

재난 사건별 이슈 생존 주기 유형 분석

유호선*, 김현진**, 오효정***

Analysis of Issue Life Cycle Types of Disaster Events

Hosuon Yoo*, Hyeon-Jin Kim**, and Hyo-Jung Oh***

이 논문은 2018년도 전북대학교 연구기반 조성비 지원에 의하여 연구되었음.
본 연구는 2017년 한국연구재단의 지원을 받아 수행된 연구임(NRF-2016R1A2B1008000)

요 약

본 연구는 소셜미디어에서 나타나는 재난·재해 사건의 이슈 생존 주기 유형을 분석하고자 한다. 분석을 위해 재난연보를 기반으로 최근 10년간 발생한 대형 사회재난 12건과 자연재해 6건을 선정하여 각각의 재난이 언급된 뉴스와 트윗을 추출하였다. 이를 토대로 재난 사건별 생존 기간을 계산하였고, 주별 또는 월별로 미디어별 생산량을 도식화하였다. 이를 통해 재난 사건별 이슈 생존 주기의 유형을 두 가지로 분류하였다. 첫 번째 돌발형 생산 유형은 별도의 전조 징후 없이 재난 발생과 동시에 이슈의 생산이 폭발적으로 일어났고, 규칙적으로 감소하는 양상을 보였다. 두 번째 유형은 주기적 생산 유형으로 태풍과 같은 자연재해나 신종 인플루엔자와 같은 감염병 재난에 해당하는 것으로, 매해 주기적인 생산 빈도를 나타내는 것으로 파악되었다. 이러한 특성을 반영한 회귀 트리 모델을 활용하여 생존 기간 영향 요인을 추출하였고, 해당 알고리즘을 활용하여 생존 기간을 예측 및 검증하였다.

Abstract

This study aims to analyze the types of issue life cycles of disasters in social media. Based on the national disasters report during last 10 years, twelve social disasters and six natural disasters were selected as target major issues. News and tweets related to these disasters were automatically collected and the survival period of disaster events was calculated. The media production amount was plotted weekly or monthly. This study classified the types of issue life cycles by disaster events into two categories: First, an unexpected production type showed an explosive occurrence of the issue at the same time as the disaster occurred, and it showed a regular decreasing pattern. The second is a periodic production type which corresponds to natural disasters such as typhoons or infectious diseases such as swine flue. A regression tree model was built by reflecting the factors of the issue survival time and the survival period was predicted and verified using the algorithm.

Keywords

social disaster, natural disaster, issue attention cycle, social media, twitter

* 전북대학교 기록관리학과

- ORCID: <https://orcid.org/0000-0002-0745-1707>

** 한국전자통신연구원 스마트콘텐츠연구실

- ORCID: <https://orcid.org/0000-0001-9557-766X>

*** 전북대학교 기록관리학과 교신수

문화융복합아카이빙 연구소 (교신저자)

- ORCID: <https://orcid.org/0000-0001-8067-2832>

· Received: Jan. 17, 2018, Revised: Feb. 09, 2018, Accepted: Feb. 12, 2018

· Corresponding Author: Hyo-Jung Oh

Graduate School of Archives and Records Management, Chonbuk National

University, 567 Baekje-daero, Deokjin-gu, Jeonju, Jeollabuk-do, Korea (54896)

Tel.: +82-63-270-3208, Email: ohj@jbnu.ac.kr

1. 서 론

우리나라는 2014년 ‘세월호 침몰’이라는 대참사를 겪었고, 우리 국민의 고통과 슬픔이 채 가시지 않은 채 2016년 ‘경주지진’과 2017년 ‘포항지진’을 겪었다. 유례없는 재난·재해 사건들이 연달아 발생함에 따라 우리 국민의 고통과 슬픔은 가중되었고, 이에 따라 우리나라 국민들의 재난안전에 대한 기대는 매우 높아졌다. 또한 국민 안전 보장의 측면에서 재난 사건의 경우 그 특성상 단 한 건의 사고로도 큰 인명피해와 재산피해를 동반할 수 있기 때문에 효과적인 재난관리를 위한 대책 마련의 필요성이 제기되고 있다[1].

재난의 발생빈도와 이로 인한 피해 규모는 날이 커지고 있고[2], 또한, 일련의 재난 발생 현황을 살펴보면 비슷한 유형의 재난이 때와 장소를 달리 하여 반복적으로 발생했다. 현대의 재난 발생은 다양한 원인들을 포함하고 있지만, Kim and Choi는 재난의 원인들이 개선되지 못하고 반복적으로 일어나는 원인을 재난 이슈의 탄생과 죽음이라는 관점에서 접근하였다. 일반적으로 대형 재난 사건이 발생하면 미디어가 주도적으로 정보를 제공하고 여론을 형성하게 되는데, 재난 보도 행태에 관한 기존 연구들은 언론의 단발적 보도행태와 냄비근성을 지적해왔다[3].

