

미디어 텍스트 분석 기반의 공급망 리스크 모니터링 시스템의 개발*

최동업(제1저자)

중앙대학교 대학원 경영학과 석사과정
(dongyc1231@cau.ac.kr)

서용원(교신저자)

중앙대학교 경영경제대학 경영학부 교수
(seoyw@cau.ac.kr)

오늘날의 기업들은 무역 갈등의 심화, 전염병, 경기 침체, 전쟁 및 각종 자연재해 등으로 인한 공급망 리스크에 노출되어 있다. 이러한 배경에서 공급망 리스크와 관련된 정보를 수집하고 동향을 파악하는 것의 중요성이 증대되고 있으며, 뉴스 기사와 같이 실시간으로 발생하는 미디어 텍스트를 분석하는 것은 공급망 리스크와 관련된 최신 정보를 빠르게 수집하는 유용한 방법으로 대두되고 있다. 하지만, 공급망 리스크와 관련된 텍스트를 분석하는 연구는 초기 단계이며, 최근 활용도가 증가하는 딥러닝 기반의 자연어 처리 기법을 적용한 텍스트 분석은 미비한 상황이다. 이에 본 연구에서는 뉴스 기사 분석을 활용하여 공급망 리스크와 관련된 정보를 수집, 도출하는 인공지능 기반의 공급망 리스크 모니터링 시스템을 개발한다. 이를 위해 사전 학습 언어모델인 KoBERT에 기반해 공급망 리스크 관련 기사만을 수집하는 필터링 모델을 수립하고, 수집된 기사의 공급망 리스크 유형을 LDA 토픽 모델링 기반으로 식별하여 학습 데이터를 구축하였다. 이후, BOW(Bag of Words)와 KoBERT를 사용한 딥러닝 기반의 공급망 리스크 분류 모델을 개발하여 수집된 기사의 공급망 리스크 유형을 예측하였다. 분석 결과, KoBERT 기반의 공급망 리스크 관련 기사의 필터링 정확도가 92.2%의 높은 성능을 보이는 것으로 나타났으며, 리스크 유형 분류 모델에서도 KoBERT 기반의 공급망 리스크 유형 분류 모델이 BOW 기반 모델에 비해 높은 분류 성능을 보이는 것을 확인하였다.

주제어: 공급망 리스크, 텍스트 마이닝, 언어모델, KoBERT, LDA 토픽모델링

1. 서론

오늘날의 기업들은 무역 갈등의 심화, 전염병, 경기 침체, 러시아-우크라이나 전쟁과 같은 국지전과 각종 자연재해 등으로 인한 다양한 공급망 리스크에 노출되어 있다. 국내 수출 기업의 대부분이 물류난, 원자재 가격 상승, 물품 수급 차질과 같은 공급망 문제를 경험한 바가 있는 것으로 파악되었다(한국무역

협회, 2022). 불확실성이 심화하는 환경 속에서 기업가치의 극대화를 위해 공급망 리스크를 식별하고, 관리하는 것은 필수적인 상황이다(최병일 등, 2023). 실제로, 기업들은 정부 차원의 공급망 조기경보 시스템 운영을 통한 선제 대응의 강화에 많은 관심을 보이고 있고, 공급망 이슈의 빠른 전달과 국내외 시장 정보의 적시 공유에 대한 높은 수요를 보이는 것으로 나타났다(한국무역협회, 2022).

때문에, 공급망 리스크와 관련된 정보 수집과 동향

논문투고일: 2023. 8. 31 논문수정일: 2023. 9. 24 게재확정일: 2023. 10. 4

* 이 논문은 2021년도 중앙대학교 연구장학기금 지원에 의한 것임.



© 한국생산관리학회. 이 저작물은 Creative Commons 저작자표시-비영리 4.0 국제 라이선스에 따라 이용할 수 있습니다. (<https://creativecommons.org/licenses/by-nc/4.0/deed.ko>)

파악에 대한 중요성과 관심이 증대되고 있다. 공급망 리스크를 파악하는 것은 공급망 리스크 관리의 첫 단계로서 대응 전략 수립에 역할을 할 수 있기 때문이다. 정부 차원에서도 산업 경쟁력을 위해 공급망 리스크를 관리하는 것의 필요성이 대두되고 있다(홍성태와 김창희, 2023). 이러한 배경에서 대한민국 정부는 공급망 리스크 동향을 파악하고, 공급망 관련 이상징후가 발생하면 이를 신속히 전파하기 위한 글로벌 공급망 분석센터의 출범을 발표한 바 있다(산업통상자원부, 2022).

이러한 배경에서 텍스트 마이닝을 활용해 뉴스 기사, 트위터와 같이 미디어에서 실시간으로 발생하는 텍스트에서 공급망 리스크와 관련된 최신 정보를 수집하고, 이를 공급망 리스크 관리에 활용하는 연구가 수행되고 있다(Chu 등, 2020; Ganesh와 Kalpana, 2022). 하지만, 공급망 리스크 관리 분야에서의 텍스트 분석 연구는 충분히 이루어지지 않은 상황으로, 더 많은 연구가 필요한 상황이다.

또한, 기존 연구에서는 주로 키워드의 등장 여부에 기반하여 분석을 수행하였으며, 최근 주목받는 BERT, GPT 등의 대규모 텍스트가 학습된 언어모델을 활용한 텍스트 분석 기법의 적용은 충분히 이루어지지 않은 상황이다. 기존 연구에서 사용한 방식은 단순 오타 혹은 키워드의 누락 등으로 인해 잘못된 분석되는 텍스트가 발생하는 한계가 존재한다(Su와 Chen, 2018). 언어모델에서는 단어의 등장 여부뿐만 아니라 앞 단어 혹은 앞, 뒤 단어를 동시에 고려하기 때문에 이러한 문제를 개선할 수 있다.

따라서, 본 연구는 미디어에서 등장하는 인공지능경망과 언어모델 KoBERT를 활용하여 고도화된 공급망 리스크와 관련된 뉴스 기사를 수집하고, 수집된 기사의 공급망 리스크 유형을 실시간으로 식별할 수 있는 모니터링 시스템을 개발하고 제안하고자 한다.

이를 위해 우선, 공급망 리스크 관련 검색 키워드로 뉴스를 수집한다. 두 번째로, 수집된 문서 기반의 학습 데이터를 구축하여 KoBERT 기반의 공급망 리스크 관련 기사 필터링 모델을 개발한다. 세 번째로, 수집된 기사의 공급망 리스크 유형을 LDA 토픽 모델링 기반으로 식별하여 학습 데이터로 구축한다. 이에 기반하여 BOW(Bag of Words)와 KoBERT를 사용한 딥러닝 기반의 공급망 리스크 분류 모델을 개발하여 기사의 공급망 리스크 유형을 예측하였으며, 두 모델의 분류 성능을 비교한다.

II. 관련 연구

2.1 미디어 텍스트 분석을 통한 공급망 리스크 모니터링

미디어에서 나타나는 공급망 리스크 모니터링 관련된 연구는 뉴스 기사와 트윗을 수집하는 방법을 제안하는 방식과 수집된 텍스트 데이터를 통해 공급망 리스크와 관련된 정보를 제공하는 방식으로 주로 수행되었다. Ganesh와 Kalpana(2022)는 공급망 리스크 관련 트윗을 실시간으로 모니터링하는 시스템을 제안하였고, 수집된 데이터의 단어 빈도 분석과 연관 규칙 마이닝 분석을 통해 주요 공급망 리스크 요인과 관련된 키워드를 제시하였다. Aboutorab 등(2022)은 공급망 리스크 관련 기사를 수집하고, 강화학습을 통한 기사 점수 지표 개발로 공급망 담당자에게 더 도움이 될 수 있는 기사를 전달하는 방법을 제시하였다.

