

# 뉴스 기사 텍스트 마이닝과 네트워크 분석을 통한 폭염의 사회·경제적 영향 유형 도출: 2012~2016년 사례

정재인<sup>1)</sup> · 이경준<sup>1)</sup> · 김승범<sup>2)\*</sup>

<sup>1)</sup>기상청 영향예보추진팀, <sup>2)</sup>국립기상과학원 재해기상연구부

(접수일: 2020년 5월 18일, 수정일: 2020년 7월 1일, 게재확정일: 2020년 7월 19일)

## Text Mining and Network Analysis of News Articles for Deriving Socio-Economic Damage Types of Heat Wave Events in Korea: 2012~2016 Cases

Jae In Jung<sup>1)</sup>, Kyoungjun Lee<sup>1)</sup>, and Seungbum Kim<sup>2)\*</sup>

<sup>1)</sup>Impact forecast Promotion Team, Korea Meteorological Administration, Seoul, Korea

<sup>2)</sup>High Impact Weather Research Department, National Institute of Meteorological Sciences, Gangneung, Korea

(Manuscript received 18 May 2020; revised 1 July 2020; accepted 19 July 2020)

**Abstract** In order to effectively prepare for damage caused by weather events, it is important to proactively identify the possible impacts of weather phenomena on the domestic society and economy. Text mining and Network analysis are used in this paper to build a database of damage types and levels caused by heat wave. We collect news articles about heat wave from the SBS news website and determine the primary and secondary effects of that through network analysis. In addition to that, based on the frequency with which each impact keyword is mentioned, we estimate how much influence each factor has. As a result, the types of impacts caused by heat wave are efficiently derived. Among these types of impacts, we find that people in South Korea are mainly interested in algae and heat-related illness. Since this technique of analysis can be applied not only to news articles but also to social media contents, such as Twitter and Facebook, it is expected to be used as a useful tool for building weather impact databases.

**Key words:** Text-mining, natural language processing, network analysis, heat wave, weather impacts

### 1. 서 론

최근 기후가 변화함에 따라 위험기상으로 인한 사회·경제적 피해가 점차 증가하고 있다. 2015년 재해연보(MPSS, 2016)에 따르면, 우리나라에서 1961~1970년 기간 동안 기상 재해로 인해 연평균 1천 3백억 원의 피해가 발생하였으나, 2001~2010년 기간 동안은

2조 1천억원의 피해가 발생하는 등 기상으로 인한 피해는 꾸준히 증가하는 경향을 보인다(Fig. 1).

이러한 피해는 기술이 고도화되어 매우 정확한 날씨 예보가 제공됨에도 불구하고 발생하기도 한다. 2016년 1월 23일 제주 지역에는 강풍을 동반한 폭설이 예보되었지만, 항공기 결항을 예상하지 못한 승객들이 3일간 공항에 고립되는 사태가 발생하였다. 이렇듯 정확한 예보가 제공되었음에도 불구하고 큰 피해가 발생하는 것은 날씨가 어떤 영향을 미칠지에 대한 이해가 부족하기 때문이다. 이 시점에서 우리는 수치적으로 정확한 예·특보 보다는 국민의 안전과 재산을 보호하기 위해서 어떤 정보를 제공해야 하는지 고민할

\*Corresponding Author: Seungbum Kim, High Impact Weather Research Department, National Institute of Meteorological Sciences, KMA, Gangneung 25457, Korea  
Phone: +82-33-913-1050, Fax: +82-33-644-2657  
E-mail: kmaskim@korea.kr

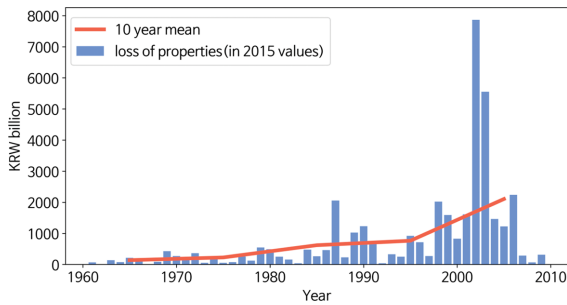


Fig. 1. Annual loss caused by natural disaster in South Korea (1961~2010) [source: MPSS (2016)].

필요가 있다(WMO, 2015).

위험기상이 미치는 잠재적 영향에 대한 정보를 제공하기 위해서는 영향 정보 데이터베이스를 구축하는 것이 중요하다. 영향 정보 데이터베이스에는 위험기상이 사회·경제에 어떤 영향을 미치는지, 어떤 영향에 의한 피해가 크게 나타나는지, 수요자는 어떤 영향 유형에 주로 관심을 가지는지에 대한 정보가 포함되어야 한다. 하지만 그 중요성에 비해 아직까지 활용도 높은 위험기상 영향 데이터베이스 구축은 미흡한 실정이다.

국내에서 위험기상 영향을 파악하기 위해 사용할 수 있는 자료로는 대표적으로 행정안전부에서 매년 발행하는 재해연보가 있다. 재해연보는 재해 피해 및 복구와 관련된 중요한 통계 정보를 제공한다. 다만 통계에서 5개 기상 현상(태풍, 호우, 대설, 강풍, 풍랑)만을 다룬다는 점, 자료 수집 방법 상 오랜 시간이 소요된다는 점 등의 제한점이 있다. 또한 재해연보에서 제공하는 인명 피해 및 피해·복구액 정보로는 국민들이 어떤 영향에 관심을 가지는지 파악하기 어렵다. 재해연보를 비롯한 위험기상 피해 자료는 위험기상에 의한 1차 피해 사례를 다루기 때문에 1차 피해에 이어지는 2·3차 피해나, 영향 간의 연관 관계는 파악할 수 없다는 점도 기상 영향 데이터베이스로 사용하기에는 한계로 작용한다.

본 연구에서는 기존의 기상 영향 자료의 한계점을 보완하고, 보다 상세한 데이터베이스를 구축하기 위한 방법 중 한가지로 텍스트 마이닝(Text-mining)과 네트워크 분석(Network analysis) 방법을 제시하고자 한다.

텍스트 마이닝은 기계 학습(Machine learning), 자연어 처리(Natural language processing) 등의 방법을 결합하여 비정형화된 텍스트로부터 정보를 얻어내는 기법을 말한다(Gupta and Lehal, 2009). 특정 규격으로 갖추어진 데이터베이스가 아닌 소셜 미디어(social media), 뉴스 기사, 메신저 대화까지도 분석이 가능하며, 방대한 양의 자료로부터 기존에 없던 새로운 정

보를 얻어낼 수 있다는 장점이 있다. 이 분석 방법은 최근 의학, 패션 등 여러 분야에서 점차 활발히 활용되고 있는 추세로, 특히 소비자가 직접 게시한 소셜 미디어 및 온라인 상의 게시글을 분석함으로써, 수요자의 선호도 및 트렌드를 파악하고 마케팅에 활용하는 연구가 활발히 논의되고 있다(Mittermayer, 2004; Rickman and Cosenza, 2007; Nassirtoussi et al., 2014).

네트워크 분석 기법은 개념들 간의 연결 관계를 분석하여, 분석 대상의 구조를 파악하고 내재되어 있는 의미를 심도 있게 이해할 수 있도록 하는 기법이다. 텍스트 네트워크 분석은 네트워크 분석을 텍스트에 적용한 것으로, 텍스트 내 단어들의 구조적 배열과 등장 빈도를 통해 텍스트의 의미를 파악하고 단어 간 관계를 분석하는 것이다. 만약 서로 다른 두 단어가 같은 텍스트 내에 빈번하게 등장하고, 가까이 위치한다면 이들은 강한 관계성을 가지게 된다. 객관적이고 정형화된 규칙을 통해 관계성을 분석하므로 연구자의 주관성을 최소화할 수 있으며, 분석 속도가 빠르다는 장점이 있다.