한편 현대 사회에서 이슈와 여론의 형성은 소셜 미디어를 통해 더욱 활발히 일어나고 있다. 소셜 미디어는 재난 현장의 소식을 언론보다 빨리 전달하고, 뿐만 아니라 재난으로 인한 정신적 공황과 아픔을 함께 나누는 공감의 도구로써까지 그 역할을 확장해 나가고 있다[4]. 이 중 특히 트위터는 여러 소셜 매체 중에서도 불특정 다수의 사람들에게 신속하게 전파 가능한 특성으로 인해 과거 여러 재난 발생사례에서 재난 정보의 전파에 많이 활용되어 왔다. 이에 재난 이슈에 관한 여론을 조사하기 위해 매달 재난 트윗을 분석하거나 재난 관리에 활용하는 방안에 대한 다양한 연구가 활발히 진행 중이고 소셜 빅데이터 분석의 필요성이 제기되고 있다[5].

본 연구는 소셜미디어상에서 사회재난 이슈의 생존 주기를 분석하고, 요인을 도출한 Yoo, Myung and Oh[6]의 후속연구로써, 그 범위를 자연재해까지

확장하였다. 국민의 사회적 관심도를 확인할 수 있는 지표중의 하나인 소셜미디어를 활용하여 재난 사건별로 그 생존 주기의 양상을 파악하고자 하였다. 이를 위해 재난 사건별로 유형을 분류하였고, 기계학습 도구를 활용하여 각 재난의 생존 주기에 공통적으로 영향을 미치는 지표를 가늠하였다.

II. 관련 연구

Downs는 이슈가 사회적 관심을 받아 생성과 절정, 쇠퇴에 이르기까지를 일련의 단계를 제시하였고, 많은 후속 연구들이 언론 보도량을 측정하여 이슈 생존 주기를 밝혀왔다[7][8]. 앞선 연구들은 주로 기사 수를 분석하여 운영함수(Spline function)로 계량화하여 생존 주기를 분석하였다. 이 방법론은 일정 기간 동안 증가와 감소를 반복하는 이슈를 대상으로 사회적 관심의 흐름을 분석하는 데는 유용하지만 재난과 같이 잠복기나 성숙기가 없다가 사건 발생 이후 폭발적으로 증가하는 유형을 가진 이슈 생존 주기의 분석에 적용하기는 한계가 있었다[9].

이를 해결하기 위해 Kim and Choi는 별도의 생존율 공식을 차용하여 재난 이슈의 생존 주기를 분석하였다. 구체적으로 재난 이슈별 언론 매체의 기사 수를 계량하여 재난 이슈 생존 주기를 분석하였다. 분석 결과를 통해 재난 이슈의 생존 주기는 사건 발생 이후 폭발적으로 증폭하며 4~5주 이후에 급속하게 감소하는 패턴을 보인다는 점과 자연재난 이슈에 비해 사회재난 이슈의 생존 기간이 길다는 특성을 파악하였다[3].

Yoo, Myung and Oh는 소셜미디어에서 사회재난의 이슈 생존 주기를 분석하였는데, 생존 기간을 계산하여 생존 기간에 영향을 미치는 요소들을 도출해내었다. 구체적으로 최근 10년간 발생한 대형 사회재난을 대상으로 트윗을 수집하였고, 계산한 생존 기간과 각 요인들 간의 상관관계 분석을 활용하여 인명 피해 규모와 3.4주 트윗 발생량이 전체 트윗 생존 기간에 큰 영향을 미친다는 점을 밝혔다[6].

본 연구는 연구 범위 측면에서 뉴스에 보도된 사회재난에 한정되어 있던 기존 연구를 확장하여 자연재해와 사회재난을 아울러 각 재난 사건별로 소셜미디어에서의 생존 특성을 파악하였고, 이를 토대

로 유형을 분류하고자 한다. 또한 연구 방법 측면에서 기계학습을 도입하여 각 생존 특성의 공통점을 파악하고, 통계적으로 검증하고자 했다는 점에서 기존 연구들과 차별성을 가진다.

III. 재난 이슈 생존 주기 분석

3.1 연구방법

본 연구에서는 재난 사건별 이슈 생성 추이를 비교·분석하기 위해 다음과 같은 방법으로 연구를 진행하였다. 첫째, 분석대상을 구체화하기 위해 재난 사건을 선정하였다. 사회재난과 자연재해로 나누어 각각 행정안전부에서 발행한 2016 재난연감과 2016 재해연보를 활용하여 선정하였다. 구체적으로 이슈화가 많이 된 재난을 우선적으로 선별하기 위해 재난안전 대책본부가 운영된 재난·재해를 선정하였다.