공급망 리스크와 관련된 텍스트 데이터를 수집하고, 이를 분석하여 공급망 리스크의 패턴을 파악하

거나, 리스크를 평가하는 연구도 진행되어왔다. Su와 Chen(2018)은 공급망 리스크 관련 트윗을 수집하여 공급자 선정 과정에서 공급자가 노출되어 있는 리스크에 관한 정보를 제공하는 연구를 수행하였다. Chu 등(2020)은 글로벌 공급망 리스크 관련 뉴스 기사를 수집하여 공급망 리스크의 변화 패턴을 평가하기 위해 감성 분석을 수행하였고, 특정 기업과 관련된 감성 분석 패턴과 해당 기간의 기업 주가의 비교를 수행하였다. Murakami(2022)은 금속 공급망의 리스크를 평가하기 위해 주요 금속들과 관련된 기사를 수집하여 감성 분석을 수행하였고, 이를 통해 금속별 감성 분석 점수를 비교하고, 이슈를 분석하였다.

2.2 공급망 리스크 유형 분석 관련 연구

공급망 리스크의 유형을 분석한 연구는 공급망 리스크 관리 분야를 중심으로 한 연구와 공급망 리스크 관련 텍스트 분석 연구로 나누어 살펴보았다.

공급망 리스크 관리 분야의 선행연구에서 리스크 유형의 분석은 공급망 내 리스크 식별을 위하여 주로 수행되었다. Sodhi와 Chopra(2004), Tang과

Tomlin(2008), Tummala와 Schoenherr (2011)과 Samvedi 등(2013)의 연구에서는 공급망에서 발생할 수 있는 리스크 요인에 중점을 두어 공급망 리스크 유형을 제안하였다. Harland 등(2003)과 Wu 등(2006)의 연구에서는 공급망 리스크를 외부적 리스크와 내부적 리스크로 분류한 후 유형을 제안하였다. 공급망 리스크 관리 분야의 연구에서 제안된 공급망 리스크 유형을 정리하면 <표 1>과 같다.

한편, 공급망 리스크 텍스트 분석 연구에서는 공급망 리스크 관리 분야의 선행연구에서 제안된 리스크 유형을 바탕으로 텍스트 데이터의 분류에 적합한 공급망 리스크 유형이 제안된 바 있어, 본 연구의 유형분류에 참고할 수 있다. 공급망 리스크 관련 텍스트 분석 연구에서 제시된 공급망 리스크 유형을 정리하면 <표 2>와 같다.

본 연구에서는 선행연구에서 제안된 공급망 리스크의 유형에 기반하여 기사 내 등장하는 공급망 리스크 유형을 선정하였다. 이후, 선정된 리스크 유형에 기반한 기사 내 공급망 리스크 유형 식별과 리스크 유형 분류 모델 개발을 수행하였다.

<표 1> 공급망 리스크 관리 분야 선행연구의 공급망 리스크 유형

번호	저자	연도	공급망 리스크 유형
1	Harland 등	2003	Strategic, Operations, Supply, Customer, Asset impairment, Competitive, Reputation, Financial, Fiscal, Regulatory, Legal
2	Sodhi와 Chopra	2004	Disruptions, Delays, Systems, Forecast, Intellectual Property, Procurement, Receivables, Inventory, Capacity
3	Tang과 Tomlin	2008	Supply, Process, Demand, Intellectual Property, Behavioral, Political/Social
4	Tummala와 Schoenherr	2011	Demand, delay, disruption, inventory, manufacturing (process) breakdown, physical plant (capacity), supply (procurement), system, sovereign and transportation risks
5	Samvedi 등	2013	Supply, Demand, Process, Environmental risks

〈표 2〉 공급망 리스크 관련 텍스트 분석 연구의 공급망 리스크 유형

번호	저자	연도	공급망 리스크 유형
1	Su와 Chen	2018	Environmental, Sociopolitical, Financial, Economics
2	Chu 등	2019	Supply, Demand, Logistics, Environmental, Political, Financial, System, Information
3	Chu 등	2020	Supply, Demand, Logistics, Environmental, Political, Financial, System, Operational aspects
4	Ganesh, A. D.와 Kalpana, P	2022	Supply, Demand, Logistics, Environmental, Political, Quality
5	Aboutorab 등	2022	Financial shock, Trade dispute, Geopolitical conflict, Political violence, Natural catastrophe, Technological catastrophe, Disease outbreak, Humanitarian crisis, Externality

2.3 텍스트 분류 및 토픽 분석 관련 연구

텍스트의 분류를 위해서는 빈도에 기반한 전통적인 BOW(Bag Of Words) 모델이 많이 사용되어왔다. BOW는 문서에서 등장하는 단어의 등장 여부를 표현하는 기법으로서, 단어의 등장 여부를 통한 문서를 분류한다. 이에 기반하여 단어별 가중치를 두거나 주요 키워드만을 사용하는 방식으로 BOW를 수정하여 분류 알고리즘에 적용하거나 단어의 등장 여부를 점수화하여 문서를 분류하는 연구 등이 수행되었다(박단호 등, 2012; George와 Joseph, 2014; 변형호 등, 2018). 또한, 문서 분류의 성능을 높이기 위하여 단어의 등장 여부를 기계에 학습시켜 문서를 분류하는 모델을 제안하는 연구가 수행되어왔다. 김성희와 엄재은(2008)은 단어의 등장 빈도에 가중치를 부여한 TF-IDF 행렬을 인공신경망에 학습시킨 문서 분류 모델을 제안하였고, 이를 다른 문서분류 알고리즘과 비교하였다. 어균선과 이건창(2019)은 BOW를 로지스틱 회귀분석, 인공신경망, 나이브 베이저안 네트워크 등에 기반한 문서 분류 모델에 적용하여 성능을 비교하였다.

최근에는 언어모델(LM: language model)을 문서

분류에 활용하는 방안이 많이 연구되고 있다. 예로서, BERT(Bidirectional Encoder Representations from Transformer)는 2018년 구글에서 발표한 위키피디아, 도서, 웹문서 등에서 수집된 대규모의 텍스트 데이터가 학습된 사전 학습 언어모델로서(Delvin 등, 2018), 문서 분류, 예측, 감성 분석 등의 주요 텍스트 분석 작업에서 우수한 성능을 보이는 것으로 알려져 있으며, 한국어 문서에 대해서는 한국어 텍스트 데이터를 학습시킨 KoBERT 모델이 많이 사용되고 있다(SKT Brain, 2019). 최준영과 임희석(2020)은 KoBERT와 ELMO, LSTM 기반의 자연어처리 모델을 사용해 이커머스 상품평 분류를 수행하였고, 이 중 KoBERT 모델의 분류 성능이 가장 좋은 것을 밝혔다. 황상흠과 김도현(2020)은 R&D 과제정보, 특허와 같은 기술문서 분류 모델을 제안하였다. 권순보와 유진은(2022)은 온라인 진료상담 문서 분류에 KoBERT 모델을 제안하였고, FastText 모델과의 비교 결과 KoBERT 모델이 더 나은 성능을 보이는 것을 확인하였다.

한편, 수집된 텍스트를 대상으로 주요 토픽을 추출하고, 각 토픽에 대응되는 키워드와 문서를 식별하여 이슈 및 동향을 파악하기 위해 토픽 모델링 기법

이 많이 활용되고 있다(박준형과 오효정, 2017; 윤희영과 곽일엽, 2021). 토픽 모델링의 방법론으로는 단어의 등장 확률 분포에 기반하여 토픽을 구성하는 LDA(Latent Dirichlet Allocation) 기법이 널리 사용되고 있다(Blei 등, 2003). Zhao와 Wang(2018)의 연구에서는 정유 시장의 공급망 리스크 요인을 도출하기 위해 LDA 토픽 모델링이 사용되었다. Chu 등(2020)은 잠재적인 공급망 리스크 요인을 도출하기 위해 공급망 리스크 관련 선행연구 텍스트에 LDA 토픽 모델링을 수행하였다. 윤희영과 곽일엽(2021)의 연구에서는 해운물류 관련 뉴스 텍스트 데이터에 대해 LDA 토픽 모델링을 수행하여 주요 토픽을 파악하였고, 2019년과 2020년의 토픽과 주요 키워드의 차이를 비교하였다. 박기태 등(2022)은 친환경 물류 토픽을 도출하고, 기간에 따른 차이를 비교하기 위해 LDA 토픽 모델링을 수행하였다. 나진성(2022)은 컨테이너선 관련 기사자료에 LDA 토픽 모델링을 수행하여 기간별 주요 토픽을 도출하고, 키워드를 분석하였다.