현재까지 텍스트 마이닝 기법을 적용하여 기상과 관련된 위험 요소를 도출한 연구는 많지 않다. Won and Kim (2014)의 경우 텍스트 마이닝을 활용하여 ‘기후변화 적응실패’와 ‘수자원 공급위기’를 주제로 사회 위험 이슈를 도출하고 정책적 측면에서 검토하는 연구를 진행하였다. 기상 분야의 경우, 언론, 소셜 데이터를 수집 및 분석하여 자연재해에 대한 국민들의 관심 이슈를 확인하고, 정부의 재난재해 관련 계획과 비교한 연구가 있었다(GTC, 2015). Chae et al. (2016)에서는 텍스트마이닝 기법을 뉴스기사에 적용하여 홍수의 피해 양상을 분석하였다. 또한, Sun et al. (2017)에서는 텍스트 마이닝을 통해 기상연감을 분석하여, 기상청의 업무 중점이 연도별로 어떻게 변화되어 왔는지 살펴보기도 하였다. 하지만 현재까지 텍스트 마이닝과 네트워크 분석을 활용하여 위험기상에 의한 피해를 분석한 연구는 많지 않은 것으로 분석된다.

본 연구에서는 위험기상 중 분석할 기상 요소로 폭염을 선정하였다. 폭염으로 인해 국내에서 2016년 한해 동안에만 2,125명의 온열질환자가 발생하였고, 이 중 17명은 사망하는 등 폭염이 미치는 영향은 매우 치명적이다(KCDC, 2016). 또한 탄소 배출을 현재보다 감축하지 않을 경우, 21세기 말(2070~2099년)까지 한반도 평균 기온이 6.0°C 상승하고, 폭염 발생일수가 57.4일 증가할 것이라는 전망이 있다(NIMS, 2011). 이에 따라, 폭염에 의해 발생 가능한 영향을 파악하고 피해를 저감하기 위한 지속적인 노력을 하는 것은 상당한 의의가 있다.

본 연구에서는 먼저 과거 5년간(2012~2016년)의 폭염 관련 뉴스 기사를 수집하고, 텍스트 마이닝을 통해

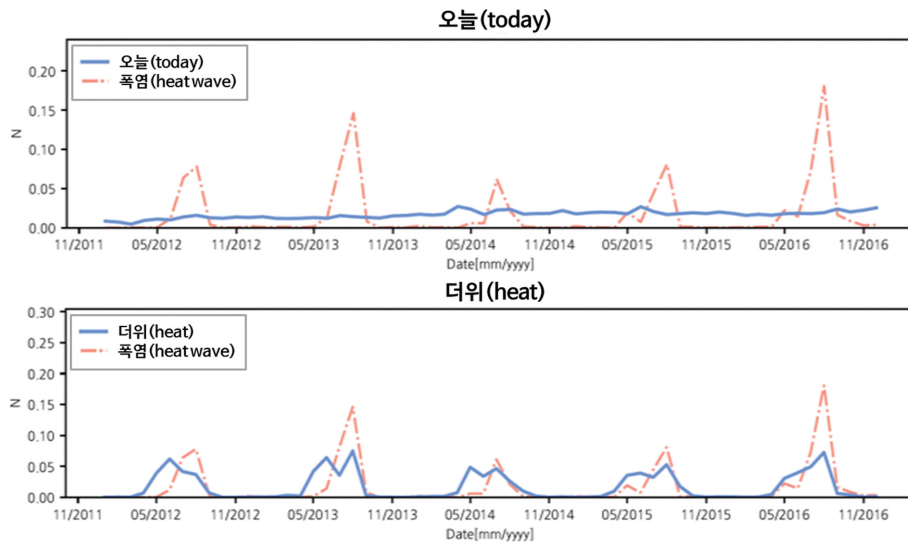


Fig. 2. Examples of word relative frequency: (top) today vs heat wave, (bottom) heat vs heat wave.

폭염에 의한 영향 유형을 도출하였다. 인터넷에서 활용 가능한 자료로 얼마만큼 양질의 데이터베이스를 구축할 수 있는지 확인하기 위해 아주 최근 자료 보다는 비교적 과거의 자료를 활용하였다. 다음으로 텍스트 네트워크 분석을 통해 도출한 영향 유형 간의 연결 관계를 살펴보았다. 네트워크 분석을 통해 1차 영향에 이어지는 2·3차 영향을 효과적으로 파악할 수 있으며, 이를 통해 폭염 발생 시 연쇄적으로 일어나는 2·3차 영향에 효율적으로 대비할 수 있을 것으로 기대한다.

## 2. 연구 방법

### 2.1 분석 자료

폭염 영향 분석을 위해 사용한 자료는 과거 5년 간 (2012년 1월~2016년 12월)의 폭염 관련 뉴스 기사 3,140건으로, SBS 뉴스 웹페이지(<https://news.sbs.co.kr>)로부터 수집하였다. 뉴스 기사는 사실을 기반으로 한 정확한 정보를 담고 있고, 맞춤법과 형식 면에서 자연어 처리가 수월하다는 장점이 있기 때문에 분석 자료로 사용하였다. 네이버 뉴스, 다음 뉴스와 같은 뉴스 플랫폼들은 같은 사례를 담은 여러 언론사의 기사가 중복 수집되어, 키워드 빈도나 기사수 분석에 영향을 주기 때문에 본 연구에서는 배제하였다. 폭염의 사회·경제적 영향과 관련된 기사만을 수집하기 위해, 인터넷 뉴스 점유율이 높은 언론사 중 검색 시 정치·경제·사회·국제 등 분야별 뉴스 필터링 기능을 제공하는 언론사를 선정하였다. 필터링 기능을 통해 해외 기사와 폭염 영향과는 관계 없는 정치 관련 기사를

배제할 수 있다. 필터링 기능의 측면에서 상위 언론사 중 연합뉴스와 SBS 뉴스가 적합하다고 판단하였으나, 연합뉴스의 경우 웹사이트에서 검색일 기준 1년 이전의 기사를 제공하지 않아 SBS 뉴스를 분석 대상으로 선정하였다.

### 2.2 자연어 처리(Natural language processing)

자연어 처리를 위해 한국어 자연어 처리에 최적화된 파이썬(Python) 패키지인 KoNLPy (Park and Cho, 2014)를 사용하였다. 수집한 3,140건의 기사 본문에 대해 자연어 처리를 시행하였고, 기사에 사용된 총 16,385개의 단어(명사)를 추출하였다.

### 2.3 폭염 연관 단어 선정

정확하고 신뢰할 수 있는 결과를 얻기 위해, 추출된 16,385개의 명사 중 분석에 사용할 폭염 연관 단어를 선별하였다. 주관성이 반영되는 것을 최소화하기 위해, 먼저 객관적인 기준에 따라 폭염 연관 단어 후보를 선별하여 목록화한 뒤, 후보 목록을 눈으로 확인하여 최종적인 폭염 연관 단어를 선정하였다.

① 적어도 10건 이상의 기사에서 언급된 단어 선별  
첫 번째 조건은 각 단어가 언급된 기사의 수를 기준으로 설정하였다. 기사 수의 기준을 높게 설정할 경우 자주 기사화되지 않는 영향 유형이 배제될 수 있어, 낮은 기준(10건)을 설정하였다. 적어도 10건의 기사에서 언급된 단어만을 선별하였고, 이 기준으로 총 16,385개의 단어 중 12,798개의 단어가 분석에서 제외되었다.