둘째, 소셜 데이터 수집은 재난 이슈의 분석 및 오픈 API를 통해 데이터 수집 또한 용이한 트위터(twitter)를 선정하여 자동 수집하였다. 수집기간은 사회재난의 경우 각 재난 사건에 대해서 재난안전 대책본부가 운영된 시점부터 2016년 12월 31일까지로 설정하였고, 자연재해의 경우 전조징후를 파악하기 위하여 대책본부가 운영되기 한 달 전부터 수집하였다. 분석의 신뢰성을 높이기 위하여 관련 트윗 수가 500이하인 재난은 제외하였다. 한편 ‘세월호 침몰사고’는 재난 자체의 특성 보다는 정치·사회적인 영향으로 국민의 관심이 유지되고 있는 현안으로 특수한 경우에 해당되므로 분석대상에서 제외하였다. 또한 자연재해의 경우에는 트윗을 수집하기 위한 키워드 선정이 용이한 태풍, 지진을 중심으로 수집하였으며 호우, 대설 등은 제외하였다.

셋째, 수집된 트윗을 토대로 생존 기간 현황을 파악하였다. Kim and Choi의 생존율 공식을 차용하여 매 주별로 계산했고, 생존율이 10%가 되는 시점까지 걸리는 기간을 생존 기간으로 보았다[3]. 구체적으로 이슈 생존율은 식 (1)과 같이 계산하고, 생존율 R_i 는 해당 주까지의 누적 이슈 생산량 $\sum_{k=1}^i a_k$ 를 전체 이슈 생산량 N_a 로 나눈 것을 의미한다.

$$R_i = \left(1 - \frac{\sum_{k=1}^i a_k}{N_a}\right) \times 100(\%), i = 1 \dots n \quad (1)$$

넷째, 트윗 생성 추이를 도식화하여 그 양상을 생존 모형에 따라 분류하여 재난 유형별 생존 특성을 파악하였다. 또한 트윗 생존 기간을 종속변인으로 하여 다양한 독립변인들과의 의사결정나무를 제작함으로써 생존 기간에 공통적으로 영향을 미치는 요인들을 탐색하였다. 의사 결정 나무(Decision Tree)란 데이터 마이닝을 위한 기계학습 중 하나로 의사 결정 규칙을 나무 형태로 도식화하여 자료의 분류 또는 예측을 수행하는 분석방법이다. 이 방법은 분석 과정이 나무 구조로 표현되므로 분석 과정이 직관적이고 결과는 이해하기 쉽다는 장점이 있다[10]. 의사 결정 나무의 다양한 알고리즘 중 본 연구는 CART(Classification and Regression Trees)에 포함된 회귀 트리를 사용하였다. 이 트리는 각 독립변수로 구성되는 앞에 도달하는 자료의 평균값을 기반으로 하여 이를 이분화 하는 과정을 반복하여, 결과적으로 각 분류별 표준편차가 최소화되도록 트리를 결정한다. 이를 위해 본 연구의 실제 데이터 분석 과정에서 CART의 R 구현 패키지인 RPART를 사용하여 회귀 트리를 구축하였다.

3.2 사회재난 생존 주기

수집된 사회재난의 트윗 생산 현황은 표 1과 같다. 최근 10년간 발생한 대형 사회재난은 총 12건으로, 트윗 생산량은 ‘메르스’, ‘신종 인플루엔자’, ‘마우나 오션리조트 붕괴사고’ 순으로 나타났고, 생존 기간은 ‘신종인플루엔자’, ‘마우나 오션리조트 붕괴사고’, ‘청루15호 침몰사고’ 순으로 나타났다. 그림 1은 구체적으로 사회재난 이슈의 트윗의 생산 추이를 살펴보기 위해 1년(52주)을 기준으로 주별 생산량을 정규화 하여 도식화한 것이다. 대부분의 사회재난 이슈들은 최초 재난 발생 시 폭발적으로 트윗이 생산되었고, 이후 급격하게 생산 빈도가 줄어드는 것을 확인할 수 있었다. 하지만 일부 재난 사건들은 오랜 기간이 지난 후에도 이슈가 다시 활발하게 생기는 재점화 현상을 확인할 수 있었다.

