

토픽 모델링과 의미망 분석을 활용한 땅밀림 관련 뉴스 기사와 연구 논문의 동향 분석

윤상후¹ · 광민서² · 김동엽^{3*}

¹전남대학교 통계학과, ²전남대학교 빅데이터융합학과, ³대구대학교 산림자원학과

Analysis of News Articles and Research Trends on Landcreep Using Topic Modeling and Semantic Network Analysis

Sanghoo Yoon¹, Minseo Kwak² and Dongyeob Kim^{3*}

¹Department of Statistics, Chonnam National University, Gwangju 61186, Korea

²Department of Big Data Convergence, Chonnam National University, Gwangju 61186, Korea

³Department of Forest Resources, Daegu University, Gyeongsan 38453, Korea

요약: 기후변화로 인한 자연재해 발생 빈도가 증가함에 따라 땅밀림 현상이 사회적 그리고 학문적 관심을 받고 있다. 본 연구는 2004년부터 2024년까지 수집된 뉴스 기사와 연구 논문의 텍스트 데이터를 대상으로 텍스트 마이닝 기법, 특히 토픽 모델링과 의미망 분석을 적용하여 땅밀림과 관련된 주요 이슈와 연구 동향을 분석하였다. 뉴스 기사 분석에서는 주로 지역 사회의 대응 노력, 정부 정책, 기술 도입이 강조된 반면, 연구 논문 분석에서는 땅밀림의 지질학적 특성, 예측 모델, 첨단 기술 활용 방안이 중심이 되었다. 두 데이터는 모두 예방적 접근의 중요성을 공유하며, 뉴스 기사는 실질적 사례를, 연구 논문은 이론적 기초와 기술적 해결책을 제시한다. 본 연구의 결과는 땅밀림 관련 연구의 초기 단계에서의 학문적 공백을 메우고, 향후 연구 및 정책 방향에 중요한 통찰을 제공할 것으로 기대되었다.

Abstract: As the frequency of natural disasters caused by climate change increases, academics and the general public alike have become increasingly concerned with landcreep phenomena. This study applies text mining techniques, particularly topic modeling and semantic network analysis, to text data collected from news articles and research abstracts published between 1995 and 2024. The analysis of news articles highlights community response efforts, government policies, and technology adoption, whereas the research abstracts focus on the geological characteristics of landcreeps, prediction models, and the application of advanced technologies. Both the practical cases in the news articles and the theoretical foundations and technical solutions in the research papers emphasize the importance of a preventive approach. This study addresses an academic gap in the early stages of landcreep research and provides important insights for future research and policy directions.

Key words: landcreep, topic modeling, semantic network analysis, news articles, research abstracts

서론

전세계적으로 기후변화 및 인간활동으로 인하여 다양한 종류의 자연재해가 새롭게 발생하고 있으며, 그 빈도와 규모가 증가하고 있다. 국토의 63%가 산림지역인 우리나라에서도 새로운 산림재해의 발생 빈도와 규모가 증가하고 있는데, 이 중에서 2010년대 중반 이후 일반 대중에게 비교적 새롭게


등장한 산림재해로는 산지토사재해의 일종인 땅밀림을 들 수 있다. 땅밀림은 토층 내 점토층의 존재나 암반층 피암면지하수의 영향으로 토층 혹은 토괴(土塊)가 느린 속도로 미끄러져 내려가는 현상으로 정의된다. 산사태와 다르게 특정 지역의 불안정한 지질이나 산림 파괴, 인간의 개발 활동이 장기간에 걸쳐서 땅밀림에 누적되는 영향을 미칠 수 있으며 그 피해의 규모 또한 보통의 산사태보다 더 클 수 있다는 점(Seo et al., 2021)에 있어서, 땅밀림에 대한 사회적 관심이 요구되는 시점이다.

땅밀림에 관한 연구는 1990년대 중반부터 2000년대 중반까지 특정 지역의 발생 사례(Woo et al., 1996; Park et al.,

* Corresponding author

E-mail: dy.kim@daegu.ac.kr

ORCID

Dongyeob Kim  https://orcid.org/0000-0001-5243-0780

2003; Park et al., 2005) 를 중심으로 진행되기 시작하였으나, 이후 약 10년간 연구가 거의 이루어지지 않았다. 하지만 2010년대 중반 이후 경주시 양북면에서의 땅밀림 사례(Park and Park, 2019)와 포항 지진에 의한 땅밀림(Ministry of Culture, Sports and Tourism, 2018)이 보고되면서 관련 연구가 다시 주목받기 시작했다. 땅밀림은 발생지역의 입지환경 특성 분석과 발생 원인 규명, 그리고 복구·복원 측면에서 2010년 중반 이후 산지토사재해 분야의 중요한 연구 분야(혹은 주제) 중 하나로 자리 잡았다. 특히, 기후변화에 따른 강우량 증가 및 인위적 산지개발 등의 영향으로 땅밀림의 발생·활동 가능성이 높아지면서, 이에 대한 학문적 연구는 더욱 시급한 과제로 대두되었다.

한편, 땅밀림과 같은 자연재해로 인한 인명 및 재산 피해를 효율적으로 저감하기 위해서는 이에 대한 원인 규명 및 특성 분석 등의 연구적인 측면도 중요하지만, 대중들의 인식 제고 역시 필수적이다. 대중이 땅밀림과 그 위험성을 올바르게 인식하고 그로 인한 피해를 사전에 예방할 수 있도록, 언론은 땅밀림의 위험성과 예방 방법에 대한 정보를 적극적으로 알리는 역할을 해야 한다. 실제로 땅밀림 현상에 대해서는 2016년 이후 언론을 통한 관련 보도가 증가하면서 대중들의 관심을 끌기 시작했으며, 지속된 관련 보도를 통하여 인근 지역 주민들이 경각심을 갖게 되었으며, 이렇게 증가된 사회적 관심은 정책 입안자들로 하여금 실질적인 대응 정책을 마련하도록 하는 계기가 되었다. 이러한 측면에서 땅밀림과 같은 산림재해 분야에서는 연구 논문뿐만 아니라 뉴스 기사를 함께 분석하는 것이 보다 폭넓은 함의와 통찰을 제공하고 이를 통하여 미래 연구의 방향성까지도 제시될 수 있다.

본 연구에서는 땅밀림 관련 연구 논문의 초록과 뉴스 기사를 수집하고 텍스트 마이닝 기법인 토픽 모델링(Topic Modeling)과 의미망 분석 방법을 적용하였다. 토픽 모델링은 다수의 문서에서 단어 출현 빈도를 확률적으로 분석하여 문서의 대표 주제어를 추출하는 분석 기법이다(Blei, 2012). 이 기법은 텍스트 데이터를 효과적으로 처리할 수 있고 특정 기간의 트렌드와 동향을 분석하는 데 유용하다. 토픽 모델링은 뉴스 기사 데이터를 분석하여 특정 시기별 주요 이슈를 도출하거나(Moon et al., 2018), 학술 논문의 초록을 분석하여 연구 동향을 파악하는 데(Park and Song, 2013; Lee and Kim, 2024) 적절한 기법으로 널리 사용되고 있다. 또한, 의미망 분석은 문서 내 단어 간의 관계를 노드와 링크로 시각화하여, 해당 단어들이 어떻게 상호작용하는지 파악하는 기법이다. 의미망 분석은 단어들의 동시출현빈도, 상관관계수, 엔그램(N-gram) 등을 통해 특정 단어나 개념들이 문서 내에서 어떻게 연결되는지, 그리고 어떤 주제들이 상호 연관되는지를 시

각적으로 명확히 보여준다(Borgatti et al., 2009). 이 방법은 학술 논문이나 뉴스 기사에서 주요 키워드를 추출하고, 이들 간의 관계를 탐색하는 데 효과적인 기법으로 활용된다(Kim et al., 2021)

토픽 모델링이나 의미망 분석을 활용한 선행연구는 대개 연구 논문이나 뉴스 기사 중 하나에 초점을 맞추어 진행되는 경우가 많다. 산림분야에서는 산지토사재해 분야(Lee et al., 2017), 산림생태 분야(Lee and Lee, 2021), 산림치유 분야(Song et al., 2022), 또한 북한 산림 분야(Lim et al., 2020) 등에 연구 논문을 대상으로 하여 연구동향 분석을 실시한 사례가 있다. 그러나 본 연구는 연구 논문 및 뉴스 기사 등 두 종류의 연구자료를 통합적으로 활용함으로써 더욱 풍부하고 신뢰성 있는 결과를 도출하여 학문적 통찰과 대중적 시각을 동시에 아우르는 분석을 실시하고자 한다. 연구 논문은 특정 주제에 대한 이론적 근거와 분석을 바탕으로 학술적 통찰을 제공하지만, 일반 대중과의 거리감이 존재한다는 한계가 있다. 반면, 뉴스 기사는 대중적 관점에서 즉각적인 사회적 반응을 반영한다. 따라서 두 데이터의 통합적 분석은 학문적 통찰과 사회적 인식을 유기적으로 연결함으로써, 연구 결과의 실질적 적용 가능성을 높이고 사회적 요구를 반영한 실용적인 연구 방향을 제시할 수 있다. 분석된 연구 결과를 기반으로 땅밀림에 대한 사회적 관심과 학문적 연구 동향을 명확히 파악하고, 최종적으로는 향후 연구의 방향성을 제시하고자 하였다.