본 연구에서는 토픽모델링 결과로 도출된 키워드의 등장 여부로 BOW를 구축하여 인공신경망에 학습시킨 문서 분류 모델과, 파인튜닝 작업을 거친 KoBERT 기반의 텍스트 분류 모델을 제안하고, 두 모델의 성능을 비교하였다. 또한, LDA 토픽 모델링 기법을 사용하여 수집된 공급망 리스크 관련 기사의 주요 토픽을 도출하였고, 이를 통해 공급망 리스크 관련 이슈를 분석하였다.

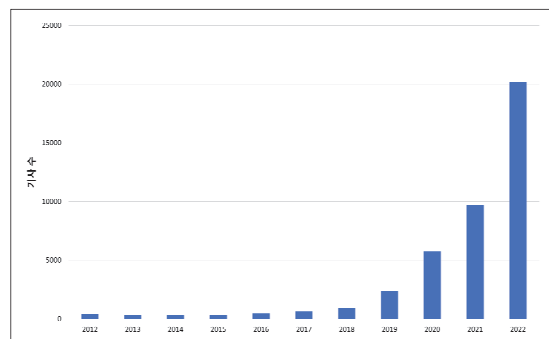
III. 웹 크롤러를 통한 기사 데이터셋 구축

실시간으로 국내 주요 언론사의 뉴스 기사를 제공

하는 서비스인 네이버 뉴스 홈페이지에서 기사를 수집할 수 있는 웹 크롤러를 Python의 BeautifulSoup 패키지를 사용하여 개발하였다. 개발된 웹 크롤러를 사용하여 2012년 1월 1일에서 2022년 12월 31일까지의 최근 11년간 네이버 뉴스 홈페이지에 게시된 주요 일간지의 뉴스 기사를 수집하였다. 검색 키워드로는 ‘공급망’, ‘공급사슬’과 ‘리스크’, ‘붕괴’, ‘위험’, ‘중단’, ‘차질’의 조합으로 총 10개의 검색 키워드를 사용하였으며, 검색 키워드별로 검색 상단에 위치하는 기사를 1일 최대 100건 수집하였다.

오류로 인해 본문이 수집되지 않는 기사와 중복으로 수집된 기사를 제외하고 총 41,569건의 기사를 수집하였다. 연도별 수집된 기사의 수는 <그림 1>과 같다. 2018년을 기점으로 수집되는 기사의 수가 증가하는 것을 확인할 수 있었다.

〈그림 1〉 연간 수집 기사 수



이렇게 수집된 뉴스 기사 데이터의 전처리를 위해 기사 본문 내에 존재하는 문장 부호와 광고성 단어, 언론사 등의 불용어를 제거하고 단어별 토큰화를 수행하였다. 이후, Mecab 라이브러리의 형태소 분석기를 사용하여 토큰화된 단어 중 일반 명사(품사 기호: NNG)와 고유 명사(품사 기호: NNP)에 해당하는 단어를 별도로 추출하였다. 수집된 뉴스 기사

데이터를 체계적으로 관리하기 위하여 본 연구에서는 Maria DBMS를 사용하여 기사 데이터셋을 데이터베이스화하였다.

IV. 공급망 리스크 관련 기사 필터링 모델 개발

4.1 공급망 리스크 관련 기사 필터링 모델 개요

수집된 기사 중 공급망 리스크와 관련이 있는 기사만을 추출하기 위해 KoBERT 모델을 활용한 문서 분류 모델을 개발하였다. 필터링 모델은 입력된 기사의 공급망 리스크 관련 여부를 판단한다. 이때, 모델의 성능을 높이고, 사용 목적에 알맞게 조정하기 위해 파인 튜닝을 수행하였다. 파인 튜닝을 위해

수집된 기사를 모델 학습에 사용하였다.

KoBERT 모델의 경우 기사 본문과 같이 길이가 긴 문서를 모델에 학습시키는 것보다 길이가 짧은 문서를 학습시키는 것이 더 나은 성능을 보이는 것으로 알려져 있다. 본 연구에서는 뉴스 기사의 제목과 기사 본문의 첫 100글자를 추출하여 필터링 모델의 파인 튜닝에 사용하였다. 뉴스 기사의 제목과 초반에 등장하는 문장이 해당 기사의 내용을 함축적으로 표현하고 있다고 판단하였기 때문이다.

4.2 파인 튜닝을 위한 데이터셋 구축

수집된 기사 중 9,597건의 기사를 직접 검토하여 공급망 리스크와 관련성이 있는 기사를 1, 관련성이 없는 기사를 0으로 라벨링하는 작업을 수행하여 파인 튜닝을 위한 데이터셋을 구축하였다. 파인 튜닝을 위해 라벨링 된 데이터의 예시는 <표 3>, <표 4>와 같다.

<표 3> 공급망 리스크와 관련된 기사 예시

기사 제목
수에즈 운하 마비... HMM 선박4척 남아공 희망봉 우회키로
中 전력난에 공장 가동 중단·정전...“‘형다’ 보다 더 위기”
반도체 수급난에...내비 뺀 차량까지 내놓는 車업체
<신종코로나> 애플 최대 하청업체, 일주일간 4조원 손실
‘시멘트 동났다’ 광주·전남 건설현장 공정 중단 확산

<표 4> 공급망 리스크와 관련되지 않은 기사 예시

기사 제목
코로나19 K-바이오 뜨자 인천 경제도 훈풍
토요타車, 지난해 세계 자동차 판매량 1위 재탈환
오세훈, 첫 국무회의부터 부동산·방역대책 놓고 충돌 <데스크 모닝픽>
로봇이 공장 점검하고 물류 하역... 현대차, 보스틴다이내믹스와 협업 본격화
<뉴욕마감>나스닥 1.2%↓...실적 앞두고 6일 만에 하락

〈표 5〉 KoBERT 기반 공급망 관련 기사 분류 모델 파인 튜닝 데이터셋

	학습 데이터	평가 데이터
공급망 리스크와 관련 없는 기사 (0)	1,907건	771건
공급망 리스크와 관련 있는 기사 (1)	4,811건	2,108건
합계	6,718건	2,879건

〈표 6〉 KoBERT 기반 공급망 관련 기사 분류 모델 분석 성능표

평가 지표	수치
Accuracy (정확도)	0.9220
Precision (정밀도)	0.9400
Recall (재현율)	0.9535
F1 Score	0.9461

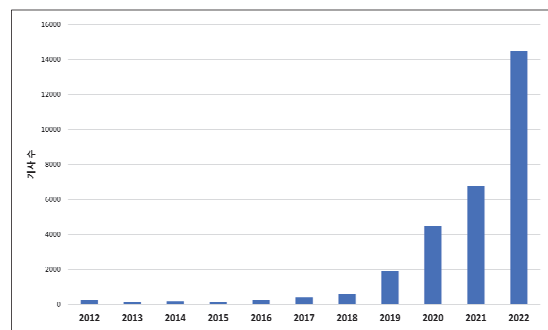
파인 튜닝을 위해 라벨링 작업을 수행한 기사 데이터 중 70%를 학습 데이터, 30%를 평가 데이터로 사용하였고, 학습 데이터와 평가 데이터의 수는 〈표 5〉와 같다.