② 폭염 기간에 자주 언급된 단어 선별

**Table 1.** List of Heat wave-related words.

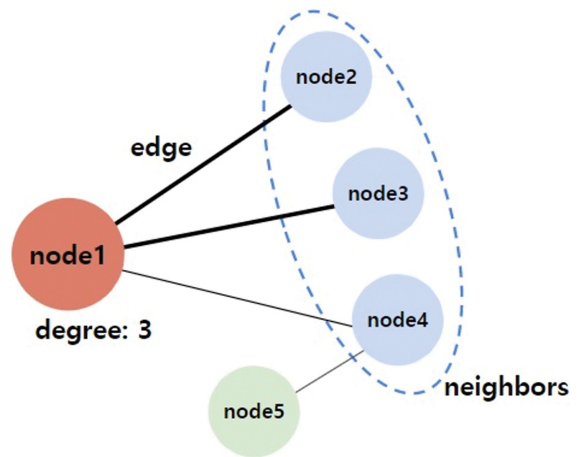
폭염 연관 단어			
변압기	적조	식수	곤충
에어컨	어류	과수원	밭일
정전	클로로필	모종	탈수
발전기	독성	감자	농사일
전기요금	물고기	참외	설사
냉방	정수	고추	열사병
절전	영산강	수박	고령자
전력경보	낙동강	오이	경련
가축	금강	가뭄	온열
축사	방류	작물	탈진
우력	하천	수확량	체온
해파리	악취	농작물	실신
갯벌	한강	옥수수	질병
폐사	수위	과수	폭염
어패류	남조류	뇌염	
수산물	수돗물	모기	
넙치	수질	벌레	
수온	녹조	일본뇌염	
가두리	강물	벌집	
어민	부패	말벌	

폭염 연관 단어의 경우 뉴스 기사에 폭염과 함께 언급되므로, 폭염 기간에 언급된 기사 수가 급증할 것이다. 따라서 폭염 관련 기사 수의 시간에 따른 변화와 각 단어를 언급한 기사 수의 시간에 따른 변화를 비교하였다(Fig. 2). Figure 2의 빨간 점선은 폭염이 언급된 기사 수의 변화를 2012년 1월부터 2016년 12월까지의 시간에 따라 나타낸 것이고, 파란 실선은 예시 단어인 ‘오늘’과 ‘더위’를 언급한 기사 수를 시간에 따라 나타낸 것이다. Figure 2의 상단 그림을 보면, 단어 ‘오늘’이 언급된 기사수의 경우 ‘폭염’과는 달리 특정 기간에 증가하거나 감소하지 않고 전반적으로 일정하게 나타난다. 반면, 하단에 나타난 ‘더위’의 경우 ‘폭염’과 유사하게 6~8월에 주로 언급되는 경향을 보인다. ‘더위’와 같이 폭염 기간에 주로 언급되는 단어만 선별한 결과 3,587개의 단어 중 3,217개의 단어가 분석에서 제외되었다.

③ 호우, 태풍 관련 단어 제외

앞서 언급한 과정 ①, ②로는 폭염과 연관되어 있지 않지만 단순히 여름철에 주기적으로 언급되는 단어를 배제하기 어렵다. 예를 들어, 우리나라의 호우와 태풍은 여름철에 집중적으로 발생하기 때문에, 이와 관련된 단어는 폭염과 같은 시기에 언급 횟수가 급증한다. 따라서 과정 ①, ②에 의해 선별된 370개의 단어 목록을 확인하여 ‘물폭탄’, ‘침수’ 등 호우, 태풍과 직접적으로 연관된 단어는 분석에서 제외하였다.

④ 장소 관련 단어 제외



**Fig. 3.** Graph theory. Circles represent nodes, lines represent edges. The degree of Node1 is 3 which means Node2, Node3 and Node4.

기사 내에 포함되어 있는 장소 관련 단어는 유용한 정보로 쓰일 수 있지만, 본 연구는 기상 현상이 미치는 영향 유형 도출을 목적으로 하므로 장소 관련 단어는 분석에서 제외하였다.

⑤ 기타 불용어 제외

다음으로, 폭염과 함께 자주 언급되지만 기상 영향과 관련한 의미를 유추할 수 없는 단어들을 제외하였다. 예를 들어, ‘여름’, ‘기승’, ‘섬씨’ 등의 경우 폭염 현상을 보도할 때 주로 쓰이는 단어들이지만, 폭염 영향과 관련한 정보는 담고 있지 않다. 이러한 기타 불용어를 분석에서 제외하였다.

⑥ 유사단어, 중복단어 정리

마지막으로 유사한 단어와 중복되는 단어를 정리하였다. ‘물고기-어류’, ‘바닷물-해수’, ‘항공기-비행기’ 등 분석에 영향을 줄 수 있는 중복단어, 유사단어들을 하나의 표현으로 통일하였다.

불용어 및 중복 단어를 정리한 결과, 총 74개의 폭염 연관 단어가 선정되었다. ‘변압기’, ‘에어컨’, ‘정전’, ‘가축’ 등 선정된 74개의 폭염 연관 단어들을 Table 1에 제시하였다.

2.4 텍스트 네트워크 분석(Text network analysis)

앞서 2.3에서 선정한 폭염 연관 단어에 대해 텍스트 네트워크 분석을 실시하였다. 각각의 폭염 연관 단어를 노드(node)로 하여 연결 관계를 파악함으로써, 기상이 미치는 주된 영향과 이어질 2차 영향(피해)을 파악할 수 있다. 텍스트 네트워크 분석을 위해 파이썬 NetworkX 패키지(Hagberg et al., 2008)를 사용하였고, 이 패키지가 제공하는 결과 그래프(graph)는

아래와 같은 정보를 담고 있다(Fig. 3):

- ① 노드(node): 분석한 단어 혹은 개념
- ② 연결선(edge): 각각의 노드를 연결하는 연결선
- ③ 이웃(neighbor): 특정 노드와 연결되어 있는 노드
- ④ 차수(degree): 노드에 연결되어 있는 연결선의 개수

### 2.5 시각화

분석을 통해 결정된 네트워크 구조는 오픈 소스(open source)로 제공되는 네트워크 시각화 및 분석 프로그램인 Gephi (Bastian et al., 2009)를 이용하여 시각화하였다. 먼저 모듈성(modularity) 분석(Newman, 2006; Blondel et al., 2008)을 통해 영향 네트워크를 하위 그룹으로 분류하였다. 모듈성은 네트워크 내에서 상대적으로 밀접한 관계를 가지고 있는 하위 집단을 찾기 위해 주로 사용되는 척도로, 모듈성이 크다는 것은 집단 내 노드들 사이의 관계가 집단 간 노드들 사이의 관계에 비해 강한 것을 의미한다. 다시 말해, 하위 그룹이 명확하게 분류될수록 모듈성이 크다는 것으로, 본 연구에서는 모듈성이 크게 나타나는 방향으로 그룹을 분류하고, 분류된 그룹을 다시 모듈성을 기준으로 점차적으로 하위 분류하였다.

하위 그룹을 명확히 표현하기 위해, Gephi의 ForceAtlas2 알고리즘(Jacomy et al., 2014)을 적용하였다. ForceAtlas2는 노드 간 연결 정도를 기준으로 노드의 위치를 결정하는 알고리즘으로, 허브(hub) 역할을 하는 노드를 중심에 두고 허브와 연결된 노드들이 주위를 둘러싼 형태로 배치된다. 허브 노드를 중심으로 연결 강도가 높은 노드끼리 군집화되어 배치되기 때문에, 분류한 하위 그룹들이 명확히 구분되어 표현된다.

## 3. 결 과

### 3.1 분석 방법의 타당성 검증

본 연구에서는 기상 또는 기상 영향을 나타내는 키

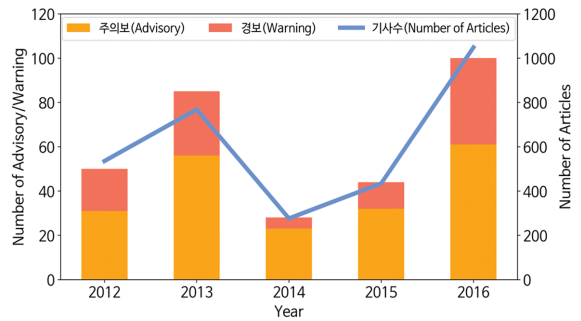


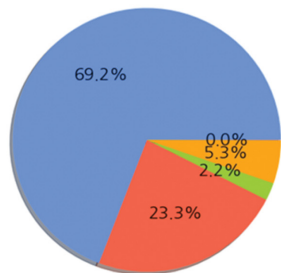
Fig. 4. Comparison of annual frequency of heat wave warning (bar) and SBS news articles about heat wave (line) in South Korea.

워드와 뉴스에 얼마나 자주 언급되는지를 통해 사람들이 해당 영향에 피해를 받는 정도를 추정하였다. 직접적인 조사를 통해 자료를 수집하는 체계가 아니어서 이에 대한 검증을 우선 수행하였다.