재료 및 방법

1. 연구자료 수집

본 연구에서는 땅밀림 현상에 대한 언론 보도를 분석하기 위해 한국언론진흥재단의 뉴스 빅데이터 분석 시스템인 빅카인즈(BIGKINDS)를 활용했다. 빅카인즈에서 ‘땅밀림’ 키워드를 포함하는 뉴스 기사를 수집하고, 한국언론진흥재단과 바이칼AI(Baikal AI)가 공동 개발한 형태소 분석기 바른(bareun)을 이용하여 기사의 핵심 키워드를 추출했다. 분석 대상 기간은 2004년 1월부터 2024년 4월까지이며, 중복 기사를 제외한 총 208건의 기사가 최종 분석에 사용되었다.

한편, 국내 학술지 정보, 논문 정보 및 참고문헌을 DB화하여 논문 간 인용 관계를 분석하는 시스템인 KCI를 활용하여, 땅밀림 관련 논문의 초록을 수집하였다. 논문 검색 키워드는 뉴스 기사 검색과 동일하게 ‘땅밀림’이고, 검색된 논문의 기간은 2004년 1월부터 2024년 4월까지로 총 40편의 논문이 수집되었다.

2. 연구방법

텍스트 마이닝(Text Mining)은 대규모 비구조화 텍스트 데이터에서 유의미한 패턴과 정보를 추출하는 과정으로, 최근 다양한 학문 분야에서 활발히 활용되고 있다(Lim et al., 2023; Yoon and Kim, 2023; Lee et al., 2024; Park and Yoon, 2024). 이 기법은 자연어 처리(NLP; Natural Language Processing)와 통계적 분석을 결합하여 텍스트 데이터를 구조화된 형태로 변환하고, 이를 바탕으로 숨겨진 정보를 도출하고 관련 통찰력을 제시한다(He et al., 2017). 토픽 모델링이나 감성 분석과 같은 고도화된 텍스트 마이닝 기법은 단순히 정보를 요약하는 것을 넘어서는 심층적인 통찰을 제공할 수 있다(Aggarwal and Zhai, 2012). 본 연구에서는 토픽 모델링과 의미망 분석을 적용하여, 땅밀림 관련 연구자료를 분석하였다. 토픽 모델링과 의미망 분석을 수행하기 위해 R 4.3.2 프로그램을 활용하였다. 토픽 모델링은 topicmodels 패키지를 사용하여 문서 집합 내의 주요 주제를 추출하였고(Grün and Hornik, 2011), 의미망 분석은 igraph 패키지를 통해 단어 간의 관계를 시각화하였다(Csardi and Nepusz, 2006).

1) 토픽 모델링

토픽 모델링은 문서를 다양한 주제의 확률적 혼합체로 표현하는 분석 방법으로서, 문서 내 숨겨진 주제를 발견하는데 중점을 둔다. 이 기법은 문서에서 각 주제(topic)를 단어의 분포로 모델링하고, 통계적 추론을 통해 잠재적 주제를 추출한다. 주로 언론 보도나 문서에서 특정 이슈를 식별하는 데 사용되며, 문서 내 단어 출현 빈도를 바탕으로 주제를 추출하는데 집중한다(Blei, 2012).

과거에는 주로 LSA(Latent Semantic Analysis)이나 pLSA(Probabilistic Latent Semantic Analysis) 기반 토픽 모델링을 활용하였으나, 최근에는 기존 LSA 및 pLSA의 한계점을 개선하여 문서 내 숨겨진 주제를 효과적으로 발견하는 방법으로 LDA(Latent Dirichlet Allocation)가 널리 활용되고 있다(Blei et al., 2003). LSA는 문서 내 단어들의 잠재적 주제를 추출하기 위해 특잇값 분해(singular value decomposition)를 사용하는 방법인데(Landauer et al., 1998), 문서 내 단어 분포가 가우시안 분포(Gaussian distribution)를 따른다는 가정이 실제 텍스트 데이터의 특성과 일치하지 않을 수 있다는 한계점이 있다. 또한, pLSA는 단어 빈도 대신 단어 출현 확률을 바탕으로 분석함으로써 LSA의 한계점을 어느 정도 극복할 수는 있으나(Hofmann, 1999), 문서 수준에서 확률 생성 모델을 제공하지 못하는 한계가 있다. 반면에 LDA는 문서 내 숨겨진 주제를 추출하는 생성 모델로서 pLSA보다 더 복잡한 구조를 가지고 있으나, 과적합(overfitting)의 위험이 적고 일반화 성능은 더 뛰어나다. LDA는 문서 작성

과정을 모델링하며, 각 문서에 포함될 주제와 단어를 선택하는 매개변수(parameter)를 설정함으로써, 관찰된 변수인 문서와 단어를 통해서 보이지 않는 변수인 주제와 그 비율을 추론할 수 있다. 즉, LDA는 대규모 비정형 텍스트 데이터에서 주요 주제를 효과적으로 추출하고, 각 주제에 대한 단어 분포를 확률적으로 파악하여 분석 결과의 해석 가능성이 높다는 장점이 있다. LDA의 확률 분포는 다음과 같이 정의된다.

$$p(\theta, Z, W | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(Z_n | \theta) p(W_n | Z_n, \beta) \quad (1)$$

여기서, θ =문서의 토픽 분포, Z =단어의 토픽 할당, W =관측된 단어, α, β =사전 매개변수이다. 식 (1)은 사전 매개변수 α, β 를 기반으로 문서의 토픽 분포(θ), 단어들의 토픽 할당(Z) 및 관측된 단어(W)의 각 확률을 결합확률로 계산하는 것을 의미한다. 구체적으로 식 (1)의 첫 번째 항 $p(\theta | \alpha)$ 는 문서별 토픽 분포 θ 에 대한 사전 확률로서 이는 디리클레 분포(Dirichlet distribution)를 따른다. 두 번째 항 $\prod_{n=1}^N p(Z_n | \theta)$ 는 각 단어가 특정 토픽에 할당될 확률을 의미하며, 문서의 토픽 분포(θ)에 의존한다. 마지막 항 $\prod_{n=1}^N p(W_n | Z_n, \beta)$ 는 특정 토픽에 의해 단어가 생성될 확률로서, 이는 토픽의 단어 분포(β)에 의해 결정된다.

Figure 1은 LDA 연산 과정을 그래프 모델로 표현한 그림으로서, 문서 생성 과정을 시각적으로 설명한다(Blei et al., 2003). Figure 1은 LDA 연산 과정을 나타내는 그래프 모델이다. 이 모델에서 상자는 반복적인 과정을 나타내며, D 는 문서의 수, N 은 각 문서의 단어 수, K 는 토픽 수를 의미한다. 연산 과정은 사전 매개변수 α 와 η 를 기반으로 각 문서의 토픽 분포 θ_d 와 토픽의 단어 분포 β_k 를 샘플링하는 것으로 시작한다. 이어서 각 단어 $W_{d,n}$ 에 대해 토픽 $Z_{d,n}$ 을 문서의 토픽 분포 θ_d 에서 샘플링하고, 선택된 토픽 $Z_{d,n}$ 에 따라 단어 $W_{d,n}$ 을 $\beta_{Z_{d,n}}$ 에서 생성하는 과정을 반복한다(Park and Song, 2013). 최종적으로 깃스 샘플링을 활용하여 θ, Z, β 를 각각 추정하며, 이를 통해 문서의 잠재적인 토픽 구조를 이해하고 추출할 수 있다.