4.3 KoBERT를 활용한 기사 필터링 모델 성능 및 필터링 결과

KoBERT 기사 분류 모델의 학습은 Epoch를 10으로 설정하여 수행하였고, 주요 평가 지표별 기사 분류 성능은 〈표 6〉과 같다. 모든 평가 지표의 수치가 약 0.92 이상으로 높은 성능을 보이는 것을 볼 수 있었고, 해당 모델을 사용하였을 때, 높은 정확도로 공급망 리스크와 관련된 기사를 추출할 수 있음을 확인하였다.

전체 기사 데이터 41,589건 중 학습 및 평가에 사용된 데이터를 제외한 나머지 기사에 필터링 모델을 적용해 분류를 수행하였다. 분류 결과, 전체 기사 중 29,670건이 공급망 리스크와 관련된 기사로 분류되었다. 연도별 공급망 리스크 관련 기사의 수는 〈그림 2〉와 같다.

〈그림 2〉 연간 공급망 리스크 기사 수



연도별 공급망 리스크 관련 기사의 수는 2019년을 기점으로 크게 증가하는 것을 볼 수 있다. 이는 2019년 발생한 한·일 무역분쟁, 2020년 발생한 코로나19 팬데믹으로 인하여 공급망 리스크와 관련된 관심이 증가하여 기사의 수가 증가한 것으로 해석할 수 있다. 또한, 러시아·우크라이나 전쟁, 미·중 무역 갈등과 금리 인상으로 인한 인플레이션과 같은 공급망 리스크 관련 이슈가 계속 발생하면서 공급망 리스크와 관련된 기사의 수가 매년 증가하는 것을 확인할 수 있다. 이러한 상황을 고려하면, 앞으로 공급망 리스크와 관련된 기사의 수는 매년 증가할 것으

로 예측할 수 있다.

V. 토픽 기반의 공급망 리스크 유형 식별 모형 수립

5.1 공급망 리스크 관련 기사 분류를 위한 공급망 리스크 유형 선정

기사에서 등장한 공급망 리스크 유형에 관한 정보 도출과 수집된 기사의 공급망 리스크 유형을 분류하기 위한 모델의 학습 데이터셋 구축을 위하여 수집된 기사의 공급망 리스크 유형을 식별하였다.

우선, 기사 내 등장하는 공급망 리스크 유형을 선행연구에 기반하여 선정하였다. 공급망 리스크는 유형에 따라 공급망에 미치는 영향이 다르고, 대응 전략이 다르게 수립된다(조지성 등, 2022). 따라서, 공급망 리스크 관련 기사의 리스크 유형 정보는 기사에서 등장한 공급망 리스크의 특징과 주요 대응 전략을 파악하고, 이에 기반한 적절한 공급망 리스크 관리 전략을 수립하는 것에 도움을 줄 수 있다.

수집되는 기사의 공급망 리스크 유형 정보를 제공하기 위해서는 기사 내 공급망 리스크의 유형을 선정할 필요가 있다. 이를 위해, 공급망 리스크 유형을 분석한 선행연구와 공급망 리스크 관련 텍스트 분석 연구에서 제시된 주요 공급망 리스크 유형을 검토하여 적합하다고 판단된 리스크 유형을 선정하였다.

공급망 리스크 유형의 선정에는 기사에서 등장하는 공급망 리스크 관련 이슈를 유형화하기에 적합한 것과 직관적으로 공급망 리스크의 유형을 알 수 있는가를 중점적으로 검토하였다. 또한, 공급망 외부에서 발생하는 이슈가 다수 보도되는 기사의 특성을

반영하여 공급망 외부에서 나타나는 리스크 요인을 주요하게 고려하여 선정을 수행하였다.

이에 따라 선행연구에서 제시된 공급망 리스크 유형과 본 연구의 목적을 고려하여 기사 내 공급망 리스크의 유형을 7가지로 선정하였다. 공급망 리스크 유형은 ‘Supply and Demand’, ‘Social’, ‘Political’, ‘Environmental’, ‘Financial’, ‘Logistics’, ‘Technological’이며, 리스크 유형에 대한 개요는 <표 7>과 같다.

5.2 토픽 모델링을 위한 최적 토픽 수 결정

29,670건의 기사를 모두 검토하여 직접 공급망 리스크 유형을 식별하는 것은 어려움이 있다고 판단되어 문서를 주제에 따라 분류하는 작업을 수행할 필요가 존재하였다. 이를 위해 본 연구에서는 LDA 토픽 모델링을 수행해 기사를 주요 토픽에 기반하여 분류하였다. 도출된 토픽에 해당하는 기사와 토픽 내 키워드를 검토하여 공급망 리스크 유형에 따라 토픽을 분류하는 작업을 수행하였다. 그 후, 각 토픽에 해당하는 공급망 리스크 유형에 따라 기사의 리스크 유형을 식별하였다.

LDA 토픽 모델링은 사전에 설정한 토픽의 수에 따라 문서 집합 내 토픽을 나눈다. 따라서, 적절한 토픽의 수를 결정해야 한다. 본 연구에서는 토픽의 수를 결정할 때 대중적으로 사용되는 지표인 Coherence score를 기준으로 토픽의 수를 결정하였다. Coherence score는 토픽 내 등장 단어들의 유사도를 측정하는 지표로 Coherence score가 높을수록, 토픽 내에 의미가 유사한 단어들이 잘 모여있음을 나타낸다. 본 연구에서는 토픽의 수가 10개에서 40개 사이일 때의 Coherence score를 계산하였다. 토픽 수에 따른 Coherence score는 <그림 3>과 같다. 본 연구는 이

〈표 7〉 공급망 리스크 유형 및 개요

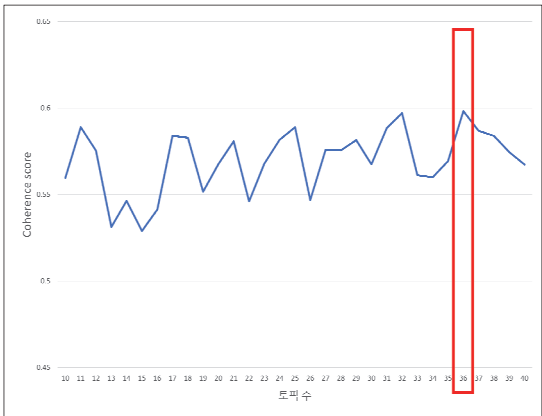
공급망 리스크 유형	개요	선행연구
Supply and Demand	공급난 및 수요 급증 등과 관련된 리스크	Ganesh와 Kalpana, 2022 Chu 등, 2020 Chu 등, 2019
Social	팬데믹, 기업의 사회적 책임, 노사 갈등 등의 사회적 요인과 관련된 리스크	Su와 Chen, 2018
Political	전쟁 및 지정학적 리스크와 국내외 정책의 변화 등의 정치적 요인과 관련된 리스크	Su와 Chen, 2018 Ganesh와 Kalpana, 2022 Chu 등, 2020 Chu 등, 2019
Environmental	지진, 홍수와 같은 자연재해 및 환경 오염, 기후 위기 등의 환경적 요인과 관련된 공급망 리스크	Ganesh와 Kalpana, 2022 Chu 등, 2020 Su와 Chen, 2018
Financial	경기, 환율, 물가 등의 재무적 요인과 관련된 리스크	Su와 Chen, 2018 Chu 등, 2020 Chu 등, 2019
Logistics	물류난, 운임 상승 등의 물류적 요인과 관련된 리스크	Ganesh와 Kalpana, 2022 Chu 등, 2020
Technological	사이버 보안, 산업 핵심 기술 등 기술적 요인과 관련된 리스크	Aboutorab 등, 2022

중 Coherence score가 가장 높았던 36개를 토픽의 수로 결정하였다.

