상위 토픽 비율의 변화량을 산출하였다. 이후 산출된 데이터를 바탕으로 변화량 분포를 분석하고, 해당 연도 내에서 상위 10%에 해당하는 토픽은 Hot 토픽, 하위 10%에 해당하는 토픽은 Cold 토픽으로 분류하였다.

2021년부터 2025년까지의 연도별 토픽 변화 추세에 대한 분류 결과는 다음과 같다. 2021년에는 ‘교통사고 및 대중교통 안전(토픽 11)’과 ‘영상·이미지 기반 객체 탐지(토픽 9)’가 Hot 토픽으로 분류된 반면, ‘스마트 시티 인프라(토픽 7)’ ‘차세대 모빌리티(토픽 10)’는 Cold 토픽으로 나타났다. 2022년에는 ‘가구 특성 및 인구 이동패턴(토픽 5)’과 ‘보행환경 및 도시 경관(토픽 6)’이 Hot 토픽으로 분류된 반면, ‘도시 쇠퇴 및 지방 소멸(토픽 2)’, ‘부동산 시장 동향(토픽 4)’은 Cold 토픽으로 나타났다.

2023년에는 ‘부동산 시장 동향(토픽 4)’과 ‘차세대 모빌리티(토픽 10)’가 Hot 토픽으로 부상한 반면, ‘범죄예방형 도시재생(토픽 3)’과 ‘보행환경 및 도시 경관(토픽 6)’은 Cold 토픽으로 전환되었다. 2024년에는 ‘범죄예방형 도시재생(토픽 3)’과 ‘스마트 시티 인프라(토픽 7)’가 Hot 토픽으로 재부상하였으며, ‘도시 환경 오염 및 재난·재해(토픽 9)’와 ‘부동산 시장 동향(토픽 4)’은 Cold 토픽으로 분류되었다. 2025년에는 ‘범죄예방형 도시재생(토픽 3)’과 ‘차세대 모빌리티(토픽 10)’가 Hot 토픽으로 나타난 반면, ‘가구 특성 및 인구 이동패턴(토픽 5)’과 ‘스마트 시티 인프라(토픽 7)’는 Cold 토픽으로 분류되었다.

도시 분야 인공지능(AI) 활용 연구의 연도별 토픽 변화 추이를 종합적으로 분석한 결과, 시기별로 연구 주제가 뚜렷한 변화를 보이는 것으로 나타났다. ‘차세대 모빌리티(토픽 10)’와 ‘범죄예방형 도시재생(토픽 3)’은 최근 2년 이상 연속하여 Hot 토픽으로 분류되었으며, 이는 해당 주제에 대한 연구가 지속적으로 확대되고 있

음을 시사한다.

반면, ‘부동산 시장 동향(토픽 4)’과 ‘스마트 시티 인프라(토픽 7)’는 시기에 따라 Hot과 Cold를 반복하는 변동성을 보였으며, 이는 정책 환경의 변화나 사회적 이슈에 민감하게 반응하는 주제 특성을 반영하는 것으로 해석된다. 또한, ‘가구 특성 및 인구 이동패턴(토픽 5)’과 ‘도시 환경 오염 및 재난·재해(토픽 9)’는 최근 들어 지속적으로 Cold 토픽으로 분류되고 있으며, 이들 분야에 대한 연구 비중은 점차 감소하는 경향을 보이고 있다.

4. 결론

이 연구는 대규모 텍스트 데이터를 정량적으로 분석하여, 국내 도시 분야에서 인공지능(AI) 관련 연구의 동향을 체계적으로 파악하는 것을 목적으로 한다. 이를 위해 한국학술지인용색인(KCI)에 등재된 논문 중 ‘도시’와 ‘AI’ 관련 키워드를 포함한 문헌을 대상으로, 텍스트마이닝 기법 중 하나인 BERTopic 기반 토픽 모델링 분석을 수행하였다. 분석 결과의 요약과 시사점은 다음과 같다. 첫째, 연구 주제가 특정 분야에 집중되기보다 다양한 영역으로 확산되고 있는 주제 다양화의 경향이 확인되었다. 특히 연도별 추세 분석을 통해 일부 토픽은 연속적인 성장 또는 감소를 보였으며, 일부는 급격한 변동 양상을 나타냈다. 이는 AI 기술의 고도화와 함께 도시 문제 해결을 위한 AI 활용 연구가 점차 다양화되고 있으며, 나아가 도시 문제의 복잡성과 인공지능 기술의 범용성이 동시에 확대되고 있음을 시사한다. 둘째, ‘차세대 모빌리티’와 ‘범죄예방형 도시재생’은 최근 2년간 급격히 부상한 주제로 나타났다. 이는 2023~2024년 산업통상자원부와 한국자동차연구원을 중심으로 자율주행차, UAM, 스마트 모빌리티 분야의 대규모 R&D 사업이 활발히 추진되고 있으며, CPTED(범죄예방환경설계)가 국정과제로 채택되어 스마트 방범 인프라 구축이 본격화된 정책적 흐름이 연구에 반영된 결과로 해석할 수 있다. 셋째, ‘부동산 시장 동향’과 ‘스마트시티 인프라’ 관련 주제는 시기별로 뚜렷한 변동성을 보였다. 이는 부동산 규제 정책의 변화, 금리 인상·인하, 스마트시티 시범사업 추진 등 외부 정책 환경 변화에 대한 민감한 반응으로 이해된다. 이 연구는 국내 도시 분야에서 인공지능 기술이 어떻게 활용되고 있는지를 텍스트마이닝 기반 정량 분석을 통해 체계적으로 고찰하였다는 점에서 의의가 있다. 또한 향후 도시 관련 연구 기획 및 정책 수립을 위한 기초 자료로 활용될 수 있는 가능성을 제시하였다. 그러나 다음과 같은 한계를 지닌다. 첫째, 자료 수집 시점의 한계로 인해 2025년 3월 이후에 발표된 논문 데이터는 분석에 포함되지 못하였다는 시의성의 제한이 존재한다. 둘째, 분석 대상에서는 인공지능 기술의 구체적 유형이나 적용 방식에 대한 세부적인 내용을 충분히 반영하지 못하였다. 향후 연구에서는 전처리 기법의 고도화를 통해 AI 기술의 종류별 분류 및 활용 목적에 따른 세분화된 분석이 가능하도록 개선함으로써, 보다 심층적인 도