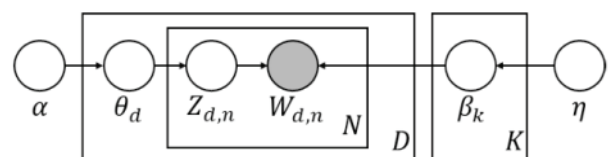


Figure 1. The graph model of LDA.

2) 의미망 분석

의미망 분석은 개념이나 단어 간의 관계를 네트워크 형태로 표현하고 분석하는 방법이다(Sowa, 1984). 이 분석 기법은 개념 간의 상호작용과 관계를 시각적으로 표현함으로써, 복잡한 정보나 지식 구조를 보다 명확하게 이해하는데 활용될 수 있다. 의미망 분석은 주로 시맨틱 네트워크(semantic network)와 개념적 그래프(conceptual graph)와 같은 구조를 통해 개념 간의 연결을 시각화하여, 개념의 관계를 추론하고 새로운 정보나 지식을 발견하는 데 유용하다. 궁극적으로 특정 주제나 도메인 내에서 중요한 개념들이 어떻게 연결되는지 파악할 수 있다. 또한, 네트워크 중심성과 연결 강도를 평가하는 기법을 통해 주요 개념을 식별하고, 그들의 상호작용을 보다 체계적으로 분석할 수 있다(Borgatti et al., 2009).

이 연구에서는 의미망 분석 기법 중 엔그램(N-gram) 방법을 적용하였다. 이것은 텍스트 데이터를 연속된 n개의 단어로 묶어 의미 있는 패턴을 추출하는 기법이다(Shannon, 1948). 이 방법은 텍스트 내에서 자주 등장하는 단어 쌍(2-그램), 세 개의 단어(3-그램), 또는 더 긴 연속적인 단어들의 패턴을 파악하는 데 사용된다. 엔그램 모델은 자연어 처리에서 특히 중요한 역할을 하며, 단어 간의 순서와 상호 연관성을 고려하는 통계적 방법을 사용한다. 엔그램 방법은 최초 개발 당시에는 자연어 처리에 주로 활용되었지만, 점차 활용성이 확대되어서 자동 번역, 텍스트 요약, 감정 분석 등 다양한 응용 분야에서도 활용되고 있다(Manning et al., 1999). 특히, 엔그램 기반의 언어 모델은 특정 단어가 주어진 문맥에서 등장할 확률을 예측할 수 있어, 문장 구조와 의미를 더 정확하게 모델링한다. 본 연구는 N=2인 바이그램을 이용하였다.

결과 및 고찰

1. 연구자료 수집 결과

수집된 땅밀림 관련 뉴스 기사 208건의 주요 키워드를 정리하면 Table 1과 Figure 2와 같다. 땅밀림 관련 언론 기사에서 추출된 주요 키워드의 빈도 및 비율은 다음과 같다. 비율은 키워드의 빈도를 뉴스 기사 총 건수(208건)로 나눈 값이다. ‘지역(area)’ 키워드가 172회(82.7%)로 가장 높은 빈도를 보였으며, 이는 땅밀림 현상이 특정 지역에서 발생하거나 지역적 영향을 미치는 경우가 많음을 시사한다. 이어서 ‘산사태(landslide)’가 165회(79.3%), ‘발생(occurrence)’이 157회(75.5%), ‘땅밀림(landcreep)’이 155회(74.5%)로 높은 언급 빈도를 나타냈다. 그 외 ‘피해(damage)’ (142회, 68.3%), ‘산림청(Korea Forest Service)’ (126회, 60.6%), ‘위험(danger)’ (112회, 53.8%), ‘관리(management)’ (109회, 52.4%), ‘복구(rehabilitation)’ (107회, 51.4%), ‘조사(survey)’ (107회, 51.4%)

Table 1. Top 12 keywords from news articles.

Keyword	Frequency	Ratio
지역(area)	172	82.7%
산사태(landslide)	165	79.3%
발생(occurrence)	157	75.5%
땅밀림(landcreep)	155	74.5%
피해(damage)	142	68.3%
산림청(Korea Forest Service)	126	60.6%
위험(danger)	112	53.8%
관리(management)	109	52.4%
복구(rehabilitation)	107	51.4%
조사(survey)	107	51.4%
계획(plan)	99	47.6%
검사(inspection)	98	47.1%



Figure 2. The wordcloud of news articles.

등 땅밀림과 관련된 다양한 측면의 키워드들이 50% 이상의 빈도를 보였다. 이는 언론에서 땅밀림 현상 자체의 발생과 함께 그로 인한 피해, 그리고 관련 기관의 관리 및 복구 노력에 대해 비중 있게 다루고 있음을 보여준다.

한편, 땅밀림에 관한 연구 동향을 조사하기 위해 수집된 총 40건의 논문 초록의 주요 키워드를 정리하면 Table 2와 Figure 3과 같다. 연구논문에서는 ‘땅밀림(landcreep)’ 키워드가 36회(90.0%)로 압도적으로 높은 빈도를 보였는데, 이는 논문 조사 시 키워드로 ‘땅밀림’을 이용했기 때문이다. 다음으로 ‘연구(research)’가 34회(85.0%), ‘발생(generation)’이 32회(80.0%), ‘결과(result)’가 29회(72.5%), ‘분석(analysis)’이 28회(70.0%)의 높은 빈도를 보였다. 또한 ‘지역(area)’ (27회, 67.5%), ‘특성(characteristic)’ (21회, 52.5%) 등의 키워드도 50% 이상의 빈도를 나타내었다. 반면 ‘산사태 취약 지역(landslide-prone area)’(18회, 45.0%), ‘수행(conduct)’ (18회, 45.0%), ‘산사태(landslide)’(17회, 42.5%), ‘지형(topography)’(16회, 40.0%), ‘제공(offer)’(15회, 37.5%) 등의 키워드는 상대적으로 낮은 빈도를 보였다. 이러한 결과는

Table 2. Top 12 keywords from research abstracts.

Keyword	Frequency	Ratio
땅밀림(landcreep)	36	90.0%
연구(research)	34	85.0%
발생(generation)	32	80.0%
결과(result)	29	72.5%
분석(analysis)	28	70.0%
지역(area)	27	67.5%
특성(characteristic)	21	52.5%
산사태 취약지역(landslide-prone area)	18	45.0%
수행(conduct)	18	45.0%
산사태(landslide)	17	42.5%
지형(topography)	16	40.0%
제공(offer)	15	37.5%



Figure 3. The wordcloud of research abstracts.

분석된 자료가 땅밀림 현상 자체의 발생 원인, 특성, 그리고 이에 대한 연구 및 분석 결과에 중점을 두고 있음을 시사한다.

Table 3에서 제시된 바와 같이 2016년 이후 땅밀림 관련 뉴스 기사의 수가 증가하였고, 2018년 이후에는 연구 논문의 수도 급증한 것으로 나타났다. 이러한 증가는 당시 국내에서 발생한 주요 재난과 밀접한 관련이 있는 것으로 분석된다. 구체적으로 2016년 경주 지진과 2017년 포항 지진의 영향으로 해당 지역 인근에서 땅밀림이 발생하였다는 보도로 인하여, 국민적 경각심이 더욱 커진 것으로 추정된다. 특히 포항 지진의 경우, 액상화와 함께 땅밀림 현상이 한반도에서 처음으로 관측된 사례로 기록되어, 관련 연구 및 언론 보도의 증가에 영향을 미쳤다.

이러한 상황 인식에 따라 산사태 관리 당국은 땅밀림 위험 지역에 대한 조사 및 예방 대책을 강화하기 시작했다. 특히 산림청은 전국 땅밀림 우려지역을 추출하여 매년 현장조사를 시행하고 있으며, 무인원격감시시스템을 구축하고 땅밀림 위험지도를 개발하려는 등 선제적인 예방 및 대응 체계 구축에 적극적인 노력을 기울이고 있다. 이는 땅밀림 재해로 인한

Table 3. The number of news articles and research abstracts by year.

Year	News articles	Research abstracts
2004	1	1
2005	1	3
2006	1	1
2007	0	0
2008	0	0
2009	1	0
2010	0	0
2011	1	0
2012	0	0
2013	11	0
2014	4	0
2015	2	3
2016	22	2
2017	22	0
2018	26	3
2019	5	8
2020	14	4
2021	46	7
2022	18	4
2023	31	2
2024	2	2
Sum	208	40

인명 및 재산 피해를 최소화하려는 국가적 노력의 일환으로 볼 수 있다.