표 1. 사회재난 이슈 현황

Table 1. Status of social disaster issues

Disasters	Period	# of Issues			Survival Period	
		Twitter	News	Ratio	Twitter	News
Novel swine-origin influenza A (신종인플루엔자)	09.11.4.~09.12.10.	31728	48010	0.66	291	292
Yongheung fire forest (용흥지구 산불피해)	13.3.9.~13.3.10.	2516	3774	0.67	1	156
Cargo shinking (화물선 청루15호, 파나마침몰)	13.10.15.~13.10.30.	165	761	0.22	22	67
Gymnasuim crash (경주 마우나오션리조트 체육관 붕괴사고)	14.2.17.~14.2.21.	4416	10790	0.41	24	34
Subway station srach (상왕십리역 열차추돌사고)	14.5.2.	2625	4053	0.65	4	5
Fire Accident at the Goyang (고양중합터미널 지하공사장 화재사고)	14.5.26.	1534	3970	0.39	16	52
Fire Accident at the Jangsung Convalesent hospital (장성 호사랑요양병원 화재사고)	14.5.28.	823	5584	0.15	11	52
Fire Accident at the Damyang Pension (담양대덕 펜션 화재사고)	14.11.15.	576	2592	0.22	1	14
Deep-sea fishing vessel Oryong501 Sinking (원양어선 제501오룡호 침몰사고)	14.12.1.	2283	4589	0.50	9	6
Fire Accident at the Ejeongbu-si Apartment (의정부시 대봉그린아파트 화재)	15.1.10.	1091	7323	0.15	2	39
Mers (메르스)	15.5.28.~15.7.28.	280598	382791	0.71	15	32
Overtum accident of Fishing Boat (늑시어선 돌고래호 전복 사고)	15.9.5.~15.9.30.	2325	7377	0.32	6	3

* source: Yoo, Myung and Oh[6]

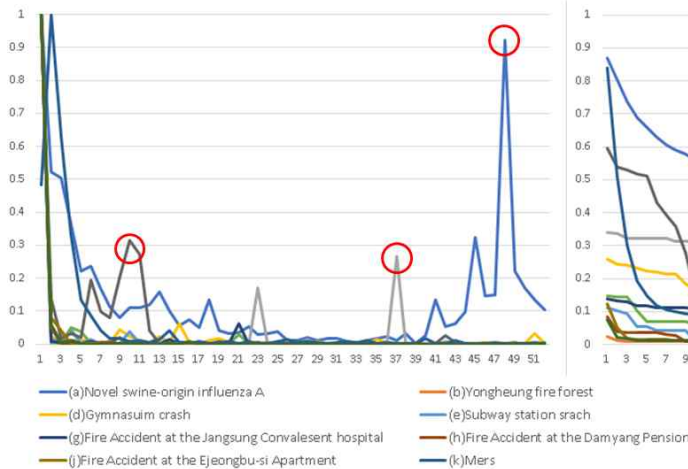


그림 1. 사회재난 트윗 빈도 정규화 그래프
Fig. 1. Tweet frequency normalization graph

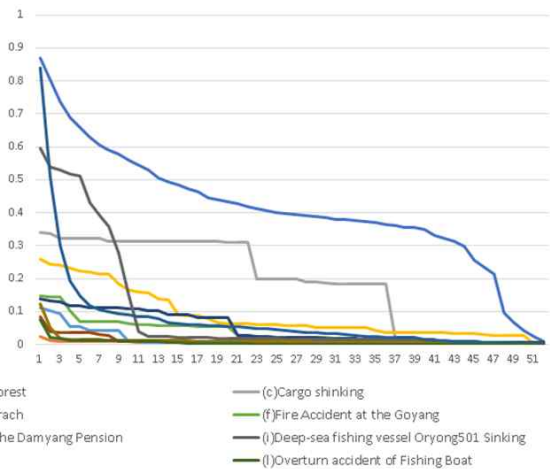


그림 2. 사회재난 이슈 생존율 그래프
Fig. 2. Social disaster issue survival rate graph

그림 2는 이슈 생존율 그래프로 이슈의 생존율은 재난 발생 이후 점차 줄어들어 생존율이 0.1 이하가 되는 시점까지를 생존 기간으로 볼 수 있다. 여기서 그래프의 곡선이 완만할수록 생존 기간이 길게 유지되었다고 볼 수 있는데 위 그래프에서는 ‘신종인플루엔자’, ‘청루15호’의 생존율이 오랫동안 꾸준히 지속되는 것으로 나타난다. 실제로 ‘신종인플루엔

자’의 생존 기간은 291주로 모든 사회재난 중에서 가장 길었고, ‘청루 15호’의 생존 기간 역시 22주로 높게 나타났다. 또한 해당 재난들은 그림 1의 표기된 부분과 같이 이슈의 재점화가 큰 폭으로 일어났고, 이와 같은 이슈의 재점화 양상이 이후의 생존 기간에 불규칙하게 영향을 미칠 수 있다는 점을 파악할 수 있었다.