5.3 토픽 모델링 결과

공급망 리스크와 관련되었다고 분류된 29,670건의 기사에 대하여 LDA 토픽 모델링을 수행하여 기사 데이터 내에 존재하는 토픽과 토픽 별 주요 키워드를 도출하였다. 2012년에서 2022년까지의 공급망 리스크 관련 기사에 대한 36개의 토픽과 토픽별 주요 키워드는 〈표 8〉과 같다. 토픽별 주요 키워드는 토픽 내 등장 빈도 상위 10개의 단어로 구성되어 있다. LDA 토픽모델링을 수행하면 각 기사 내 토픽의 비중을 계산할 수 있는데, 본 연구에서는 가장 큰 비중을 가진 토픽을 해당 기사의 주요 토픽이라 판단하여 이를 기사별 토픽으로 입력하였다. 각 토픽명은 토픽별 주요 키워드와 토픽에 해당하는 기사를 검토하여 결정하였다.

〈그림 3〉 토픽 수에 따른 Coherence score



<표 8> 토픽명 및 주요 키워드

번호	토픽명	주요 키워드
0	경제 지표 관련 보도	수출, 무역, 수입, 적자, 증가, 수지, 한국, 기록, 감소, 지난해
1	환율, 주식시장	투자, 증시, 시장, 하락, 증권, 지수, 연구원, 주가, 상승, 코스피
2	금리 인상 및 인플레이션	금리, 인상, 인플레이션, 연주, 물가, 상승, 기준, 시장, 정책, 경기
3	코로나19로 인한 공급망 위험	코로나, 세계, 확산, 사태, 바이러스, 신종, 감염증, 글로벌, 위기, 영향,
4	브렉시트로 인한 소비재 공급난	업체, 영국, 제품, 부족, 소비자, 공급망, 판매, 유통, 공급, 시간
5	한·일 무역분쟁	일본, 수출, 한국, 규제, 정부, 소재, 조치, 부품, 산업, 품목,
6	식량 공급망 위기	식량, 곡물, 가격, 세계, 식품, 수출, 농업, 생산, 수입, 위기
7	노사갈등	파업, 노조, 화물, 연대, 임금, 건설, 시멘트, 요구, 안전, 노동자
8	요소수 등 원자재 공급난	요소수, 수입, 공급망, 국내, 수급, 정부, 공급, 품목, 생산, 수출
9	자연재해 및 기후 위기	지역, 발생, 피해, 대만, 시설, 이번, 사고, 가뭄, 규모, 시위
10	공급망 위기 관련 정부 대응 정책 발표	지원, 중소기업, 정부, 수출, 국내, 해외, 확대, 조사, 자금, 대상
11	기후 위기 및 탄소 중립	에너지, 탄소, 석탄, 기후, 배출, 중립, 전력, 발전, 변화, 수소
12	소재·부품·장비 산업 지원 정책 발표	산업, 기술, 투자, 개발, 사업, 전략, 지원, 분야, 경쟁력, 회장
13	IRA 및 전기차 보조금 관련 정책	배터리, 전기차, LG, SK, 시장, 전기, 공장, 자동차, 생산, 업체
14	원전 관련 정책으로 인한 원전 공급망 붕괴	원전, 원자력, 전력, 건설, 에너지, 태양광, 정부, 호기, 정책, 산업
15	천연자원 등 원자재 공급망	자원, 리튬, 광물, 희토류, 원자재, 배터리, 가격, 생산, 호주, 니켈
16	국내 경제 위기	금융, 위기, 금리, 시장, 부채, 통화, 은행, 상황, 달러, 환율
17	자동차 공급망 공급난 및 중단	자동차, 생산, 공장, 부품, 업체, 차량, 가동, 현대차, 중단, 업체,
18	미·중 무역 갈등과 글로벌 공급망 위협	한국, 세계, 글로벌, 공급망, 정책, 무역, 국가, 산업, 필요, 변화
19	공급망 리스크 관련 대담, 칼럼	문제, 생각, 세계, 상황, 나라, 정도, 얘기, 사람, 부분, 교수
20	정부 및 정치권 발표 관련 보도	정부, 정책, 재정, 대통령, 국민, 후보, 일자리, 민주당, 의원, 규제
21	러시아·우크라이나 전쟁	러시아, 우크라, 유럽, 제재, 전쟁, 침공, 우크라이나, 에너지, 독일, EU
22	경제, 제조업 관련 지표 보도	경기, 전망, 분기, 성장, 소비, 증가, 올해, 감소, 회복, 투자
23	물류 관련 보도	물류, 컨테이너, 선박, 항만, 운송, 화물, 운임, 항공, 공급망, 해운
24	공급망 리스크 관련 정부 발표	회의, 장관, 정부, 대응, 공급망, 통상, 협력, 대통령, 이날, 논의
25	코로나19의 확산	코로나, 백신, 확진, 방역, 오미크론, 변이, 확산, 집중, 마스크, 감염
26	사이버 보안	정보, 보안, 공격, 서비스, 데이터, 디지털, 사이버, 기술, 시스템, 관리
27	기업 실적 발표	분기, 시장, 매출, 실적, 삼성전자, 이익, 영업, 수요, LG, 사업
28	증권 관련 보도	지수, 시장, 하락, 뉴욕, 증시, 이날, 주가, 분기, 시간, 기록
29	ESG 관련 공급망 리스크	ESG, 사회, 인권, 노동, 관련, 관리, 경영, 환경, 공급망, 평가
30	국내 지정학적 리스크	대통령, 한국, 북한, 외교, 대만, 안보, 정상, 관계, 바이든, 국가
31	물가 상승 및 인플레이션	물가, 상승, 가격, 소비자, 인상, 유가, 원자재, 국제, 금등, 이후
32	반도체 산업 공급망 강화	반도체, 삼성전자, 생산, 공장, 대만, TSMC, 업체, 투자, 장비, 파운드리
33	중국 봉쇄로 인한 공급망 차질	생산, 공장, 애플, 봉쇄, 베트남, 아이폰, 상하, 중단, 업체, 공급망
34	경제 전망 관련 보도	세계, 전망, 성장, 올해, IMF, 한국, GDP, 국가, 예상, 하향
35	미, 중 무역 갈등 및 경제 제재	무역, 대통령, 화웨이, 트럼프, 바이든, 관세, 정부, 국가, 행정부, 통신

5.4 기사 토픽과 공급망 리스크 유형 대응

각 토픽과 본 연구에서 제안한 공급망 리스크 유형을 대응시키기 위해, 토픽명, 주요 키워드, 토픽에 해당되는 기사를 검토하여 해당하는 공급망 리스크 유형과 연결하는 작업을 수행하였다. 검토 결과, 토

픽 19, 20, 24는 공급망 리스크와 관련된 사설, 정부 발표와 관련된 토픽으로 특정 공급망 리스크 유형에 해당하는 기사로 분류하기에 어렵다고 판단하여 리스크 유형화에서 제외하였다. 33개 토픽의 공급망 리스크 유형을 연결한 결과는 다음 <표 9>와 같다.

3개의 토픽에 해당하는 기사를 제외한 27,935건

<표 9> 토픽별 해당 공급망 리스크 유형

공급망 리스크 유형	번호	토픽명
Supply and Demand	4	브렉시트로 인한 소비재 공급난
	6	식량 공급망 위기
	8	요소수 등 원자재 공급난
	15	천연자원 등 원자재 공급망
	17	자동차 공급망 공급난 및 중단
	33	중국 봉쇄로 인한 공급망 차질
Social	3	코로나19로 인한 공급망 위협
	7	노사갈등
	25	코로나19의 확산
	29	ESG 관련 공급망 리스크
Political	5	한·일 무역분쟁
	10	공급망 위기 관련 정부 대응 정책 발표
	12	소재·부품·장비 산업 지원 정책 발표
	13	IRA 및 전기차 보조금 관련 정책
	14	원전 관련 정책으로 인한 원전 공급망 붕괴
	18	미·중 무역 갈등과 글로벌 공급망 위협
	21	러시아·우크라이나 전쟁
	30	국내 지정학적 리스크
	35	미, 중 무역 갈등 및 경제 제재
environmental	9	자연재해 및 기후 위기
	11	기후 위기 및 탄소 중립
Financial	0	경제 지표 관련 보도
	1	환율, 주식시장
	2	금리 인상 및 인플레이션
	16	국내 경제 위기
	22	경제, 제조업 관련 지표 보도
	27	기업 실적 발표
	28	증권 관련 보도
	31	물가 상승 및 인플레이션
	34	경제 전망 관련 보도
Logistics	23	물류 관련 보도
Technological	26	사이버 보안
	32	반도체 산업 공급망 강화

의 기사에 대해 토픽에 해당하는 공급망 리스크 유형을 라벨링하였다. 공급망 리스크 유형별 기사의 수는 <표 10>과 같다.