연도별 뉴스 기사 수와 연구 논문 수의 상관관계를 Pearson 상관계수를 통해 분석한 결과, 상관계수는 0.482($p=0.027$)로 낮은 수준의 양의 상관관계를 보였다. 이는 연구 논문 출판과 뉴스 기사 보도 사이에 어느 정도 관련성이 있음을 나타낸다. 다만 뉴스 기사는 사회적 관심을 즉각적으로 반영하는 반면, 연구 논문은 연구 수행 및 출판에 시간이 소요된다. 따라서 이 두 변수 간의 상관성은 단순한 인과적 관계로 완전히 설명하기 어려우므로, 시간적 지연효과(time lag effect)를 고려한 시차적 상관관계 분석(lagged correlation analysis)을 통해 시간의 흐름에 따른 영향을 추가적으로 검증할 필요가 있다.

2. 토픽 모델링 적용 분석 결과

1) 연구자료별 최적 토픽 수(K) 결정 결과

토픽 모델링의 결과는 사용자에게 의해 설정되는 토픽 수(K)에 민감하게 영향을 받는다. K가 지나치게 작으면 하나의 토픽에 여러 개의 상이한 주제가 섞여서, 결과의 해석이 어렵고

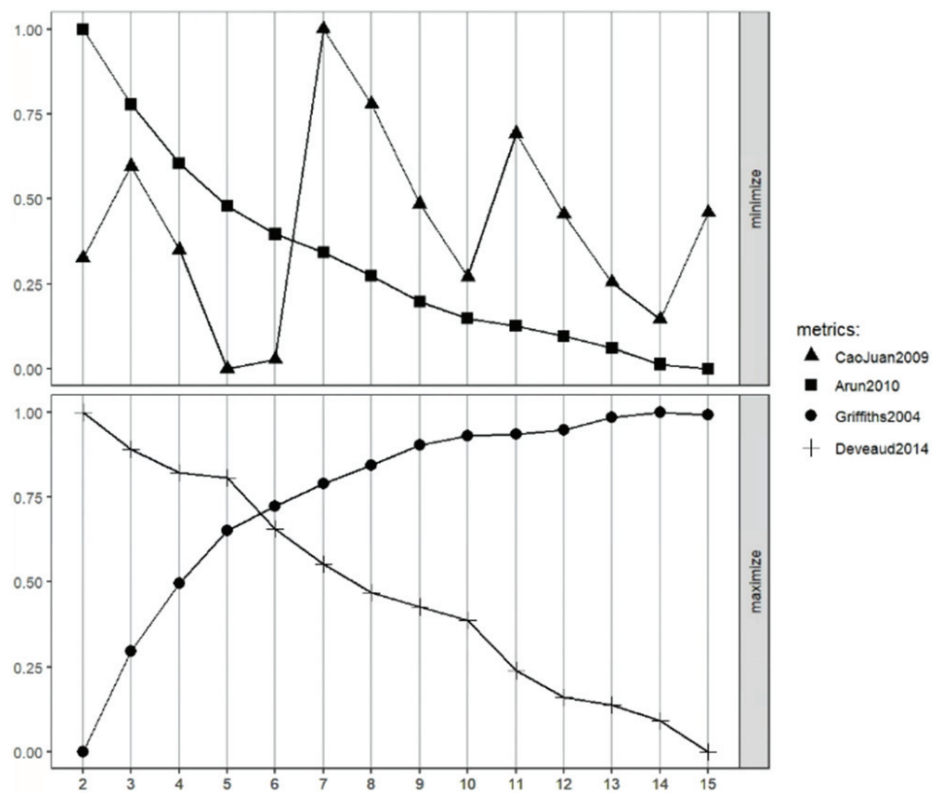


Figure 4. LDA tuning index according to the number of topics on news articles.

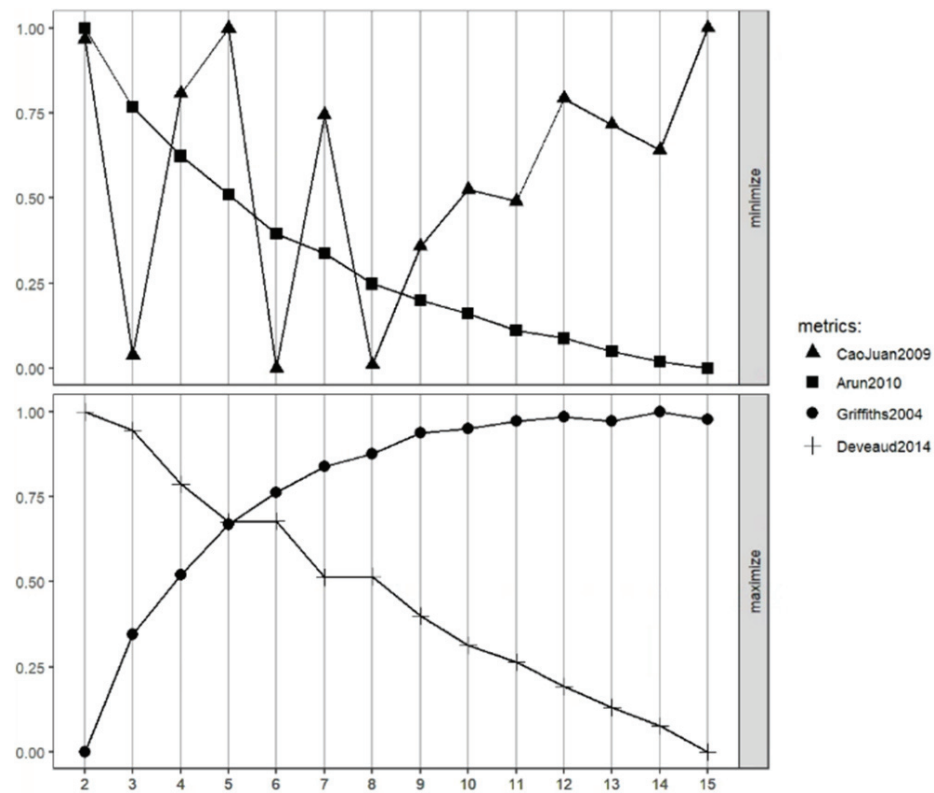


Figure 5. LDA tuning index according to the number of topics on research abstracts.

분석의 유용성이 떨어진다. 반대로 K가 너무 크면 유사한 주제들이 불필요하게 세분화되어, 결과의 간결성과 효율성이 저하된다. 따라서 K는 주제 간의 경계를 명확히 구분하면서도, 각 토픽이 충분히 해석 가능한 수준에서 결정되어야 한다. 최적의 K를 결정하기 위하여 LDA 모델에서 토픽 수(K)를 바꿔가며 모델 성능을 평가하기 위해 사용하는 지표를 LDA 튜닝 지표(LDA tuning index)라고 하며, 현재까지 다양한 방법의 튜닝 지표가 제안되어 왔다. 대표적으로 CaoJuan 지표(Cao et al., 2009), Arun 지표(Arun et al., 2010), Griffiths 지표(Griffiths and Steyvers, 2004), 그리고 Deveaud 지표(Deveaud et al., 2014)가 있다. 이 중 CaoJuan 지표와 Arun 지표는 그 값이 작을수록, 반면에 Griffiths 지표와 Deveaud 지표는 그 값이 커질수록 최적의 K를 찾게 된다. 본 연구는 R 4.3.2의 'ldatuning' 패키지를 활용하여 토픽 수의 범위를 2부터 15까지로 설정하고, 상위 4개의 평가 지표를 고려하여 최적의 토픽 수를 결정하고자 하였다(Nikita and Chaney, 2016).

네 가지 LDA 튜닝 지표를 종합적으로 검토한 결과, 뉴스 기사 데이터의 최적 토픽 수를 6개로 결정하였다(Figure 4). CaoJuan 지표 자체로만 보았을 때 5개 토픽에서 중복성과 혼잡도가 가장 낮게 나타났으나, Griffiths와 Deveaud 지표가 6개 부근에서 토픽의 설명력과 내부 응집력이 유의하게 증가하는 경향을 보였다. 또한, 뉴스 기사 수가 증가함에 따라 주제의 세분화가 필요하고 트렌드 파악의 정밀성이 중요하다는 도메인적 특성을 고려하여, 토픽 수를 한 단계 확장함으로써 토픽의 분해능(resolution)과 의미적 응집도를 높이는 방향을 우선시하였다. 연구 논문 데이터의 경우에는 CaoJuan 지표가 5개에서 국소적 최소값을 나타내며, Griffiths 지표 또한 이 지점에서 국소적 정점을 보이고 이후 토픽 수 증가에 따른 유의미한 설명력 향상이 관찰되지 않았다(Figure 5). 더욱이

Deveaud 지표 역시 토픽 수 5개에서 응집도의 급격한 증가가 확인되어, 토픽 수가 이보다 늘어날 경우 중복성과 혼잡도가 심화될 것으로 판단되었다. 이러한 정량적 지표와 정성적 해석을 종합하여, 연구 논문 데이터의 경우 최적 토픽 수를 5개로 결정하였다.