3.3 자연재해 생존 주기

표 2는 자연재해의 트윗 생산 현황과 생존 기간을 나타낸 것이다. 자연재해의 트윗 생존 기간은 사건별로 그 차이가 사회재난에 비해서 더 극심하게 나타났다. 태풍 ‘곤파스’의 경우 1년 이상의 생존 기간을 가졌지만 나머지 재해 사건들은 1달 미만의 생존 기간을 가지는 경우가 대부분이었고, 9.12 지진의 경우에만 높은 생존 기간이 나타났다. 태풍 ‘텐빈·볼라벤’의 경우 트윗 생산량은 최대였지만 생존 기간은 짧았다. 이는 사회재난 뿐만 아니라 자연재해의 경우에도 전체 트윗 생산량이 트윗 생존 기간에 큰 영향을 미치지 못한다는 점을 시사했다.

태풍 ‘곤파스’의 경우 전체 생산량은 낮았지만 꾸준히 생산되었고, ‘9.12’ 지진의 경우 이슈가 크게 된 만큼 상대적으로 수집기간이 적었음에도 불구하고 트윗 생산량과 생존 기간 모두 길게 나타났다. 반면 뉴스 생존 기간의 경우 모든 재난이 40주 이상으로 나타남으로써 뉴스에서 지속적으로 언급되고 있었다. 이는 사회재난과 마찬가지로 자연재해 역시 트위터와 뉴스미디어의 생존 기간은 큰 차이를 보임을 의미하며, 뉴스에 비해 트윗의 생존 기간은 차이가 크고 평균 생존 기간은 짧은 편인 것으로 분석되었다. 또한 이렇듯 재난 이슈들은 지속적으로 뉴스에 언급된다고 하더라도 트윗 생산량과 생존기간에 직접적인 영향은 미치지 못한다는 점을 확인할 수 있었다.

표 2. 자연재해 이슈 현황

Table 2. Status of social natural issues

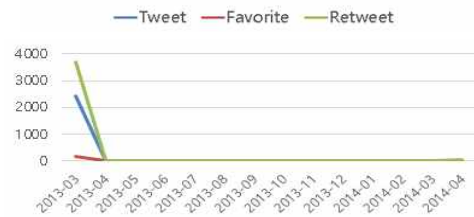
Disasters	Period	# of Issues	Survival Period
Typhoon KOMPASU (곤파스)	10.9.1. ~ 10.9.3.	5330	67
Typhoon MUIFA (무이파)	11.8.6. ~ 11.8.10.	2621	2
Typhoon TENBIN, BOLAVEN(텐빈, 볼라벤)	12.8.25. ~ 12.8.30.	25910	3
Typhoon SANBA(산바)	12.9.16. ~ 12.9.17.	1945	2
9.12 Earthquake	16.9.12.	15186	21
Typhoon Chaba(차바)	16.10.3. ~ 16.10.6.	3761	3

IV. 재난별 생존 주기 유형 분석

4.1 사회재난

4.1.1 돌발형 생산 유형

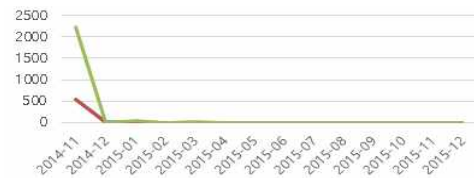
앞서 살펴본 대로 사회재난은 특별한 전조 징후 없이 이슈가 폭발적으로 생산되었다가 급격히 줄어드는 패턴을 보였고, 이와 같은 이슈주기 유형을 Lee, Shin and Kang은 “돌발형” 패턴이라고 정의하였다[11]. 대표적으로 위의 그림 3의 (a) 포항 산불, (b) 의정부 아파트 화재, (c) 담양 펜션 화재, (d) 돌고래호 전복 사고 등은 이러한 유형을 매우 정확하게 따르고 있다.