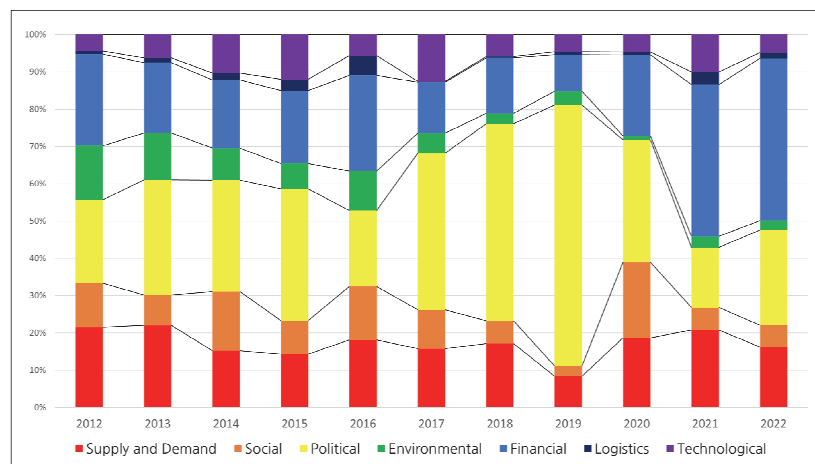
연도별 공급망 리스크 유형별 기사 수의 비중을 시각화한 결과는 <그림 4>와 같다. 연간 공급망 리스크 유형별 기사의 수를 살펴보면, 모든 기간에서 Political과 Financial 리스크에 해당하는 기사의 비중이 큰 것을 볼 수 있다. 연도별 비중이 큰 주요

기사를 살펴보면, 2019년에는 한·일 무역 갈등으로 인한 Political 리스크 관련 기사가 주요하게 나타나는 것을 확인할 수 있다. 2020년에는 코로나19로 인한 Social 리스크 관련 기사의 비중이 증가하는 것을 확인할 수 있고, 2021년과 2022년에는 금리 인상과 물가 상승 관련 이슈로 인한 인플레이션과 관련된 Financial 리스크 관련 기사의 비중이 증가하는 것을 확인할 수 있다.

<표 10> 공급망 리스크 유형별 기사 수

번호	공급망 리스크 유형	기사 수
1	Supply and Demand	4,815
2	Social	2,279
3	Political	7,873
4	Environmental	782
5	Financial	9,950
6	Logistics	507
7	Technological	1,729
	합계	27,935

<그림 4> 연간 공급망 리스크 유형별 기사 비중



VI. 텍스트의 공급망 리스크 유형 분류 모델 개발

실시간으로 수집된 공급망 리스크 관련 기사의 공급망 리스크 유형을 분류하기 위하여 KoBERT 모델을 기반의 공급망 리스크 유형 분류 모델과 BOW 기반의 인공신경망 공급망 리스크 유형 분류 모델을 개발하였다.

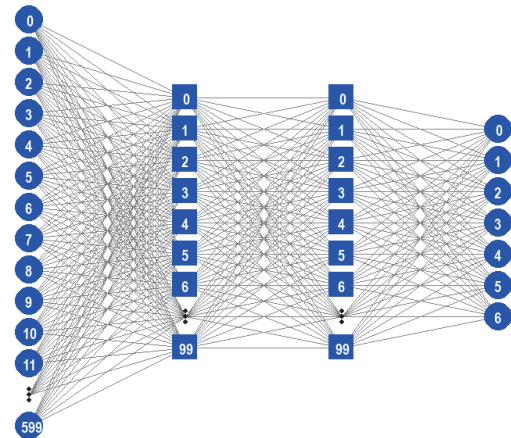
6.1 BOW 기반 인공신경망 공급망 리스크 유형 분류 모델 개발

BOW 기반의 인공신경망을 공급망 리스크 유형 분류 모델은 LDA 토픽 모델링을 통해 도출된 각 토픽의 주요 키워드들의 등장 빈도 기반의 BOW를 모델에 학습시켜 개발하였다. 33개 토픽의 20개씩의 키워드로 총 660개의 키워드의 등장 여부를 계산하였다. 토픽별 키워드에 중복이 존재하나, 이를 고려하지 않고 빈도를 계산하였다. 이때, 특정 키워드가 한 문서 안에서 다수 등장하여 등장 빈도의 극단적인 차이가 발생하는 것을 고려하여 4회 이상 등장한 키워드는 모두 4회 등장한 것으로 정규화하였다.

전체 기사 27,935건 중 70%에 해당하는 19,555건의 기사 데이터를 모델 학습에 사용하였고, 30%에 해당하는 8,380건의 기사 데이터를 평가에 사용하였다. 인공신경망은 입력층과 2개의 은닉층, 출력층으로 구성된 인공신경망을 만들었고, 입력층에는 키워드 개수인 660개의 입력층 노드가 들어갔고, 2개의 은닉층에는 100개의 뉴런을 두었으며 활성화 함수 ReLU를 사용하였다. 출력층에는 7개의 공급망 리스크 유형을 분류해야 하므로 7개의 뉴런을 두었으며, 활성화함수로 Softmax를 사용하였다. 수립된

인공신경망의 구조는 <그림 5>와 같다.

<그림 5> BOW 기반 인공신경망 구조



모델의 학습은 Epoch를 1,000으로 설정하여 수행하였다. 모델 학습 결과, 전체 분류 정확도는 약 0.75로 나타났다. 각 공급망 리스크 유형별 분류 성능은 <표 11>과 같다.

6.2 KoBERT를 활용한 공급망 리스크 유형 분류 모델 개발

KoBERT를 사용한 공급망 리스크 유형 분류 모델의 파인 튜닝을 위해, 제목과 본문 첫 100글자를 합친 데이터를 학습과 평가에 사용하였다. 전체 기사 데이터 중 70%에 해당하는 19,555건의 기사 데이터를 모델 학습에 사용하였고, 30%에 해당하는 8,380건의 기사 데이터를 평가에 사용하였다. 모델의 학습은 Epoch를 10으로 설정하여 수행하였다. 모델 학습 결과, 전체 분류 정확도는 약 0.82로 나타났다. 각 공급망 리스크 유형별 분류 성능은 <표 12>와 같다.