2) 뉴스 기사 분석 결과

뉴스 기사에 대한 토픽 모델링 분석의 결과, 최적 토픽 수 6개별로 분석된 주요 키워드와 전체 뉴스 기사 중 차지하는 비율을 Table 4에 나타내었다. 각 토픽에서 다루어지는 뉴스 기사 내용을 요약하여 정리하면 다음과 같다.

토픽 ①은 주로 기후변화와 집중호우로 인해 발생하는 산사태와 땅밀림에 대한 예방 및 대응 노력에 관한 내용을 다루고 있다. 산림청 및 여러 지자체는 여름철 집중호우에 대비하여 산림 재해 예방 사업을 추진하고 있음을 강조하고 있다. 또한, 산림청은 전국적으로 땅밀림 우려지역에 대한 정밀 조사를 진행하고, 피해 예방을 위한 다양한 대책을 마련하고 있다.

토픽 ②는 경주와 장수 지진 발생 후 산림청 및 지방 정부의 대응과 산림 및 재해 예방 노력을 다루고 있다. 경주 지진 후 산림청은 진앙지 반경 50 km 내 산림시설 및 다중이용시설을 점검한 결과, 피해가 없음을 확인했다. 또한, 장수 지진 발생에 따른 산사태 취약지 현장 점검을 실시하여 2차 피해를 예방하고 있다. 이 외에도 중부지방산림청은 해빙기 땅밀림 지역에 대한 국가안전대진단을 실시하여 복구 사업지를 특별 점검하고 있다.

토픽 ③은 한국의 여러 지역에서 발생한 땅밀림 현상과 그로 인한 문제들을 다루고 있다. 대구 신서동에서 땅밀림 위험 A등급 지역이 확인되어 긴급 복구 작업이 진행되고 있으며, 경남은 전국에서 가장 많은 땅밀림 우려지역이 분포되어 있어 위험이 높다. 울릉군에서는 LPG 가스 배관망 사업이 땅밀림

Table 4. The result of topic modeling on news articles.

No.	Topic	Keyword	Ratio
①	Prevention and response efforts	management, prevention, site, countermeasure, react, nationwide, calamity, danger, plan, forest, information, evacuation	29.3%
②	Landslide-prone area inspection and prevention	inspection, earthquake, site, safety, vulnerable, facilities, propulsion, expert, input, emergency, on this day, target	24.1%
③	Need for landcreep prevention measures	inspection, phenomenon, ground, people, Gyeongnam, result, designation, reality, vicinity, find, official, professor	18.8%
④	Need for landcreep area management	restoration, disaster, mountainous area, typhoon, heavy rain, concentration, influence, our country, landcreep area, geology, downpour, past	12.0%
⑤	Unmanned landslide monitoring system	installation, scale, damage, Mt. Halla, primary factor, surveillance, continuous, area, progress, service, sediment, measurement	9.1%
⑥	Suncheon city securing national funds	business, budget, Suncheon, secure, backup, reflection, member, furtherance, Suncheon City, congress, national expenses	6.7%

현상으로 인해 차질을 빚고 있다. 청남대와 김해 등에서도 산사태와 땅밀림 위험이 높아 정밀 조사와 예방 대책이 필요하다.

토픽 ④는 산림 관리의 중요성을 강조하고 있다. ‘경일포럼’ 시리즈는 국립공원 내 땅밀림지 관리, 여름철 재해 예방, 산사태와 땅밀림 위험지 관리, 태풍 대비, 강우 집중 시기 관리의 필요성 등 다양한 주제를 다룬다. 특히, 경상도 지역이 땅밀림과 산사태의 위험이 가장 높다고 경고하며, 위험지 관리를 철저히 할 것을 촉구한다. 또한, 재해 예방에 충분한 재정을 투자해야 한다는 의견이 제기되었으며, ‘태화산 산사태 예방대책’과 같은 구체적인 예방 대책 추진 사례도 소개된다.

토픽 ⑤는 한국에서 땅밀림 및 산사태 예방과 관련된 다양한 노력과 기술 도입을 다루고 있다. 산림청은 전국 25곳에 땅밀림 무인 원격감시시스템을 설치하여 선제 대응 체계를 구축하고 있다. 이 시스템은 땅밀림 위험 지역을 실시간으로 모니터링하여 조기 경보를 가능하게 한다(Seo et al., 2018; Kim et al., 2023). 부산지역에도 3곳에 설치되었으며, 이를 통해 땅밀림 피해를 예방하려는 노력이 강화되고 있다. 서울시는 도시 관리에 최첨단 항공 레이저 측량 기술을 도입하여

보다 정확한 지형 데이터를 확보하고 있다. 즉, 땅밀림과 산사태의 조기 발견 및 대응을 위한 기술적 발전을 다룬다.

토픽 ⑥은 순천시의 예산 관련 소식을 다루고 있다. 순천시 산림재해 예방과 관리를 위해 사회간접자본(SOC) 핵심 예산을 확보하였으며 이를 통해 지역 주민들의 생활 환경 개선에 기여할 것으로 기대한다.

연도별 뉴스 기사와 특정 토픽 간의 관련성을 분석한 결과는 Table 5이다. 토픽 ①은 2016~2017년과 2021년에 언급 빈도가 높아, 기후 변화와 산림재해 예방에 대한 관심이 집중되었음을 나타낸다. 토픽 ②는 2016년 이후 꾸준히 언급되며, 2022~2023년에 기사 수가 급증해 지진과 재해 예방 노력이 주목받았다. 토픽 ③은 2018년 경주시 양북면의 땅밀림 발생과 2021년 산림청의 땅밀림 예방·대응정책 발표에 따라 많은 사회적 관심을 끌었다. 토픽 ④는 2013년에 큰 관심을 받았고, 이후 꾸준히 산림 관리와 재해 예방 중요성이 논의되었다. 토픽 ⑤는 2018년에 산림청에서 전국 25개소에 땅밀림 무인원격감시시스템을 설치한 것을 계기로 땅밀림 감시 기술과 관련된 논의가 활발하였다. 토픽 ⑥은 순천시의 예산 확보 및 산림재해 관리가 주요 이슈였던 2020년에 관련 기사 수가 급증했다. 이처럼 토픽별 뉴스 기사 건수는 시기별 사회적 이슈와 관련 대응 정책에 의해 결정되는 것으로 생각되었다.

Table 5. The number of news articles related to topics by year.

Topic	①	②	③	④	⑤	⑥	Total
2004	0	0	0	0	1	0	1
2005	0	0	0	0	1	0	1
2006	0	0	0	1	0	0	1
2007	0	0	0	0	0	0	0
2008	0	0	0	0	0	0	0
2009	0	0	0	0	1	0	1
2010	0	0	0	0	0	0	0
2011	0	0	0	1	0	0	1
2012	0	0	0	0	0	0	0
2013	0	0	3	8	0	0	11
2014	1	0	3	0	0	0	4
2015	0	0	0	1	1	0	2
2016	13	5	3	0	1	0	22
2017	13	4	3	1	0	1	22
2018	1	7	5	1	12	0	26
2019	2	2	0	1	0	0	5
2020	1	0	2	1	1	9	14
2021	21	5	13	3	0	4	46
2022	4	10	2	1	1	0	18
2023	5	17	4	5	0	0	31
2024	0	0	1	1	0	0	2
Sum	61	50	39	25	19	14	208
Ratio	29.3%	24.1%	18.8%	12.0%	9.1%	6.7%	100%

3) 연구 논문 분석 결과

연구 논문에 대한 토픽 모델링 분석의 결과, 최적 토픽 수 5개별로 분석된 주요 키워드와 전체 연구 논문 중 차지하는 비율을 Table 6에 나타내었다. 각 토픽에서 다루어지는 연구의 내용을 정리하면 다음과 같다.