피해가 클수록 자주 기사화된다는 가정을 확인하기 위해, 먼저 5년간(2012~2016년) 폭염 특보 발표 횟수와 폭염 관련 SBS 뉴스 기사 수를 비교하였다(Fig. 4). Figure 4에 따르면 폭염 특보 발표 횟수가 가장 적었던 2014년에는 폭염 관련 기사 수 또한 가장 적었고, 폭염 특보 발표 횟수가 가장 많았던 2016년에는 폭염 관련 기사가 2014년에 비해 세 배 이상 많이 보도되었다. 폭염 특보 발표 횟수와 폭염 관련 뉴스 기사 수의 경향성이 정확히 일치하는 것은 기사 수를 척도로 폭염 정도를 추정하는 것이 타당하다는 것을 의미한다.

피해액이 클수록 뉴스에 자주 언급된다는 가정의 타당성을 확인하기 위해, 과거 4년간(2012~2015년) 재해연보에 제시된 위험기상(태풍, 호우, 강풍, 대설, 풍랑)별 피해액과 뉴스 기사 수를 비교하였다(Fig. 5). 뉴스 기사 수를 산정할 때, 태풍의 경우 호우와 강풍을

Damages of meteorological disaster by cause (2012~2015)



Number of SBS news articles (2012~2015)

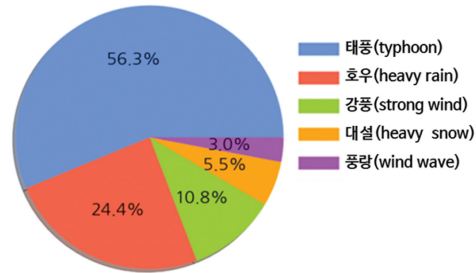


Fig. 5. Comparison of damages of meteorological disaster by cause (left) and number of SBS news articles containing each meteorological keyword (right) in South Korea (2012~2015).

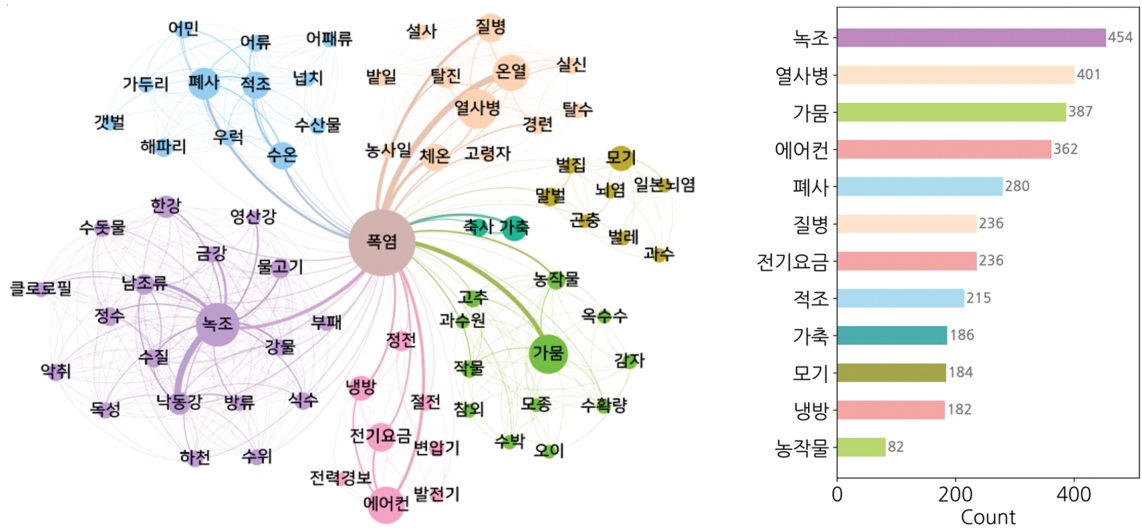


Fig. 6. Heat wave impact type diagram (left) and keyword frequencies (right) in South Korea during the year of 2012~2016.

동반하기 때문에 ‘태풍’ 키워드로 검색한 뉴스 기사에 ‘호우’와 ‘강풍’이 함께 언급된다는 문제가 있었다. 뉴스 기사 수가 중복으로 처리되는 것을 방지하기 위해, 호우·강풍 관련 기사의 경우 태풍이 함께 언급된 기사를 제외하였다. Fig. 5에 따르면 각각의 위험기상 현상별 피해액과 뉴스 기사 수의 양상이 전반적으로 유사하게 나타나는 것을 알 수 있고, 피해액이 클수록 뉴스에 자주 언급된다는 가정이 타당하다는 것을 확인할 수 있다. 따라서 본 연구에서는 뉴스 기사 수를 피해 정도를 추정하는 척도로 사용하였다.

3.2 영향 유형도

Figure 6의 왼쪽 그림은 폭염으로 인해 나타나는 영향의 유형들을 나타낸다. 군집화 된 각각의 하위 그룹은 서로 다른 색으로 표현하였고, 연결선의 두께는 두 노드 간 관계가 강할수록 굵게 표현하였다. 각 노드의 위치는 결과를 명확하게 전달하기 위해 임의로 배열한 것이다. 노드의 크기는 각각의 단어가 기사에 언급된 빈도를 나타내며, 이 중 주요 단어의 언급 횟수는 Fig. 6의 오른쪽 그림에 막대그래프와 함께 별도로 제시하였다.

3.1절에 제시한 바와 같이 기사에 자주 언급된다는 것은 폭염으로 인해 많은 사람이 피해를 입거나, 피해의 정도가 크거나, 혹은 당시의 사회·정책적인 이슈와 관련이 있기 때문인 것으로 추정하였다.

Figure 6의 왼쪽 그림을 보면, 폭염에 의한 영향 유형은 크게 어류 등 수산물 피해, 온열질환, 해충 피해, 가축 폐사, 농작물 피해, 전력 문제, 녹조 문제(Fig. 6 하늘색으로 표현된 수산물 피해 그룹부터 시계방향

순)의 7개 그룹으로 군집화되었다. 각 단어의 언급 빈도(노드의 크기와 Fig. 6 오른쪽 그림 참고)로 미루어 볼 때, 사람들이 관심 가지는 영향은 녹조문제, 열사병 등의 온열질환, 가뭄으로 인한 농작물 피해, 냉방 시설 가동으로 인한 전기요금, 수산물·가축 폐사, 해충 피해 순임을 알 수 있었다.

주목할 만한 점은 폭염에 의한 영향 중 ‘녹조’가 가장 많이 언급되었다는 점이다. 녹조는 상수원 오염, 어류 폐사 등의 피해를 야기하기도 하지만 온열질환이나 가축 폐사 등에 비해 직접적인 피해의 정도나 피해액이 적다. 그럼에도 불구하고 녹조가 폭염 관련 키워드 중 가장 많이 언급된 것은 분석 자료로 사용한 2012~2016년의 사회·정책적 이슈와 관련이 있기 때문인 것으로 추정하였다. 녹조와 함께 ‘한강’, ‘금강’, ‘영산강’, ‘낙동강’이 함께 언급된 것으로 보아 4대강 정비 사업에 따른 녹조 현상과 관련된 기사가 많이 보도되었음을 알 수 있었다. 이를 확인해보기 위해 ‘Google 트렌드’(https://trends.google.co.kr)를 활용하여 시간의 흐름에 따른 폭염과 녹조의 검색량 추이를 살펴보았다. ‘Google 트렌드’는 구글에서 제공하는 서비스로, 구글 검색 데이터를 기반으로 하여 사용자가 특정 검색어를 검색한 횟수를 지수화해 대중의 관심도를 보여준다. 지수는 0~100 사이의 값을 가지며, 최대 검색량인 100을 기준으로 하여 이에 대한 상대적인 값으로 검색량을 표출한다.

Figure 7은 2010년부터 2018년까지 폭염, 녹조가 검색된 상대적인 횟수를 시간의 흐름에 따라 나타낸 것이다. 폭염의 검색 횟수는 매년 여름철에 주기적으로 증가하는 경향을 보이고, 녹조는 2012년에 검색 횟수가

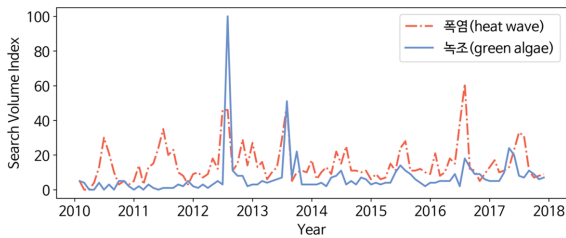


Fig. 7. Google search volume for heat wave (red) and green algae (blue) in South Korea during the year of 2010~2018.