〈표 11〉 BOW 기반 인공신경망 공급망 리스크 유형 분류 모델 분석 성능

	Precision	Recall	F1-score	Support
Supply and Demand	0.7237	0.6404	0.6795	1,485
Social	0.6183	0.6381	0.6280	688
Political	0.7628	0.7798	0.7712	2,343
Environmental	0.5660	0.5128	0.5381	234
Financial	0.8232	0.8490	0.8359	2,967
Logistics	0.6026	0.6714	0.6351	140
Technological	0.6717	0.6845	0.6780	523
Accuracy			0.7527	8,380
Macro avg	0.6812	0.6823	0.6809	8,380
Weighted avg	0.7516	0.7527	0.7515	8,380

〈표 12〉 KoBERT 기반 공급망 리스크 유형 분류 모델 분석 성능

	Precision	Recall	F1-score	Support
Supply and Demand	0.7691	0.7933	0.7810	1,461
Social	0.7650	0.6711	0.7150	684
Political	0.8231	0.8155	0.8193	2,368
Environmental	0.6895	0.6453	0.6667	234
Financial	0.8726	0.8921	0.8823	2,957
Logistics	0.7160	0.8227	0.7657	141
Technological	0.7533	0.7364	0.7448	535
Accuracy			0.8171	8,380
Macro avg	0.7698	0.7681	0.7678	8,380
Weighted avg	0.8164	0.8172	0.8164	8,380

6.3 공급망 리스크 유형 분류 모델 성능 비교

두 모델의 평균 분류 성능을 비교한 결과는 〈표 13〉과 같다. 모든 성능 지표에서 KoBERT 기반의 분류 모델이 더 나은 성능을 보이는 것을 확인할 수 있다. KoBERT 모델에서는 BOW 기반 모델의 학습에 사용된 키워드를 포함한 더 많은 단어의 등장 여부와

등장 순서, 주변 단어 등이 동시에 학습된다. 이러한 차이로 인하여 KoBERT 모델이 더 나은 성능으로 공급망 리스크를 분류할 수 있는 것으로 해석할 수 있다. 또한, 언어모델을 활용한 텍스트 분석 기법의 적용이 공급망 리스크 분야의 텍스트 분석을 더 고도화할 수 있음을 보여준다.

〈표 13〉 공급망 리스크 유형 분류 모델 성능 비교

	KoBERT 기반 분류 모델	BOW 기반 분류 모델
Accuracy	0.8171	0.7527
Precision	0.8164	0.7516
Recall	0.8172	0.7527
F1-score	0.8164	0.7515

Ⅶ. 결론

7.1 연구 결과와 시사점

본 연구는 텍스트 마이닝을 통해 미디어에서 등장하는 공급망 리스크 관련 뉴스 기사를 수집해 관련 동향을 파악할 수 있는 모니터링 시스템을 개발하고, 분석 결과를 제시하였다.

연구 결과를 요약하면 다음과 같다. 우선, 뉴스 기사를 실시간으로 수집할 수 있는 웹 크롤러를 개발하여 2012년에서 2022년에 게시된 기사를 일차적으로 수집하였다. 언어모델 KoBERT 기반의 공급망 리스크 관련 기사 필터링 모델을 개발하였고, 전반적인 성능 평가 지표에서 0.92 이상의 높은 분류 성능을 보이는 것을 확인하였다.

수집된 기사의 공급망 리스크 유형 정보를 제공하기 위하여 선행연구에 기반해 기사에서 등장하는 공급망 리스크의 유형을 Supply and Demand, Social, Political, Environmental, Financial, Logistics, Technological로 총 7가지 유형을 제안하였다.

필터링 된 공급망 리스크 관련 기사의 리스크 유형을 식별하기 위해 LDA 토픽 모델링을 수행하였다. 토픽 모델링 결과 총 36개의 토픽을 도출하였으며, 이를 통해 수집된 기사의 공급망 리스크 유형을 식별할

수 있었다. 기사의 공급망 리스크 유형을 실시간으로 분류할 수 있는 모델을 개발하기 위하여 KoBERT 기반의 문서 분류 모형과 BOW 기반의 인공신경망 문서 분류 모형을 개발하였고, 두 모델의 분류 성능을 비교하였다. 분석 결과, KoBERT 기반 모델의 분류 성능이 더 뛰어난 것을 확인할 수 있었다.

본 연구는 실시간으로 공급망 리스크와 관련된 기사를 수집하여 공급망 리스크 유형을 분류하여 시스템 이용자에게 전달할 수 있는 시스템을 제시하였다. 본 연구의 결과를 통해 다음의 시사점을 제공하였다.

본 연구의 학술적 시사점은 다음과 같다. 우선, 공급망 리스크 관련 텍스트 분석에서 충분히 적용되지 않았던 딥러닝 기반의 자연어처리 기법을 공급망 모니터링 시스템에 적용하였고, 높은 분석 성능을 보이는 것을 밝혔다. 이를 통해 향후 연구에서 딥러닝 기반의 자연어처리 기법을 사용한 고도화된 공급망 리스크 모니터링이 가능할 것을 시사하였다는 점에서 의의가 있다. 두 번째로 실시간으로 수집되는 공급망 리스크 관련 기사를 분석하여 기사 내 공급망 리스크 유형을 제안하였다는 점에서 의의가 있다. 이러한 연구 방법을 활용한다면, 특정 기간의 주요 공급망 리스크 유형 도출과 패턴 분석과 같은 공급망 리스크 동향 파악 연구에 기여할 수 있을 것으로 기대된다.

실무적 시사점은 다음과 같다. 우선, 실시간으로 수집된 공급망 리스크 관련 기사를 통하여 기업 혹은 정부의 공급망 리스크 동향 파악에 도움을 줄 수 있

는 정보를 제공할 수 있다는 점에서 의의가 있다. 본 연구에서 수행된 기사 수집 및 필터링 방법을 활용한다면, 공급망 리스크와 관련된 정보를 빠르게 파악할 수 있을 것으로 기대된다. 두 번째로 공급망 리스크 관리 전략 수립에 도움을 줄 수 있는 공급망 리스크 유형 정보를 제공할 수 있다는 점에서 의의가 있다. 기사에서 등장하는 공급망 리스크의 유형을 빠르게 식별하여 리스크 유형에 따른 예방 및 대응 전략 수립에 도움이 될 것으로 기대된다. 마지막으로, 시계열 분석을 통해 공급망 리스크 동향의 트렌드를 파악할 수 있다는 점에서 의의가 있다. 특정 리스크 유형과 관련된 기사의 비중 증감을 파악하여 주요하게 나타나는 공급망 리스크를 파악할 수 있을 것으로 기대된다.

7.2 연구의 한계점 및 향후 연구 방안

본 연구의 한계점은 다음과 같다. 우선, 전반적인 공급망 리스크 동향을 파악하기 위해 공급망 문제의 원인이 되는 이슈가 보도된 기사와 공급망 문제가 발생한 결과가 보도된 기사를 함께 수집하였다. 향후 연구에서 두 유형의 기사를 나누어 수집, 분석한다면, 공급망 문제의 잠재적인 요인을 사전에 파악하고, 공급망 문제 중 중요하게 보도되는 문제를 분석할 수 있을 것으로 기대된다.

두 번째로 공급망 리스크 유형 분류를 위해 사용된 토픽별 주요 등장 키워드의 한계가 존재한다. 학습에 사용된 키워드 중에는 특정 토픽을 대표한다고 보기 어려운 보편적인 단어가 일부 존재하고, 각 토픽의 키워드가 중복되는 경우도 존재하였다. 향후 연구에서는 공급망 리스크 관점에서 특정 이슈, 토픽, 리스크 유형을 대표할 수 있는 유의미한 키워드를 도출하는 방법이 고려되어야 할 것으로 보인다.

세 번째로 언어모델을 텍스트 분석에 적용하기 위한 파인 튜닝에서 각 공급망 리스크 유형의 기사를 충분히 수집하여 학습을 수행한다면, 공급망 리스크 유형 간 분류 성능을 향상시키고, 유형 간 분류 성능의 차이를 줄일 수 있을 것으로 보인다. 또한, 향후 연구에서는 본 연구에서 사용한 KoBERT와 더불어, GPT, BARD, HyperCLOVA X 등 언어모델의 분석 성능을 비교하여 공급망 리스크 관리 분야의 텍스트 분석에 적합한 언어모델을 밝히는 연구가 수행될 수 있을 것으로 기대된다.

마지막으로 새로운 유형의 공급망 리스크의 등장을 고려하지 않고, 공급망 리스크 유형을 제안하였다는 한계가 존재한다. 향후 연구에서는 새로운 공급망 리스크 유형이 등장하였을 때, 이를 감지하고 새로운 공급망 리스크 유형, 요인의 등장을 인공지능이 자체적으로 판별할 수 있는 모델을 개발하는 연구가 수행될 수 있을 것으로 기대된다.