우선 토픽 ①은 전기비저항을 활용한 땅밀림 지역의 특성 분석과 관련된 연구이다. 주요 연구로서 전기비저항을 이용하여 풍화층 및 땅밀림 모델의 깊이분해능과 감도분포를 분석하고, 이를 통하여 땅밀림이 발생할 수 있는 지역의 지반 특성을 파악하려는 시도가 있었다. 또한, 점토광물 함유량을 고려한 전기비저항 자료의 해석을 통하여 땅밀림 위험이 높은 지역을 식별하고, 복합 지구물리탐사 기법을 통해 활동연약대를 규명하고자 하는 연구도 있었다. 이러한 연구는 산지 지반의 안정성 분석과 땅밀림 예방을 위한 중요한 정보를 제공하며, 대규모 산지토사재해 예방을 위한 실용적인 방법론을 제시하고자 한 것이다.

토픽 ②는 땅밀림의 발생과 복구·복원과 관련된 연구로서, 주로 땅밀림이 발생한 비탈면 및 산림지의 환경적 특성 변화에 주목하고 있다. 주요 연구로서 땅밀림지에서 지하수와 강우가 미치는 영향, 복구사업 후 유기물층과 토양층의 이화학적 특성 변화를 다룬 것이 있었다. 또한, 산림환경 특성을 분석하여 땅밀림 발생을 예측하고, 복원지의 환경적 변

화가 땅밀림 예방에 미치는 효과를 규명하고자 한 연구도 있었다. 일부 연구에서는 땅밀림 발생원인 및 대책 구조물의 안정성을 평가하는 내용도 포함되어 있어서, 땅밀림이 발생한 지역의 복구·복원과 관련된 실용적인 지침을 제공하는 데 중요한 기여를 할 수 있을 것으로 생각되었다.

토픽 ③은 땅밀림 모니터링 및 위험 평가와 관련된 최신 기술을 활용한 연구이다. 이 연구는 주로 지상 라이더(terrestrial LiDAR), 드론, 무인 항공사진측량, 기계학습(machine learning) 등의 기술을 이용하여 땅밀림 위험지 및 피해 범위를 정확하게 조사하고 분류하는 방법을 다룬다. 기계학습 기법을 활용한 위험 등급 분류와 수치해석을 통한 땅밀림 위험도 평가, 항공 라이더(airborne LiDAR)를 통한 도로 인접 사면의 변위 특성 분석 등을 포함한 연구는 땅밀림 발생 가능성을 예측하고 효과적으로 모니터링할 수 있는 방법들을 제시하고 있다. 또한, 북한의 산지토사재해 발생 경향과 복구 사례를 분석하여 땅밀림 대응 방안을 제시하는 연구도 포함되어 있다. 이러한 연구는 땅밀림 재해를 예방하고 관리하는 데 중요한 기술적 접근을 제공할 수 있다.

토픽 ④는 땅밀림지의 토양 특성, 균열, 인장력, 진행 방향 및 지질적 요소와 관련된 연구이다. 이 연구들은 주로 땅밀림이 발생한 지역의 토양 물리적 특성, 균열 형상, 암석 및 토양의 구성 등을 분석하여 그 발생 원인과 특성을 파악하고자 하였다. 예를 들어, 울주, 양평, 양산, 포항, 진주 등 다양한 지역의 땅밀림지에서 토양 물리성 및 지질 변화, 인장균열 분석 등을 통해 땅밀림을 일으키는 주요 요인들을 규명하였다. 또한, 물리탐사와 실험적 연구를 통해 땅밀림의 메커니즘을 규명하려는 시도가 이루어지고 있다. 이러한 연구는 땅밀림 예측 및 예방에 중요한 역할을 하며, 지역적 특성에 맞는 대응 방안을 제시하는 데 기여할 수 있다.

토픽 ⑤는 땅밀림지의 지형적 및 지질적 특성을 분석한 연구를 포함한다. 주로 땅밀림지의 규모와 발생 원인, 지질 및

토양 특성, 산사태 위험도 분석을 다루고 있다. 구체적으로는 GIS와 항공사진, 무인 항공 사진측량을 활용하여 땅밀림지의 특성과 영향을 평가하고, 지형적 요소(경사도 등)와 지질적 요소(지질층위 구분 등)가 땅밀림 발생에 미치는 영향을 분석하였다. 이러한 분석을 통해 궁극적으로는 땅밀림을 예측하고 피해를 저감하는데 중요한 정보를 제공할 수 있다.

연도별 토픽별 연구 논문 편수로 땅밀림 연구의 시간적 흐름을 파악한 결과는 Table 7이다. 토픽 ①, ②, ③은 2018년 이후 지속적으로 관심을 받고 있으며, 특히 토픽 ①은 최근 5년간 가장 많은 연구 논문을 차지하고 있다. 이는 땅밀림의

Table 7. The number of research abstracts related to topics by year.

Topic	①	②	③	④	⑤	Total
2004	0	0	1	0	0	1
2005	0	0	1	0	2	3
2006	0	0	0	1	0	1
2007	0	0	0	0	0	0
2014	0	0	0	0	0	0
2015	1	0	0	1	1	3
2016	2	0	0	0	0	2
2017	0	0	0	0	0	0
2018	0	1	1	1	0	3
2019	2	1	1	2	2	8
2020	1	0	1	2	0	4
2021	1	5	1	0	0	7
2022	3	0	1	0	0	4
2023	0	1	1	0	0	2
2024	2	0	0	0	0	2
Sum	12	8	8	7	5	40
Ratio	30.0%	20.0%	20.0%	17.5%	12.5%	100%

Table 6. The result of topic modeling on research abstracts

No.	Topic	Keyword	Ratio
①	Electrical resistivity based analysis	resistivity, electricity, grasp, distribution, clay, commitment, data, analysis, bottom, exploration, ground, stability, check	30.0%
②	Occurrence and restoration	underground, landslide, slope, rainfall, judgment, top, influence, restore, adjacency, damage, forest land, steep slope	20.0%
③	Landcreep monitoring and risk assessment	sediment, utilization, disaster, mountainous area, conjugation, monitoring, damage area, inspection, classification, accuracy, aviation, apply, possible	20.0%
④	Soil and fracture analysis and failure characteristics	soil, crack, characteristic, seal, direction, geology, composition, breakdown, landcreep, into the landcreep, rock, nature	17.5%
⑤	Geomorphological and geological analysis	topography, landcreep area, region, change, practice, rating, microscopic, our country, cause, eroded soil, offer, contour line	12.5%

주요한 발생 원인으로서 지질학적 특성 규명을 위해서, 직접적인 시추 조사를 하지 않고도 대면적의 지질특성을 상대적으로 간편하게 분석할 수 있는 전기비저항 탐사 방법이 많이 활용되고 있는 최근 학문적 경향이 반영되고 있기 때문이다. 이렇듯 땅밀림의 발생 원인 규명 및 적합한 대책 마련을 위해서 새롭게 개발되는 다양한 방법론을 적용·시도하는 분야가 새로운 연구 주제가 대두되고 있음을 의미한다.

3. 의미망 분석 결과

1) 뉴스 기사 분석 결과

Figure 6은 뉴스 기사를 대상으로 N-gram(N=2) 기반의 의미망 분석을 수행하여, 주요 키워드들의 동시 출현 및 네트워크 상의 관계를 시각화한 것이다. Table 1에서 제시된 바와 같이, ‘area’(지역), ‘landslide’(산사태), ‘occurrence’(발생), ‘landcreep’(땅밀림) 등은 단어 빈도와 비율 측면에서 높은 비중을 차지하는 주요 키워드였다. 그러나 의미망 분석은 단순히 단어의 출현 빈도로는 드러나지 않는 키워드 간 연계성과 맥락적 중요성을 강조한다. 예를 들어, Table 1에서 높은 빈도로 나타난 단어들 중 하나인 ‘landslide’는 의미망에서 네트워크의 중심에 위치하고, ‘heavy rain’(집중호우), ‘vulnerable areas’(취약 지역), ‘cause of damage’(피해 원인)와 같은 키워드와 밀접하게 연결되며 주제의 핵심을 형성한다. 이는 단어 자체의 빈도뿐만 아니라, 두 단어가 함께 출현하는 관계의 구조적 중요성이 높다는 점을 시사한다. 또한, N-gram 의미망에는 ‘unmanned remote monitoring system’(무인원격 감시시스템)과 같이 Table 1의 단순 빈도로는 비중이 크지 않았

던 키워드가 주요 노드로 부각되고 있다. 이는 N-gram 네트워크 분석을 통해 땅밀림의 예방·대응을 위한 기술적 수단이 뉴스 담론에서 의미 있게 연결되고 있음을 보여준다. 한편, ‘management’(관리)와 ‘inspection’(점검) 같은 단어는 Table 1에서도 높은 빈도로 나타나며, 의미망에서도 중요한 연결망을 구성해 정책적·실무적 대응 방안의 중요성이 재확인되었다. 또한, ‘KFS’(Korea Forest Service, 산림청), ‘Hadong-gun’(하동군), ‘Jeonbuk’(전북), ‘Gyeongnam’(경남) 등 기관명과 지명이 주요 노드로 포함되어 지역적 맥락과 정부 기관의 역할이 동시 출현을 통해 강조되고 있음을 알 수 있다. 마지막으로, Table 1에서 높은 비율을 보인 ‘landslide’(산사태) 역시 땅밀림과의 유기적 연결을 통해 관련 자연재해들 간 비교·분석의 맥락을 하동군, 전북, 경남과 같은 지역을 통한 보도로 다루고 있다. 이처럼 의미망 분석은 Table 1에서 제공된 단어 빈도의 한계를 보완하여, 뉴스 기사 전반에서 주요 키워드들의 구조적 연관성과 담론의 주요 초점을 보다 심층적으로 파악할 수 있는 유용한 수단임을 확인할 수 있다.