급증한 이후 폭염과 함께 여름철에 주로 검색되는 경향을 보이는 것을 알 수 있다. 2012년 이후 녹조 현상에 대한 관심이 상당히 증가하였다고 볼 수 있는데, 이는 4대강 정비 사업이 2011년 10월에 완공된 것과 시기가 맞물린다. 4대강 정비 사업이 녹조와 무관하다는 주장과 녹조를 심화시켰다는 주장은 대립이 계속되고 있지만(Park, 2012), 폭염 관련 기사에서 녹조 키워드가 가장 많이 언급된 것은 이 사업 이후 대중의 녹조 문제에 대한 관심이 매우 높아졌다는 것을 의미한다.

녹조 문제 다음으로 자주 언급된 것은 ‘열사병’이다. ‘열사병’과 함께 언급된 ‘발열’, ‘농사일’, ‘고령자’, ‘탈진’ 등의 연관 단어로 미루어 볼 때 온열질환이 고령자에게서(특히 발열, 농사일을 하는 고령자) 주로 발생함을 알 수 있었다. 실제 통계를 기반으로 한 Kim et al. (2014) 연구에서도 21년간(1991~2011년) 발생한 폭염 사망자 중 58.7%는 60세 이상의 고령층이며, 무직을 제외하면 농림·어업종사자가 전체 폭염 사망자 중 가장 높은 직업 비율을 차지한다고 제시한 바 있다.

‘가뭄’ 또한 ‘열사병’과 비슷한 빈도로 자주 언급되었다. 함께 언급된 단어로 미루어 볼 때, 폭염과 가뭄이 이어지면서 고추, 감자, 옥수수, 참외, 오이 등에서 농작물 피해가 발생하였다고 추정할 수 있으며, 실제로 이러한 피해 사례를 다룬 기사를 상당수 확인할 수 있었다. SBS뉴스 2012년 6월 21일 기사 「계속되는 가뭄에 농작물 작황 ‘뚝’...물가 비상」에서는 폭염과 가뭄으로 감자, 양파, 마늘 등 수확기 작물이 피해를 입어, 수확량이 감소해 가격이 급등하는 현상에 대해 보도하였고, 2017년 6월 21일 기사 「농사 접어야 할 판...극심한 가뭄에 말라 죽어가는 농작물」에서도 가뭄과 폭염으로 발생한 논 농사와 감자 농사 피해에 대해 다루었다.

다음으로 많이 언급된 영향 단어는 ‘에어컨’으로, ‘전기요금’ 또한 상당수 언급된 것으로 보아 냉방 시설 가동으로 인한 전기요금 걱정이 반영된 것으로 보인다. 실제로 극심한 폭염 피해가 발생했던 2016년에는 누진세로 인한 전기요금 폭탄에 대해 다룬 기사를

상당수 찾아볼 수 있고(2016년 9월 20일 SBS 기사 「‘폭염’에 전기료 2배 298만 가구...5배 넘는 ‘요금폭탄’은 24만」 등), 이러한 관심은 2017년에 누진세 개편으로 이어졌다. ‘전기요금’ 이외에도 ‘변압기’, ‘발전기’, ‘정전’ 등의 단어도 확인할 수 있는데, 이는 폭염 시 냉방시설 가동으로 인한 전력 수요 증가로 정전 등의 전력 문제가 발생한 것으로 추정할 수 있다. 과거 국내 피해 사례를 살펴보면 전력 과부하로 인한 정전은 2차 피해를 야기하기도 하였다(2012년 8월 6일 SBS 기사 「전력 과부하로 곳곳 정전...경전철 한때 멈춰」).

‘에어컨’ 다음으로 자주 언급된 단어는 ‘폐사’이다. ‘폐사’와 함께 ‘적조’, ‘수온’도 상당수 언급되었고, 이와 더불어 ‘어패류’ 등 수산물 관련 단어와 ‘가두리’ 등이 언급되었다. 이는 매년 어민에게 큰 피해를 입히는 고수온 및 적조로 인한 수산물 폐사 문제가 반영된 것으로 보인다. 단어 ‘폐사’는 수산물과의 연관성이 강해 수산물 폐사 그룹에 포함되었지만, 가축의 폐사 또한 이에 해당한다. 국민안전처(현 행정안전부)에 따르면(MOIS, 2017), 2011~2016년 평균 612만여 마리의 어류와 210만여 마리의 가축이 폭염으로 폐사하는 등 피해 규모가 상당하여 언급 빈도 또한 높게 나타날 것으로 예상하였지만 전기 요금 등의 전력 문제와 달리 피해 대상이 어민과 농민에게 한정되어 있어 전력 문제에 비해 대중의 관심이 적게 나타난 것으로 판단된다.

마지막으로 ‘모기’, ‘말벌’ 등 병해충 관련 단어들이 가장 적은 빈도로 언급되었다. ‘모기’와 ‘일본뇌염’간의 연결선이 상대적으로 진하게 나타나는 것으로 보아, 모기가 매개하는 일본뇌염에 대한 관심이 높은 것으로 추정되며, 말벌에 의한 벌 쏘임 피해나 해충에 의한 과수 피해도 발생함을 짐작할 수 있다. 실제로 한반도의 기온이 올라가면서 동남아 외래종 등 말벌과 해충이 번성하여, 이로 인한 피해 사례가 상당수 보도된 바 있다(2012년 9월 21일 SBS 기사 「매아닌 모기떼 기승...일본뇌염 주의보」, 2016년 8월 11일 SBS 기사 「폭염에 동남아 말벌 번성...도심 물리는 이유」 등).

이처럼 영향 유형도에 제시된 폭염 영향들은 모두 실제 피해 사례를 통해 확인할 수 있었으며, 이는 텍스트 마이닝을 통한 영향 유형 도출과 네트워크 분석을 통한 영향 유형간 군집화가 효과적이고 정확하게 이루어진다는 것을 의미한다.

### 3.3 영향 관계도

앞서 살펴본 영향 유형도를 통해서 위험기상 현상에 의한 영향 유형과 영향 정도, 혹은 영향에 대한 사람들의 관심 정도를 직관적으로 파악할 수 있지만, 영

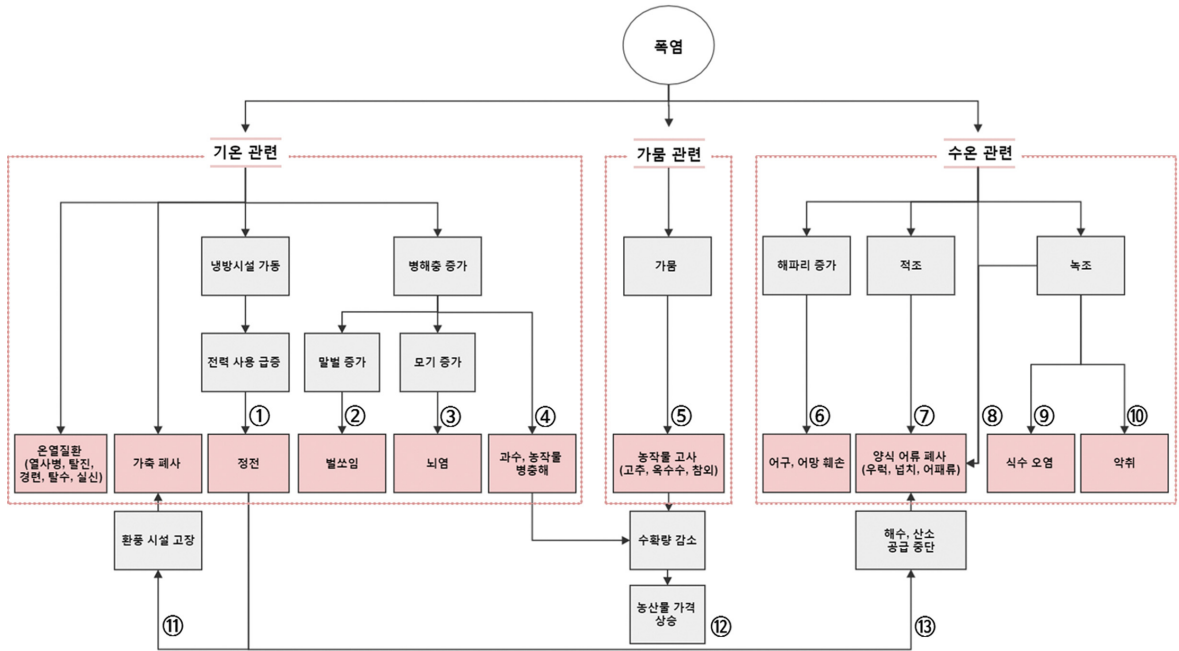


Fig. 8. Heat wave impact relationship diagram based on the data (2012~2016) in South Korea.