참고문헌

- 권순보, 유진은 (2022), “BERT와 FastText를 활용한 온라인 진로상담 문서 분류,” 한국데이터정보과학회지, 제33권 제6호, 991-1006.
- 김성희, 엄재은 (2008), “기계학습을 이용한 문서 자동분류에 관한 연구,” 정보관리연구, 제39권 제4호, 47-66.
- 박기태, 임세혁, 김민석, 최동현, 송보미 (2022), “LDA를 활용한 친환경 물류 핵심 토픽 분석 - COVID-19 팬데믹 전후 키워드 중심으로,” 한국생산관리학회지, 제33권 제3호, 463-479.
- 박단호, 최원식, 김홍조, 이석룡 (2012), “한글 형태소 및 키워드 분석에 기반한 웹 문서 분류,” 정보처리학회

- 논문지, 제19권 제4호, 263-270.
- 박준형, 오효정 (2017), "국내 기록관리학 연구동향 분석을 위한 토픽모델링 기법 비교: LDA와 HDP를 중심으로," 한국도서관·정보학회지, 제48권 제4호, 235-258.
- 변형호, 정시현, 김종권 (2018), "주제 핵심어 기반 감정 분석을 통한 프로 스포츠 팬 사이트에서의 여론 분류 기술," 정보과학회논문지, 제45권 제4호, 390-394.
- 산업통상자원부 (2022), "「글로벌 공급망(GVC) 분석센터」 출범," <https://www.korea.kr/briefing/pressReleaseView.do?newsId=156494778>.
- 여균선, 이진창 (2019), "속성선택방법과 워드임베딩 및 BOW(Bag-of-Words)를 결합한 오피니언 마이닝 성과에 관한 연구," 디지털융복합연구, 17(2), 163-170.
- 윤희영, 광일엽 (2021), "토픽모델링을 활용한 해운물류 뉴스 분석," 무역학회지, 제46권 제4호, 61-76.
- 최준영, 임희석 (2020), "자연어처리 모델을 이용한 이커머스 데이터 기반 감성 분석 모델 구축," 한국융합학회논문지, 제11권 제11호, 33-39.
- 한국무역협회 (2022), "글로벌 공급망 위기와 우리 기업의 대응현황," https://www.kita.net/cmmrcInfo/internationalTradeStudies/researchReport/focusBriefDetail.do?no=2287&logGb=A9400_20220503
- 조지성, 이연경, 권보배, 김아름, 김업지 (2022), "글로벌 공급망 리스크별 영향분석 및 대응방안 연구 - 에너지·곡물을 중심으로," [KMI] 연구보고서.
- 최병일, 류한별, 김선규 (2023), "부품 공급망 위험이 공급사슬과 해운 분야에 미치는 영향: 자동차 산업을 중심으로," 한국생산관리학회지, 제34권 제1호, 111-128.
- 홍성태, 김창희 (2023), "공급망 리스크가 국가 물류 성과에 미치는 영향에 대한 연구: 경제협력개발기구 회원국의 물류성과지수를 활용하여," 한국생산관리학회지, 제34권 제1호, 97-110.
- 황상흠, 김도현 (2020), "한국어 기술문서 분석을 위한 BERT 기반의 분류모델," 한국전자거래학회지, 제25권 제1호, 203-214.
- Aboutorab, H., Hussain, O. K., Saberi, M., and Hussain, F. K. (2022), "A reinforcement learning-based framework for disruption risk identification in supply chains," *Future Generation Computer Systems*, Vol.126, 110-122.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003), "Latent dirichlet allocation," *Journal of Machine Learning Research*, Vol.3, Jan, 993-1022.
- Chu, C. Y., Park, K., and Kremer, G. E. (2019), "Applying text-mining techniques to global supply chain region selection: considering regional differences," *Procedia Manufacturing*, Vol.39, 1691-1698.
- Chu, C. Y., Park, K., and Kremer, G. E. (2020), "A global supply chain risk management framework: An application of text-mining to identify region-specific supply chain risks," *Advanced Engineering Informatics*, Vol.45, 101053.
- Deiva Ganesh, A., and Kalpana, P. (2022), "Supply chain risk identification: a real-time data-mining approach," *Industrial Management and Data Systems*, Vol.122, No.5, 1333-1354.
- Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2018), "Bert: Pre-training of deep bidirectional transformers for language understanding."
- Harland, C., Brenchley, R., and Walker, H. (2003), "Risk in supply networks," *Journal of Purchasing and Supply Management*, Vol.9, No.2, 51-62.
- Murakami, S., and Muraoka, S. (2022), "Exploring the Potential of Internet News for Supply Risk Assessment of Metals," *Sustainability*,

- Vol.14, No.1, 409.
- Samvedi, A., Jain, V., and Chan, F. T. (2013), "Quantifying risks in a supply chain through integration of fuzzy AHP and fuzzy TOPSIS," *International Journal of Production Research*, Vol.51, No.8, 2433-2442.
- SKT Brain (2019), "Korean BERT pre-trained cased," <https://github.com/SKTBrain/KoBERT>.
- Sodhi, M. S., and Chopra, S. (2004), "Managing risk to avoid supply-chain breakdown," *MIT Sloan Management Review*, Vol.46, No.1, 53-61.
- Soumya George, K., and Joseph, S. (2014), "Text classification by augmenting bag of words (BOW) representation with co-occurrence feature" *IOSR Journal of Computer Engineering*, Vol.16, No.1, 34-38.
- Su, C. J., and Chen, Y. A. (2018), "Risk assessment for global supplier selection using text mining," *Computers and Electrical Engineering*, Vol. 68, 140-155.
- Tang, C., and Tomlin, B. (2008), "The power of flexibility for mitigating supply chain risks," *International Journal of Production Economics*, Vol.116, No.1, 12-27.
- Tummala, R., and Schoenherr, T. (2011), "Assessing and managing risks using the supply chain risk management process (SCRMP)," *Supply Chain Management: An International Journal*, Vol.16, No.6, 474-483.
- Wu, T., Blackhurst, J., and Chidambaram, V. (2006), "A model for inbound supply risk analysis," *Computers in Industry*, Vol.57, No.4, 350-365.

Development of Media Text Analysis Based Supply Chain Risk Monitoring System*

Dongyeop Choi** · Yongwon Seo***

Abstract

Recently, companies are exposed to various supply chain risks such as intensified trade conflicts, epidemics, economic and geopolitical uncertainties, and natural disasters. Thus there is increasing importance in monitoring information related to supply chain risks. Analyzing real-time media texts, such as news articles, can be utilized for monitoring up-to-date information supply chain risks. However, researches regarding analyzing supply chain risk related text are in early stages, and researches to apply modern AI techniques such as deep learning-based natural language processing to supply chain risk texts are scarce. This study aims to develop a supply chain risk monitoring system that monitors and extracts information related to supply chain risks by analyzing news articles. To collect supply chain risk related articles a filtering model based on KoBERT is developed, of which risk types are identified based on LDA topic modeling to be utilized as the train data. To predict news articles' risk types, two deep learning-based risk classification models are developed using BOW(Bag of Words) and KoBERT. The results showed high accuracy of KoBERT based model in filtering supply chain risk-related articles, and in the classification of supply chain risk types also KoBERT based model showed better performance than BOW based model.

Keywords: Supply chain risk, Text mining, Language model, KoBERT, LDA topic modeling

* This research was supported by the Chung-Ang University Research Scholarship Grants in 2021.

** Master Student, Department of Business Administration, Chung-Ang University (First Author)

*** Professor, Department of Business Administration, Chung-Ang University (Corresponding Author)



© Korean Production and Operations Management Society. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
(<https://creativecommons.org/licenses/by-nc/4.0/deed.ko>)