시 분야 AI 활용 연구 동향을 파악하는 방향으로의 확장이 필요하다.

참고문헌

- [1] 감환배·이명훈, “우리나라 도시재생 분야 연구동향 및 지식구조 분석”, 대한국토·도시계획학회논문지, vol. 51, no. 7, 통권 225호 pp. 33-55, 2016.
- [2] 박건철·이치형, “토픽 모델링을 활용한 스마트시티 연구동향 분석”, 한국인터넷정보학회논문지, vol. 20, no. 3, pp. 119 -128, 2019.
- [3] 김동현·김도국·김철화·신명선·서영덕, “사용자 사전과 형태소 토큰을 사용한 트랜스포머 기반 형태소 분석기”, 스마트미디어저널, Vol.12, No.9, pp.19-27, 2023.
- [4] 박종순·김창식, “빅데이터 연구동향 분석: 토픽 모델링을 중심으로”, 디지털산업정보학회논문지, vol. 15, no. 1, pp. 1-7, 2019.
- [5] 김창식·김남주·곽기영, “머신러닝 및 딥러닝 연구동향 분석: 토픽모델링을 중심으로”, 디지털산업정보학회논문지, vol. 15, no. 2, pp. 19 - 28, 2019.
- [6] 박한샘·김동현·장성주, “구조적 토픽 모델링 기반 스마트 시티 연구 동향 분석”, 디지털콘텐츠학회논문지, vol. 20, no. 9, pp. 1839 - 1846, 2019.
- [7] 장선영·정승현, “토픽모델링을 이용한 도시 분야 연구동향 분석”, 한국산학기술학회논문지, vol. 22, no. 3, pp. 661 - 670, 2021.
- [8] 제민희·정승현, “빅데이터와 인공지능 기술 도입에 관한 도시계획 분야 직종 간 인식분석”, 한국산학기술학회논문지, vol. 25, no. 1, pp. 469-477, 2024.
- [9] 김세진, “텍스트마이닝을 활용한 국내 문화도시 지정사업 관련 연구 동향 분석”, 한국엔터테인먼트산업학회논문지, vol. 17, no. 8, pp. 605 - 613, 2023.
- [10] 김완재·장환영·신인수, “KeyBERT와 BERTopic을 활용한 텍스트마이닝 연구동향 분석”, 한국엔터프라이즈아키텍처학회논문지, vol. 21, no. 2, pp. 159-169, 2024.
- [11] 남우민·양재훈·박건철, “BERTopic을 활용한 건설·부동산 분야 신기술 동향 분석”, 한국인터넷정보학회논문지, vol. 25, no. 6, pp. 71-80, 2024.
- [12] Mirshafiee, M., Barcomb, A., & Tan, B., “Advancing Smart Cities through Novel Social Media Text Analysis: A Case Study of Calgary,” 2023 IEEE Symposium Series on Computational Intelligence (SSCI), 2023, pp. 1-8.
- [13] Yang Liu, Fanghao Wan, “Unveiling temporal and spatial research trends in precision agriculture: A BERTopic text mining approach,” Heliyon, vol. 10, no. 17, p. e36808, 2024.
- [14] Smith, P., Gould, M., Manley, E., “The Rise and Fall of Twitter Data in Urban Analytics,” Findings, 2025.
- [15] Tandel, S. S., Jamadar, A., & Dudugu, S., “A Survey on Text Mining Techniques,” Proceedings of the 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019.
- [16] Gnanavel, S., Mani, V., Sreekrishna, M., Amshavalli, R. S., Gashu, Y. R., Duraimurugan, N., & Rao, N. S., “Rapid Text Retrieval and Analysis Supporting Latent Dirichlet Allocation Based on Probabilistic Models,” Mobile Information Systems, 2022.