항간의 관계는 파악하기 어렵다. 또한, 일련의 자동화된 과정을 통해 분석할 영향 단어를 선정하는 과정에서 일부 유의미한 단어가 누락되었을 가능성이 있다. 따라서 앞서 도출된 영향 단어들을 기반으로 뉴스 기사를 검색하여 영향 유형간 관계를 파악하고, 누락된 정보를 보완하였다.

Figure 8는 영향 관계도를 나타낸 것이다. 각각의 네모 상자는 폭염으로 인해 발생할 수 있는 현상들을 나타내며, 붉은색 네모 상자는 폭염 피해를 나타낸다. 폭염에 의한 1차 영향은 크게 세가지로 분류할 수 있다. 첫째는 높은 기온으로 인해 발생하는 피해들로 온열질환, 가축 폐사, 냉방 시설 가동으로 인한 전력 문제, 병해충 등이고, 두번째는 적은 강수량과 고온으로 인한 가뭄 피해이다. 마지막은 고수온으로 인해 발생하는 피해들로, 해파리 번식으로 인한 어구, 어망 훼손, 적조 문제, 녹조 문제가 고수온 피해에 속한다.

본 연구에서 도출한 영향 관계도에서는 1차 피해가 원인으로 작용하여 또 다른 2·3차 피해를 야기할 수 있다는 것을 보여주었다. 예를 들어, 폭염으로 인해 냉방시설 가동량이 증가하면, 전력 사용이 급증하고, 전력 과부하에 의해 정전이 발생할 수 있다. 이때 정전은 폭염에 의한 1차 피해라고 볼 수 있는데, 자칫 정전에 의해 축사의 환풍 시설 가동이 중지되면서 무더위에 가축이 폐사하는 2차 피해가 발생할 수 있다. 마찬가지로 정전으로 해수와 산소가 공급되지 못해

양식 어류가 폐사하는 피해도 발생할 수 있다. 또한 녹조 발생으로 식수가 오염되거나 악취가 발생하는 문제, 폭염으로 농작물이 고사하여 농산물 가격이 상승하는 문제 등도 2차 피해가 될 수 있다.

실제로 이러한 피해들이 발생한 사례를 기사에서 찾아볼 수 있다. 2012년 8월 6일 데일리안 기사 「'폭염특보' 충북서 닭 3만마리 폐사」에서는 전력 사용 증가로 인한 정전으로 축사 환풍기 가동이 멈추면서 닭이 집단 폐사한 사례에 대해 보도하였다. 2011년 10월 5일 연합뉴스 기사 「광주·전남 정전피해 849건 57억원 집수」에서는 폭염으로 인한 전력 사용 증가로 전력 예비율이 낮아지자 정부에서 순환 정전을 시행하였고, 이에 따라 해수·산소 펌프 가동이 중단되어 어류가 폐사한 사례에 대해 보도하였다. 이와 같은 방식으로, Fig. 8의 영향 관계도에 제시한 모든 영향의 순차적인 흐름들이 실제로 발생하였는지 확인하기 위해 피해 사례를 뉴스 기사를 통해 확인하였다. 이는 뉴스 기사 텍스트 마이닝과 네트워크 분석을 통해 정립한 영향간 관계의 정확성을 검증하기 위한 것으로, SBS 뉴스 이외에도 여러 언론사의 기사를 사용하여 발생하였던 다양한 피해 사례를 확인하고자 하였다.

Figure 8의 ①-⑬는 영향의 연쇄적인 흐름에 번호를 붙여 나타낸 것이고, Table 2는 각각의 사례가 기사에 보도된 내용을 정리한 것이다. 온열질환과 가축 폐사의 경우 폭염에 의한 1차 영향으로, 3.1절에서 관련



**Table 2.** Examples of heat wave damage.

영향 관계	뉴스 기사
① 냉방시설 가동 ↓ 전력 사용 급증 ↓ 정전	· 경남 김해시 율하동 한 아파트 단지 1천500가구에서 순간 정전이 발생하여 주민 51살 김모 씨 등 27명이 엘리베이터 안에 갇히는 사고 발생(「김해·진주서 정전 잇따라...폭염에 전력 과부하 탓」 SBS 뉴스 2016.08.12.) · 부산의 한 대규모 아파트 단지에서 전기 사용 과부하로 정전이 발생하여 아파트 주민 17명이 승강기에 갇히는 등 사고 발생(「'폭염' 속 부산 아파트단지 1206세대 전기 과부하로 '정전'」 신아일보 2017.08.07.)
② 병해충 증가 ↓ 말벌 증가 ↓ 벌쏘임	· 경남 합천군 야산에서 벌초 작업을 하던 40대 남성이 땅굴에서 나온 말벌에 등을 쏘여 사망(「합천서 벌초하다 말벌 쏘여 사망」 경남신문 2016.08.23.) · 경남 함양에서 50대 등산객이 벌에 30여차례나 쏘여 사망(「폭염에 벌쏘임 사고 주의」 KNN 2017.08.07.)
③ 병해충 증가 ↓ 모기 증가 ↓ 뇌염	· 광주에 사는 50대 남성이 작은빨간집모기가 매개하는 일본뇌염에 감염되어 의식 불명(「광주서 올해 첫 일본뇌염 환자... 50대 남성, 의식 없어」 신아일보 2016.08.31.)
④ 병해충 증가 ↓ 과수, 농작물 병충해	· 충청 이남지역에서 갈색날개매미충, 미국선녀벌레, 흑옹애, 총채벌레 등에 의한 병충해가 집중적으로 발생(「"속타는 농심"...폭염에 농작물·과수 병해충 기승」 연합뉴스 2015.08.10.)
⑤ 가뭄 ↓ 농작물 고사	· 옥천·보은지역에서 폭염과 가뭄으로 과일뿐만 아니라 고추, 깻잎, 콩, 배추 등 밭작물 피해도 발생(「지속된 폭염·가뭄에 농작물 피해 증가」 충청투데이 2016.08.25.) · 경북 안동시에서 연일 지속되는 폭염과 가뭄으로 콩 12 ha, 고추 10 ha, 생강 등 기타 4.5 ha에서 피해 발생(「안동 폭염에 가뭄 겹쳐...농작물 피해 확산」 뉴스1 2016.08.22.) · 연일 계속되는 폭염으로 벼가 말라 죽고, 콩이나 고추, 깻잎 등 밭작물 피해도 발생(「폭염*가뭄에 농작물 피해 속출」 KNN 2017.08.08.)
⑥ 수온상승 ↓ 해파리 증가 ↓ 어구, 어망 훼손	· 바다수온이 크게 올라가면서 여수항 인근에서 노무라입깃해파리 개체수가 급증하고 출현 빈도도 증가하여 어구와 어망이 크게 손상 받음(「독성 해파리떼 '몸살'..어민들 삼중고」 kbc광주방송 2016.08.23.) · 고수온 현상이 지속되면서 남해안 득량만 해역에 보름달물 해파리가 대량 출현해 정치망 어업과 자망 어업을 하는 어민에게 큰 피해를 입힘(「남해안 고수온 지속 '해파리 주의보'...어촌 골머리」 MBC 뉴스 2017.08.10.)
⑦ 수온상승 ↓ 적조 ↓ 양식 어류 폐사	· 고수온과 적조의 복합 영향으로 완도 지역 금일도, 약산도, 생일도 등 3곳 해역 424 어가에서 양식 중인 전복 4천441만2천 마리가 폐사, 358억 원의 피해가 발생(「전남 완도 전복 집단 폐사 원인, 적조·고수온 복합 영향」 노컷뉴스 2016.09.13.)
⑧ 녹조 ↓ 양식 어류 폐사	· 홍제천 합류부에서 성산대교 하부 한강 본류까지 녹조가 확산되었고, 마포구 망원 한강 공원 내 홍제천 합류부에서 물고기 폐사(「한강 녹조 발생, 폐사 물고기 등등」 뉴시스 2017.06.18.) · 낙동강 합천창녕보 상류에 녹조가 창궐하면서 강준치 등 어류의 집단 폐사 발생(「낙동강 합천보 상류에 '녹조' 나타나 물고기 떼죽음」 오마이뉴스 2015.07.01.)