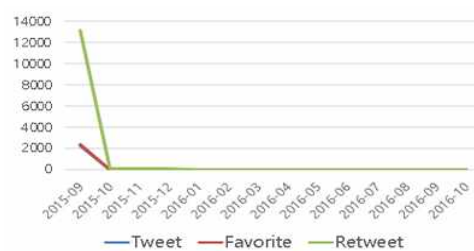
(a) Yongheong fire forest



(b) Fire accident at the Ejeongbu-si apartment



(c) Fire accident at the Damyang pension

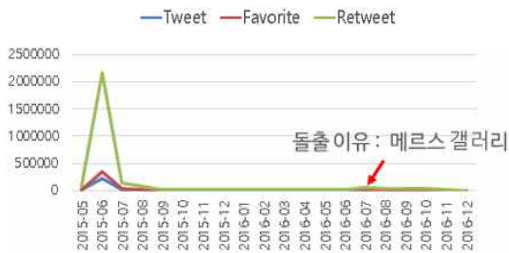


(d) Overturn accident of fishing boat

그림 3. 돌발형 생산 유형 현황
Fig. 3. Status of unexpected production type

그림 3은 이러한 유형의 재난 사건들의 월별 생산 추이와 생존 기간을 나타낸 것이다. 이들 재난은 재난 발생 이후 급격히 이슈 생산이 줄어들고 이후에 관련 이슈가 생산되지 않기 때문에 공통적으로 매우 짧은 생존 기간을 가진다. 또한 이와 같은 돌발형 생산 유형은 이슈의 생산이 규칙적으로 감소하는 양상을 보이기 때문에 해당 유형에 대한 분석을 통해 재난 사건별 생존 기간 회귀식을 도출하는 것이 가능하여 예측 모델을 설계할 수 있다.

4.1.2 예외적 돌출 유형



(a) Mers



(b) Fire accident at the Jangsung convalescent hospital



(c) Deep-sea fishing vessel Oryong501 sinking



(d) Fire accident at the Goyang

그림 4. 예외적 돌출 현황

Fig. 4. Status of exceptional extrusion

일부 재난 이슈는 재난 발생 이후에도 이슈가 재점화되는 양상을 보이는데 이는 생존 기간의 연장에 영향을 미친다. 이러한 돌출 유형은 주로 관련 이슈, 여론에 영향을 받아 형성되는데 특히 뉴스 보도 자료 중 일부는 국민의 여론을 자극하여 돌출을 심화시킬 수 있다. 오른쪽 그림 4의 (a) 장성 요양병원 화재와 고양종합터미널 화재사고의 경우 당해 10월에 있었던 2014년 유독 재난이 많이 발생한 것에 대한 보도와 연관되어 돌출되었고, 메르스의 경우에는 2016년 7월 메르스 갤러리는 온라인 커뮤니티 게시판이 형성됨에 따라 관련 이슈가 재점화되었다. 또한 오룡호의 경우 정부가 유족들의 분향소 설치 요구를 거절하고 눈 내리는 거리로 내몰렸다는 뉴스 보도와 연관하여 돌출되었다. 이를 종합하면 이슈의 돌출은 관련 이슈와 여론에 영향을 받아 일어날 수 있고, 여론 형성 과정에서 모든 보도 자료에 영향을 받는 것은 아니지만 특히 국민의 정서를 자극하여 공감대 형성이 가능한 보도들과는 연관하여 돌출되는 양상을 보인다.

4.2 자연재해

자연재해의 이슈 생산의 특징은 크게 두 가지가 있다. 첫 번째로 이슈의 전조징후가 뚜렷하게 포착된다는 점이다. 자연재해의 이슈 생산은 사회재난과는 달리 재난 발생 이전 단계부터 관련 이슈가 서서히 증가하는 전조 단계를 포착할 수 있다. 그림 5에서 재난 발생 시점 이전의 증가하는 단계를 확인할 수 있다. 두 번째로 자연재해의 생산 유형은 주기적인 생산 유형을 보인다는 것이다. 여기서 주기적이라는 의미는 이슈의 돌출이 불규칙하게 일어나는 것이 아니고 시기적으로 규칙성을 가지고 일어난다는 의미이다. 이는 특히 태풍과 같이 주기적으로 일정한 기간 동안에만 일어나는 자연재해에 두드러지게 나타난다. 예를 들어 그림 5의 태풍들의 이슈 생산 추이를 확인하면 매년 8~9월 사이에 관련 이슈가 재점화 되는 양상을 확인할 수 있다. 또한 태풍 덴빈, 볼라벤, 산바가 모두 발생한 2012년에 특히 관련 이슈는 크게 돌출하는 것을 볼 수 있는데, 매년 여름에 재점화 되는 동시에 큰 태풍들이 올 때마다 이전에 발생했던 태풍 피해에 대한 언급

이 꾸준히 이루어지고 있다는 점을 확인할 수 있다. 또한 신종인플루엔자의 경우 사회재난 임에도 불구하고 주기적인 특성을 지니고 있는데, 매년 겨울에 관련 이슈의 생산량이 폭발적으로 증가함으로써 오랜 생존 기간을 보이고 있었다. 이와 같이 자연재해 뿐만 아니라 사회재난 중에서도 “감염병” 유형은 주기적인 생산패턴을 보이는 경우가 많은데, 이러한 패턴은 AI, 구제역과 같은 재난의 경우에도 동일하게 나타난다.