2) 연구 논문 분석 결과

Figure 7은 연구 논문을 대상으로 의미망 분석을 수행한 결과이다. Table 2에서 제시된 바와 같이, 연구 논문 데이터에서 ‘landcreep’(땅밀림), ‘research’(연구), ‘generation’(발생), ‘analysis’(분석), ‘result’(결과), ‘characteristic’(특성) 등이 높은 빈도로 나타났다. 의미망 분석 결과에서도 이러한 단어들이 단순 출현 빈도를 넘어 다양한 키워드와 구조적으로 연결되어 네트워크의 중심을 이루고 있음을 확인할 수 있다. 특히, Table 2에서 가장 높은 비율을 보인 ‘landcreep’(땅밀

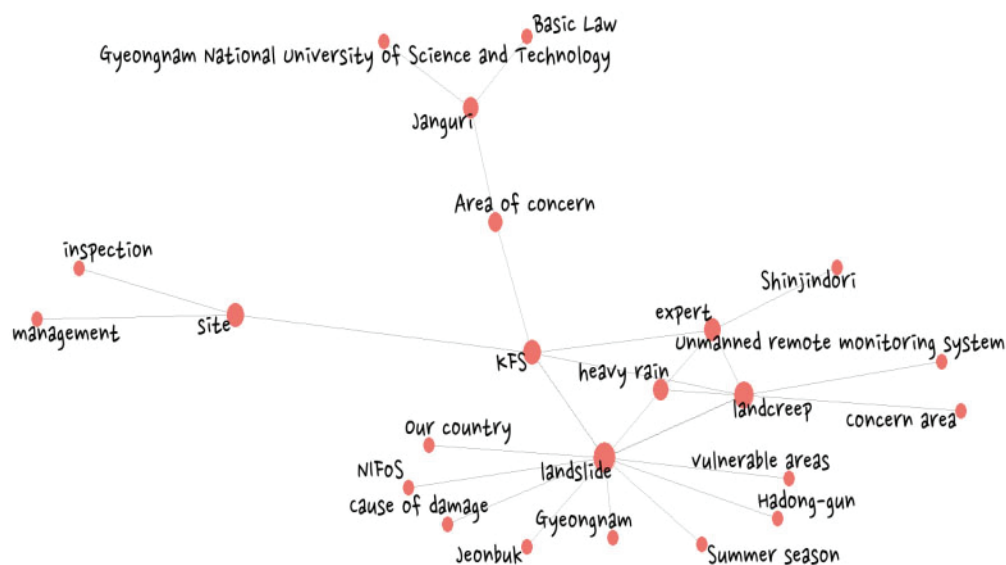


Figure 6. The network graph of news article.

리하고 있다는 점을 언급하며, 연구 논문에서도 땅밀림을 예측하고 모니터링하기 위한 첨단 기술을 활용한 시스템 구축 사례가 소개된다. 이러한 시스템들은 땅밀림의 발생 가능성을 실시간으로 감지하고 즉각적인 대응을 가능하게 하여, 땅밀림 피해를 최소화하는 중요한 역할을 한다.

차이점은 대상과 접근법에서 드러난다. 뉴스 기사는 일반 대중을 대상으로 하여 정책적 대응과 구체적인 사례를 보도하며, 보다 실질적이고 현장적인 정보를 제공한다. 반면, 연구 논문은 학술적 독자를 대상으로 하여 기술적이고 이론적인 분석을 통해 땅밀림 발생의 원인과 메커니즘을 설명하고, 예측 및 예방을 위한 첨단 기술적 방법을 제시한다. 또한, 뉴스 기사에서는 정책적 대응과 지역 사회의 협력을 강조하는 반면, 연구 논문은 기술적 기법을 통한 보다 정확한 예측과 조기 경보 시스템을 다루고 있다.

종합적으로 2010년대 중반 땅밀림에 대한 사회적 관심이 증대되어 정부의 대응·관리 대책 중심의 뉴스 기사의 양이 증가하였고, 이와 함께 땅밀림 현상의 원인 규명과 예방 대책 마련을 위한 연구 논문 역시 많이 출판되고 있음을 확인할 수 있었다. 특히, 재난관리의 4단계 즉, ‘예방-대비-대응-복구’의 단계 중에서 주로 예방 대책에 초점을 맞춘 정부 정책과 관련 기술 연구가 땅밀림 분야에서 진행되어 왔다는 점을 파악할 수 있었다. 이를 통하여 과거의 복구 중심에서 최근의 예방 중심으로의 전환이 핵심인 재난관리 대책의 패러다임 변화가 땅밀림 분야의 정책 및 연구에도 충분히 반영되어 있음을 확인할 수 있었다.

한편, 땅밀림 관련 미래의 정책 및 연구의 방향성을 설정하는데는 산사태 분야의 사례를 참고할 필요성이 있다. 우리나라에서는 2011년 서울 우면산 산사태 발생 이후, 도시생활권 중심의 인명 피해 저감을 목적으로 산사태위험지도와 KLES(Korea Landslide Early-warning System) 등이 개발·개선되었고, 이를 통하여 전국 단위의 체계화된 산사태 예방 시스템을 갖추게 되었다. 이와 마찬가지로 땅밀림 분야에서도 현재의 예방 대책 중심의 정책 및 연구 경향을 계승하여, 전국 단위의 위험지도(발생 가능성 평가지도) 구축이나 발생 예측정보 시스템 개발 등 전국 단위 땅밀림 예방 시스템 구축 관련 연구가 필요할 것으로 판단된다. 하지만, 우리나라에서 땅밀림은 산사태와는 활동 특성이 상이하고 아직까지는 발생 원인 및 특성이 충분히 규명되지 않아서, 산사태 예방 시스템과 동일한 방법론을 적용하여 땅밀림 예방 시스템을 구축하는 것은 바람직하지 않을 것으로 판단된다. 그러므로, 국내 땅밀림 발생지 혹은 발생 우려지역에 대한 현장조사 및 모니터링 등 관련 기초연구가 충분히 선행된 후, 이를 바탕으로 국내 산지환경 및 땅밀림 발생특성에 적합한 예방 시스템이 구축되어야 할 것이다. 전국 단위의 땅밀림 현장조사 및 모니

터링에 딥러닝 기반 물체 탐지 기술인 YOLO(Redmon et al., 2016; Zhou et al., 2017) 등 최신 기술이 적극적으로 도입된다면 더 정확도가 높은 예방 시스템 구축에 기여할 수 있을 것으로 기대된다.

결론

이 연구에서는 땅밀림 관련 뉴스 기사와 연구 논문의 동향을 텍스트 마이닝 기법인 토픽 모델링과 의미망 분석을 통해 분석하였다. 이를 통해 땅밀림과 관련된 주요 이슈, 연구 분야, 그리고 사회적 관심사를 체계적으로 도출할 수 있었다. 땅밀림 현상의 발생에 따른 사회적 관심은 뉴스 기사와 연구 논문의 양적 증가에 영향을 미쳤고, 땅밀림 대응·관리 정책 및 연구의 방향이 재난 관리적 측면에서 ‘예방’ 대책 중심인 것을 확인할 수 있었다. 이러한 분석 결과를 바탕으로 향후 땅밀림 관련 정책 및 연구의 방향성을 전국 단위 예방 시스템 개발·구축하는 것으로 제안하였다.