사례를 확인하였으므로 제외하였다. Table 2에 제시한 피해 사례와 같이 앞선 피해에 잇따라 발생하는 2·3 차 피해들은 영향간의 관계를 명확히 정립함으로써 피해가 발생하기에 앞서 미리 파악할 수 있고, 이러한 정보를 통해 피해에 앞서 대비할 수 있기 때문에 큰 의의가 있다.

#### 4. 요약 및 토의

본 연구에서는 웹 크롤링을 이용하여 과거 5년(2012~2016년) 간의 폭염 관련 뉴스 기사를 수집하고, 텍스트 마이닝 기법으로 이를 분석하여 폭염에 의한 영향 유형을 도출하였다. 텍스트 네트워크 분석 기법을

Table 2. Continued.

영향 관계	뉴스 기사
⑨ ↓ 녹조 ↓ 식수 오염	<ul style="list-style-type: none"> <li>· 계속되는 폭염으로 대전·세종·충주권 시민의 취수원인 대청호에 녹조가 확산되었고, 이를 저감시키기 위해 화학약품을 사용하면서 식수가 오염되어 고도의 정수과정을 필요로 함(「연일 폭염 때문에... 대청호·금강 일대 녹조 ‘식수 비상’」 대전일보 2016.08.22.)</li> <li>· 창원시민 식수원인 낙동강 창원시 본포취수장과 함안군 칠서정수장의 녹조 현상이 심해 정수 과정에서 약품을 과다 사용하고, 이에 따라 인체에 악영향을 미칠 것이라는 주장이 제기됨(「낙동강 녹조, 창원시민 식수 위협」 경향신문 2012.06.28.)</li> </ul>
⑩ ↓ 녹조 ↓ 악취	<ul style="list-style-type: none"> <li>· 대청호 상류인 옥천군 군북면 추소리수역에서 녹조와 쓰레기가 뒤엉켜 심각한 악취 발생(「대청호 상류 ‘악취진동 녹조라떼’ 비상」 충북일보 2017.08.20.)</li> <li>· 계속되는 폭염으로 영산강 일대 녹조가 갈수록 악화되어 산소 공급이 막혀 폐사된 물고기들이 곳곳에 즐비하고, 큰빗이끼벌레, 좁개구리밥이 발견된 곳에서는 죽은 물고기와 함께 심한 악취 발생(「폭염에 녹조 확산 영산강 악취 진동」 광주매일신문 2015.08.06.)</li> </ul>
⑪ ↓ 정전 ↓ 환풍 시설 중단 ↓ 가축 폐사	<ul style="list-style-type: none"> <li>· 괴산군 불정면의 한 농가에서 정전으로 인해 환풍기 가동이 멈추면서 사육 중인 닭 8만 마리 중 1만여 마리 폐사(「‘폭염특보’ 충북서 닭 3만마리 폐사」 데일리안 2012.08.06.)</li> <li>· 충주시 주덕읍의 한 양계농가에서 2시간 가량 정전이 되면서 환풍기가 작동하지 않아 사육 중이던 닭 4만 3000여 마리 중 1만 2000여 마리 폐사(「뽕통터위 가축들 죽어간다」 충청투데이 2013.08.15.)</li> </ul>
⑫ ↓ 농작물 고사 /과수, 농작물 병충해 ↓ 수확량 감소 ↓ 농산물 가격 상승	<ul style="list-style-type: none"> <li>· 폭우와 폭염이 이어지면서 각종 병해충이 기승을 부려 수확을 앞둔 채소와 과일 생산량이 급감. 작황 부진으로 소비자 가격이 상승하면서 상추, 시금치, 오이, 배추 등의 물가가 2~3배 상승(「폭염·폭우가 물고온 병해충... 농작물 수확 반타작」 국민일보 13면 TOP 2017.08.27.)</li> <li>· 충청지역에서 폭염과 가뭄이 지속되면서 농산물 공급량이 줄어 시금치·배추·부추·양상추·수박 등 일부 채소 가격 폭등(「충청지역 기록적인 폭염·가뭄 탓...농산물 가격 ‘폭등’」 동양일보 2016.09.01.)</li> <li>· 기록적인 폭염으로 농작물 피해가 발생하여 사과는 전년 대비 64%, 배추는 2배 이상 가격 폭등(「추석 앞두고 기록적 폭염에 채소·과일값 ‘폭등’」 MBC 뉴스 2016.08.19.)</li> </ul>
⑬ ↓ 정전 ↓ 해수, 산소 공급 중단 ↓ 어류 폐사	<ul style="list-style-type: none"> <li>· 폭염으로 전력 사용이 증가하여 전력 예비율이 낮아지자 순환 정전을 시행. 정전으로 인해 영광군 영광읍의 한 장어 양식장에서 1억4천400만원, 또 다른 양식장에서 6억원이 넘는 피해가 발생(「광주·전남 정전피해 849건 57억원 접수」 연합뉴스 2011.10.05.)</li> </ul>

이용하여 도출한 영향 유형 간 관계를 파악하였고, 폭염 영향 키워드가 기사에 언급되는 횟수를 척도로 삼아 폭염의 피해 정도를 추정하였다. 이러한 추정이 타당한지 확인하기 위해 연도별 폭염 특보 발표 횟수와 폭염 관련 SBS 기사 수를 비교하였고, 기사 수가 폭염의 발생 정도를 정확히 반영하는 것을 확인하였다. 또한, 재해연보에 제시된 위험기상 현상별 피해액(2012~2015년)과 동 기간 동안의 기상 현상 키워드를 포함하는 뉴스 기사 수를 비교하여, 기상 현상이 언급되는 기사 수와 피해액 또한 연관을 보임을 확인하였다.

분석 결과를 살펴보면, 폭염에 의한 영향으로 녹조·적조 등에 의한 수산물 폐사 및 어민 피해, 온열질환, 병해충 피해, 가축 폐사, 농작물 피해, 전력 문제 등이 도출되었고, 이 중 녹조 문제, 온열질환(열사병), 가뭄, 전기요금에 대한 언급 빈도가 특히 높게 나타

났다. 분석에 앞서 폭염 피해 중 인명과 관련된 온열 질환이나 상대적으로 피해액이 많은 수산물 및 가축 폐사에 대한 언급 횟수가 높을 것이라고 예상하였으나, 분석 결과 녹조 문제에 대한 언급이 가장 많은 것을 확인할 수 있었다. 이는 4대강 정비사업이라는 사회적 이슈와 맞물리며 대중의 관심이 높아 졌기 때문으로 추정된다. 또한, 수산물 폐사나 가축 폐사에 비해 전기 요금 등 전력과 관련된 내용이 자주 언급되는 것은, 폭염으로 인한 피해액보다는 피해를 받는 대상의 범위가 대중적이라는 점이 영향을 미친 것으로 해석된다.

네트워크 분석을 통해 도출된 영향 유형들은 관련 기사를 참고하여 상호간의 관계를 파악하기 용이하도록 영향 관계도를 통해 정리하였다. 영향 유형도가 위험기상이 미치는 영향의 종류와 그 영향 정도에 대한

정보를 제공하였다면, 피해의 원인과 결과, 또 다시 결과가 원인으로서 작용하는 일련의 흐름은 영향 관계도를 통해 파악할 수 있었다. 도출된 영향들은 실제로 많은 피해를 주고 있음을 과거 피해 사례를 통해 확인할 수 있었으며, 이렇듯 영향간의 관계를 파악하는 것은 연쇄적으로 발생하는 2·3차 피해에 선제적으로 대비할 수 있다는 점에서 큰 가치가 있다.