4.3 생존 주기 모형 분석 결과

재난 이슈의 생존 주기는 재난 특성에 따라 사회재난은 돌발형 유형으로, 자연재해는 주기적 생산 유형을 중심으로 나타났다. 또한 주기적 생산 유형을 가지는 자연재해와는 달리 규칙적인 생존율의 감소를 보이는 사회재난의 경우 생존 요인과 생존 기간간의 상관관계를 나타내고 있었다. 기존 연구에서 이에 대한 상관관계는 피해규모 요인과 0.998, 3-4주 트릿과 각각 0.930, 0.982의 상관계수를 가졌으며, 0.01 수준(양쪽)에서 유의하였다[6]. 해당 연구 결과에서 특히 사회재난의 생존 기간은 몇몇 요인에 강력한 영향을 받으며 비례관계로 나타날 수 있음을 알 수 있었다.

본 연구에서는 이를 토대로 구체적으로 어떤 요인들이 생존 기간에 관련이 있는지 도출하기 위해 기계학습을 통한 회귀 트리 모델을 제작하였다. 구체적으로는 주기적인 특성을 지닌 신종 인플루엔자를 제외한 나머지 사회재난을 대상으로 생존 기간에 영향을 줄 수 있는 요인들을 광범위하게 데이터셋으로 사용하고, 생존 기간과의 관계를 밝혔다. 사용한 요인은 Yoo, Myung and Oh의 선행연구를 참고하여 피해규모와 트릿 생산량을 중심으로 사용하기 위하여 인명피해 지표로 사망, 부상, 실종자 수와 각 주별 트릿, 리트릿, 좋아요 생산량, 각 주별 트릿 생존율을 주요 지표로 삼았다[6].

그림 6은 이를 도식화한 것으로 생존 기간에 가장 우선적으로 영향을 미치는 요인은 일주 생존율로 나타났다. 이는 생존 기간이 짧은 트릿들은 최초 일주 일안에 대부분의 트릿이 생산되므로 일주 생존율은 큰 수치를 나타내게 되기 때문으로, 구체적으로 일주 생존율이 0.83 이상일 경우 생존 기간은 5주 이하로 나타나게 되고, 0.83 이상일 경우 그 이상으로 나타나게 된다. 한편 일주 생존율이 0.83 이하인 생존 기간이 긴 재난들은 인명피해 수치에 따라 생존 기간이 상이하게 나타나는 것을 확인할 수 있었는데, 사망자 수가 더욱 결정력 있는 지표로 나타났다으며 다음으로 부상자 수에 영향을 받았다.