한편, 연구 논문의 분석 측면에서는 국내 산지토사재해의 대표적 유형인 산사태에 비해 땅밀림은 발생 사례가 희귀하여 관련 연구 논문의 수도 상대적으로 적어서, 다양한 관점에서 충분한 분석을 수행하는데 다소 한계가 있었다. 이는 해당 분야의 연구가 아직 초기 단계에 있음을 보여주며, 향후 지속적인 연구와 데이터 축적의 필요성을 내포하고 있다.

이처럼 토픽 모델링과 의미망 분석 등의 텍스트 마이닝 기법은 땅밀림과 같은 새롭게 등장하는 연구분야(혹은 주제)에 대한 현황을 파악하고 향후 방향성을 설정함에 있어서 유용하게 활용할 수 있는 것으로 판단되었다. 특히, 산지토사재해 분야에서는 연구 논문 뿐만 아니라 뉴스 기사를 함께 분석함으로써, 방재 대책의 활용성 혹은 효용성을 검토할 수 있는 계기가 된 것으로 생각되었다. 추후에도 해당 방법론이 산림 재해 분야 혹은 전체 산림 분야에서도 적극적으로 활용되어, 체계적인 연구 동향 분석 및 그에 기반한 합리적인 미래 연구 방향성 설정에 실질적인 도움이 되기를 기대해본다.

References

- Arun, R., Suresh, V., Veni Madhavan, C.E. and Narasimha Murthy, M.N. 2010. On finding the natural number of topics with latent dirichlet allocation: Some observations. In *Advances in Knowledge Discovery and Data Mining: 14th Pacific-Asia Conference, PAKDD 2010, Proceedings. Part I* 14: 391-402. Springer, Berlin Heidelberg.
- Aggarwal, C.C. and Zhai, C. 2012. *Mining Text Data*. Springer, New York
- Blei, D.M., Ng, A.Y. and Lafferty, J.D. 2003. *Latent dirichlet*

- allocation. *Journal of Machine Learning Research* 3: 993-1022.
- Blei, D.M. 2012. Probabilistic topic models. *Communications of the ACM* 55(4): 77-84.
- Borgatti, S.P., Everett, M.G. and Johnson, J. 2009. *Analyzing social networks*. SAGE Publications.
- Cao, J., Xia, T., Li, J., Zhang, Y. and Tang, S. 2009. A density-based method for adaptive LDA model selection. *Neurocomputing* 72(7-9): 1775-1781.
- Csardi, G. and Nepusz, T. 2006. The igraph software. *Complex Systems* 1695: 1-9.
- Deveaud, R., SanJuan, E. and Bellot, P. 2014. Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique* 17(1): 61-84.
- Griffiths, T.L. and Steyvers, M. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences* 101(suppl. 1): 5228-5235.
- Grün, B. and Hornik, K. 2011. Topicmodels: An R package for fitting topic models. *Journal of Statistical Software* 40: 1-30.
- He, W., Zha, S. and Li, L. 2017. Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management* 37(1): 1-11.
- Hofmann, T. 2019. Probabilistic latent semantic indexing. In : *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 50-57. <https://doi.org/10.1145/312624.312649> (2025. 6. 27.)
- Kim, J., Kim, J., Lee, C., Park, S., Lee, C. and Jeong, G. 2023. Monitoring landcreep using terrestrial LiDAR and UAVs. *Journal of the Korean Society of Engineering Geology* 33(1): 27-37.
- Kim, D., Oh, C. and Zhu, Y. 2021. Analyzing research trends in blockchain studies in South Korea using dynamic topic modeling and network analysis. *Journal of the Korean Society for Information Management* 38(3): 23-39.
- Landauer, T.K., Foltz, P.W. and Laham, D. 1998. An introduction to latent semantic analysis. *Discourse Processes* 25(2): 259-284.
- Lee, J. and Lee, C. 2021. A study on domestic research trends (2001–2020) of forest ecology using text mining. *Journal of Korean Society of Forest Science* 110(3): 308-321.
- Lee, K. and Kim, D. 2024. Analysis of research trends in the cadastrals field using topic modeling. *Journal of the Korean Society of Cadastre* 40(2): 71-82.
- Lee, S., Seo, J., Kim, J., Ryu, D., Seo, J., Kim, D. and Lee, C. 2017. International research trend on mountainous sediment-related disasters induced by earthquakes. *Journal of Korean Society of Forest Science* 106(4): 431-440.
- Lee, W., Park, C., Gang, S. and Yoon, S. 2024. Analysis of trends in train accident-related news using topic modeling. *Journal of The Korean Data Analysis Society* 26(6): 1739-1750.
- Lim, J., Kim, K., Kim, M., Yi, J.M. and Park, J.W. 2020. Trend analysis of north Korean forest science research (1962-2016) by data mining. *Journal of Korean Society of Forest Science* 109(1): 81-98.
- Lim, H., Kim, D. and Yoon, S. 2023. A news-based topic modeling of wildfire over suppression time. *Journal of the Korean Data And Information Science Society* 34(2): 245-254.
- Manning, C.D., Raghavan, P. and Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press.
- Ministry of Culture, Sports and Tourism. 2018. The Pohang earthquake and landcreep: A call to halt the movement. Policy Briefing of the Republic of Korea Website. <https://www.korea.kr/briefing/pressReleaseView.do?newSId=156273530> (2025. 6. 27.) (in Korean)
- Moon, S., Chung, S. and Chi, S. 2018. Topic modeling of news article about international construction market using Latent Dirichlet Allocation. *Journal of the Korean Society of Civil Engineers* 38(4): 595-599.
- Nikita, M. and Chaney, N. 2016. ldatuning: Tuning of the latent dirichlet allocation models parameters. R package version 0.2-0, URL <https://CRAN.R-project.org/package=ldatuning>.
- Park, J., Choi, K., Lee, S., Ma, H., Lee, J. and Woo, B. 2003. Analysis on the characteristics of the landslide in Nasamri (I) - With a special reference on geo-topographical characteristics. *Journal of Korean Forest Society* 92(3): 246-253.
- Park, J., Choi, K., Bae, J.S., Ma, H. and Lee, J. 2005. Analysis on the characteristics of the landslide in Maeri (I) - With a special reference on geo-topographical characteristics. *Journal of Korean Forest Society* 94(3): 129-134.
- Park, J. and Park, S. 2019. The geology and variations of soil properties on the slow—moving landslide in Yangbuk-myun, Gyungju-si, Gyeongsangbuk-do. *Journal of Korean Society of Forest Science* 108(2): 216-223.
- Park, J. and Song, M. 2013. A study on the research trends in library and information science in Korea using topic modeling. *Journal of the Korean Society for Information Management* 30(1): 7-32.

- Park, Y. and Yoon, S. 2024. Analysis of recent research trends in the journal of intellectual disabilities research through topic modeling. *Journal of Intellectual Disabilities* 26(4): 61-87.
- Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. 2016. You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*: 779-788.
- Shannon, C.E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27(3): 379-423.
- Sowa, J.F. 1984. *Conceptual structures: information processing in mind and machine*. Addison-Wesley.
- Seo, J., Lee, C., Woo, C. and Yu, S. 2021. The case study of slow-moving landslide rehabilitation and guideline of rehabilitation with landslide type. *The National Institute of Forest Science Research Materials No. 970*. pp. 143.
- Seo, J., Woo, C., Lee, C. and Kim, D. 2018. A study on detection and monitoring in land creeping area by using the UAV. *Journal of Korea Academia-Industrial Cooperation Society* 19(11): 481-487.
- Song, I., Lee, G. and Song, C. 2022. Analysis of recent trends in research on the effects of forest healing programs on humans in Korea. *Journal of Korean Society of Forest Science* 111(4): 651-666.
- Woo, B., Park, J., Choi, H., Jeon, G. and Kim, K. 1996. A study on the characteristics of the landslide in Hyuseok-dong (I) - Topographical characteristics and surface displacement. *Journal of Korean Forest Society* 85(4): 565-570.
- Yoon, S. and Kim, Y. 2023. Topic modeling for cosmetic surgery-related youtube content. *Journal of the Korean Data And Information Science Society* 34(6): 865-874.
- Zhou, X., Wang, D. and Xue, X. 2017. Object detection in aerial images: A review. *Remote Sensing* 9(12): 1271.

Manuscript Received : February 3, 2025

First Revision : July 4, 2025

Second Revision : September 8, 2025

Accepted : September 8, 2025