최신 기법인 텍스트 마이닝과 네트워크 분석은 위험기상에 의한 영향을 분석하는데 효과적인 도구임을 알 수 있었지만, 분석 과정에서 몇 가지 한계점을 찾을 수 있었다. 첫 번째는 연구자의 주관성을 완전히 배제하지 못한다는 것이다. 언어 분석의 특성상 분석에서 제외할 불용어를 선정할 때 주관적인 판단이 필요하다. 예를 들어, 유사한 의미를 가지지만 다르게 표현되는 단어들(‘고령자-노인’, ‘무더위-불볕더위’ 등), 특별한 의미를 도출할 수 없는 일반적인 단어들(‘여름’, ‘수축’, ‘남쪽’, ‘피해’ 등), 의미는 해석 가능하나 본 분석에서는 의미가 없는 단어들(‘보양식’, ‘아이스크림’ 등)은 단어 목록을 직접 확인하여 정리하여야 하기 때문에, 분석 결과에 영향을 미칠 가능성이 있다. 모든 연구에 동일하게 적용 가능한 불용어 사전은 구축할 수 없지만, 기상 관련 불용어, 유사어·동의어 등의 사전을 구축한다면, 앞으로의 분석에서 보다 신뢰할 수 있는 객관적인 결과를 얻을 수 있을 것으로 판단된다. 텍스트 분석에서 주관성이 개입되는 것은 본 연구에서만 나타나는 한계점은 아니다. Won and Kim (2014)에서도 역시 분석과정에서 주관성을 감소시키는 것을 최우선으로 개선할 점으로 꼽았다. 본 연구에서는 뉴스 기사를 선별하는 과정에서 연구자의 주관성이 개입되는 것을 우려하였고, 추후 주관성을 배제하고 객관적인 방법을 도입한다면 보다 더 설득력 있는 결과를 얻을 수 있을 것으로 판단된다.

두 번째는, 뉴스 기사 분석을 통해서만 생활과 밀접한 영향 유형을 분석하기 어렵다는 점이다. 인명 피해나 재산 피해 사례와는 달리, 피해 정도가 적은 ‘불편’ 사례는 쉽게 기사화되지 않기 때문에 뉴스 기사를 통한 분석은 제한적이다. 만약, 트위터, 페이스북, 인스타그램과 같은 SNS (Social Network Service) 상의 글을 활용한다면, 피해 정도가 크지 않은 생활 밀접한 영향을 분석하는데 도움이 될 것으로 보인다.

앞서 살펴보았듯이 최신의 웹 데이터를 이용한 텍스트 마이닝은 기존의 자료 수집 방법의 단점을 보완하고, 이전에 없던 새로운 유형의 자료를 생산해낼 수 있는 활용 가치가 높은 분석 방법이다. 국내·외적으로 위험기상의 사회·경제적 영향을 고려한 영향예보의 필요성이 확산되고 있는 상황에서, 텍스트 마이닝 기법은 위험기상 영향 데이터베이스를 구축하기 위한 매우 유용한 도구로서 사용될 것이다. 한계를 보완하고,

기상 관련 텍스트 분석에 최적화된 분석 알고리즘을 개발해 나간다면, 기상 데이터베이스 구축의 핵심적인 기술로 자리 잡을 것으로 기대한다.

## 감사의 글

이 연구는 기상청 영향예보추진팀 『지역특화 영향예보 서비스 고도화』(KMA2018-00723) 과제와 국립기상과학원 「재해기상 감시·분석·예측기술 개발 및 활용연구」(KMA2018-00123)의 지원으로 수행되었습니다.

## REFERENCES

- Bastian, M., S. Heymann, and M. Jacomy, 2009: Gephi: An open source software for exploring and manipulating networks. *Proc., In Third international AAAI conference on weblogs and social media*, San Jose, California, the Association for the Advancement of Artificial Intelligence, 2 pp.
- Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, 2008: Fast unfolding of communities in large networks. *J. Stat. Mech.-Theory E.*, **10**, P10008.
- Chae, Y., K. Cho, S. Lee, H. Jeon, Y. Chung, J. Lee, H. Park, and D.-K. Yoon, 2016: An analysis of the multiple impacts and policy networks of an extreme flood event in a metropolitan area. KEI 2016-14, 162 pp (in Korean).
- GTC, 2015: *Big data analysis of disaster caused by climate change*. Green Technology Center, 129 pp (in Korean).
- Gupta, V. and G. S. Lehal, 2009: A survey of text mining techniques and applications. *J. Emerging Technologies in Web Intelligence*, **1**, 60-76.
- Hagberg, A. A., D. A. Schult, and P. J. Swart, 2008: Exploring network structure, dynamics, and function using NetworkX. *Proc. The 7th Python in Science Conference (SciPy2008)*, 11-15.
- Jacomy, M., T. Venturini, S. Heymann, and M. Bastian, 2014: ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, **9**, e98679, doi:10.1371/journal.pone.0098679.
- Kim, D.-W., J.-H. Chung, J.-S. Lee, and J.-S. Lee, 2014: Characteristics of heat wave mortality in Korea. *Atmosphere*, **24**, 225-234, doi:10.14191/Atmos.2014.24.2.225 (in Korean with English abstract).
- KCDC, 2016: Annual Report on the Notified Patients with Heat-related illness in Korea. Korea Centers for Disease Control and Prevention, 51 pp [Available online

- at <http://www.nih.go.kr/contents.es?mid=a20304010700>] (in Korean).
- MPSS, 2016: Statistical Yearbook of Natural Disaster 2015. Ministry of Public Safety and Security, 184 pp (in Korean).
- Mittermayer, M.-A., 2004: Forecasting Intraday Stock Price Trends with Text Mining Techniques. *Proc. The 37th Hawaii International Conference on Social Systems*, Hawaii, IEEE, 10 pp, doi:10.1109/HICSS.2004.1265201.
- Nassirtoussi, A. K., S. Aghabozorgi, T. Y. Wah, and D. C. L. Ngo, 2014: Text mining for market prediction: a systematic review. *Expert Syst. Appl.*, **41**, 7653-7670, doi:10.1016/j.eswa.2014.06.009.
- Newman, M. E. J., 2006: Modularity and community structure in networks. *Proc. The National Academy of Sciences*, **103**, 8577-8582, doi:10.1073/pnas.0601602103.
- NIMS, 2011: Report on climate change scenario 2011. National Institute of Meteorological Sciences, 117 pp (in Korean).
- MOIS, 2017: Announcement of government-wide heat wave measures 2017. Ministry of the Interior and Safety, 15 pp (in Korean).
- Park, E. L., and S. Cho, 2014: KoNLPy: Korean natural language processing in Python. *Proc. The 26th Annual Conference on Human & Cognitive Language Technology*, SIGHCLT, 133-136 (in Korean).
- Park, S. B., 2012: Algal blooms hit South Korean rivers. *Nature*, doi:10.1038/nature.2012.11221.
- Rickman, T. A., and R. M. Cosenza, 2007: The changing digital dynamics of multichannel marketing: The feasibility of the weblog: text mining approach for fast fashion trending. *J. Fashion Marketing and Management*, **11**, 604-621, doi:10.1108/13612020710824634.
- Sun, H., C. Lim, and Y. S. Lee, 2017: Analysis of the yearbook from the Korea Meteorological Administration using a text-mining algorithm. *The Korean Journal of Applied Statistics*, **30**, 603-613, doi:10.5351/KJAS.2017.30.4.603 (in Korean with English abstract).
- Won, J.-Y., and D.-G. Kim, 2014: Deduction of social risk issues using text mining. *J. Safety and Crisis Management*, **10**, 33-52 (in Korean with English abstract).
- WMO, 2015: WMO Guidelines on Multi-Hazard Impact-based Forecast and Warning Services. World Meteorological Organization, 23 pp.