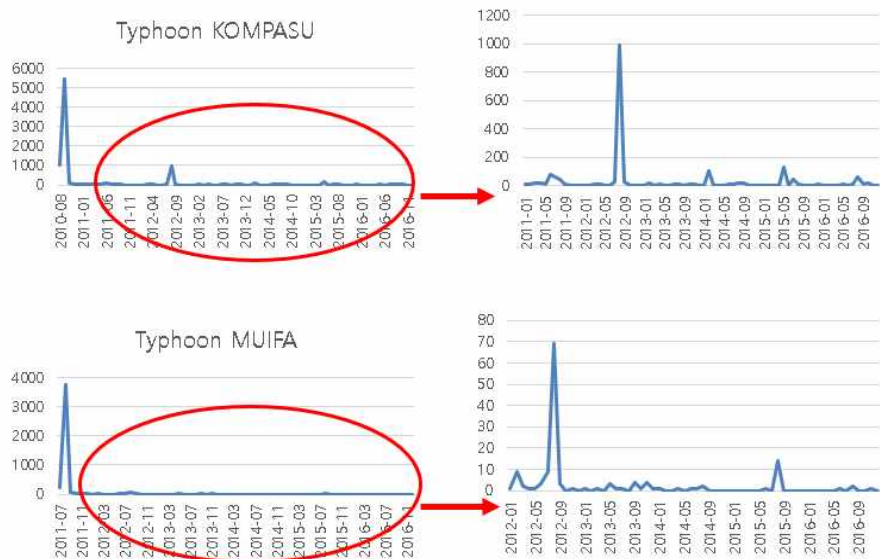


그림 5. 주기적 생산 유형 현황
Fig. 5. Status of periodic production type status

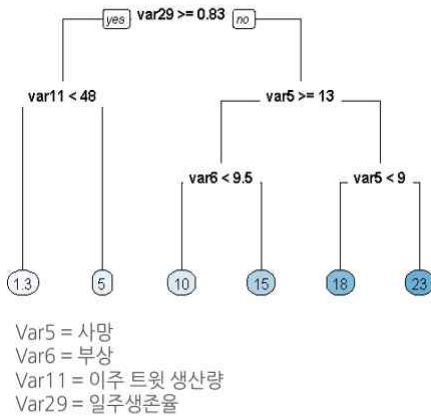


그림 6. 생존 기간 회귀 트리

Fig. 6. Survival period regression tree

표 3은 도출한 회귀 트리의 알고리즘을 적용하여 예상한 생존 기간과 실제 생존 기간을 비교한 것이다. 각 재난별로 평균 백분율 오차(MAPE)는 14%로 나타났으며, 오차가 가장 크게 나타난 돌고래호 전복사고의 경우 피해자 수가 매우 큰 사건이었음에도 트윗과 뉴스 보도상으로도 크게 이슈가 되지 않은 특이한 사건으로 보였다. 따라서 실제 재난 이슈의 생존 기간에는 위와 같은 수치적 요인들뿐만 아니라, 국민의 여론을 지속시킬 수 있는 다른 내재적인 요인들이 포함되어 있을 수 있다는 점을 의미한다.

표 3. 예상 생존 기간

Table 3. Expected survival period

Disasters	Survival Period	Expected Survival Period
Yongheung fire forest (용흥지구 산불피해)	1	1.3
Cargo sinking (화물선 청루15호, 파나마침몰)	22	23
Gymnasium crash (경주 마우나오션리조트 체육관 붕괴사고)	24	23
Subway station srach (상왕십리역 열차추돌사고)	4	5
Fire Accident at the Goyang (고양종합터미널 지하공사장 화재사고)	16	17.5
Fire Accident at the Jangsung Convalesent hospital (장성 효사랑요양병원 화재사고)	11	10
Fire Accident at the Damyang Pension (담양대덕 펜션 화재사고)	1	1.3
Deep-sea fishing vessel Oryong501 Sinking (원양어선 제501오룡호 침몰사고)	9	10
Fire Accident at the Ejeongbu-si Apartment (의정부시 대봉그린아파트 화재)	2	1.3
Mers (메르스)	15	15
Overturn accident of Fishing Boat (낚시어선 돌고래호 전복 사고)	1	5
MAE (평균 오차)		0.79 주
MAPE (평균 백분율 오차)		14%

다. 또한 해당 회귀 트리에서 분석 대상인 전체 재난 사건의 수가 많지 않아서 대조군을 설정하는데 어려움이 있었고, 이에 따라 모델의 제작과 검증이 불충분한 점은 본 연구의 한계점으로 남아있다.

V. 결론 및 제언

소셜미디어는 다른 매체보다 신속하고 확장성이 높다는 특징으로 인해 재난 발생 시 관련 상황과 정보가 대중들에게 빠르게 전파된다. 본 연구는 소셜미디어에서의 재난 이슈 언급 추이를 분석함으로써 해당 재난 이슈에 대해 여론의 생성에서부터 소멸에 이르기까지의 특성을 파악하고자 했다. 사회재난과 자연재해를 아울러 재난 이슈들이 지니는 특성을 파악하였고, 유형을 파악하여 생존 주기를 분석하였다.

연구결과에 따르면 재난 이슈는 크게 두 가지 유형으로 분류되었다. 첫 번째는 돌발형 생산 유형으로 대부분의 사회재난이 이와 같은 유형에 속하였다. 돌발형 생산 유형에 속하는 재난들은 별도의 전조 징후 없이 재난 발생과 동시에 이슈의 생산이 폭발적으로 일어났고, 규칙적으로 감소하는 양상을 보인다. 해당 유형은 피해 규모가 크거나 돌출이 크게 일어날 경우 긴 생존 기간이 나타났다.

두 번째는 주기적 생산 유형으로 태풍 등과 같은 자연재해와 감염병과 같은 매해 주기적인 생산 빈도를 나타내고 있었다. 해당 유형의 경우 재난이 발생했을 경우뿐 만 아니라 재난이 발생하지 않아도 매년 동일한 시기가 되면 해당 재난에 대한 언급이 꾸준히 나타나고 있었다.

한편 사회재난은 돌발형 유형에 따라 생산 빈도와 생존 기간이 일련의 규칙성을 띄고 나타나고 있었다. 기계학습을 활용한 생존 기간 영향 요인의 추출결과 Yoo, Myung and Oh[6]의 선행 연구 결과를 뒤따라 피해 규모와 주별 트윗 생산량이 전체 생존 기간에 큰 영향을 미치고 있었고, 이를 통해 생존 기간의 예측도 가능하였다.

본 연구는 재난 이슈에 대한 소셜미디어 여론의 생존 주기 모형을 분석하였고, 이를 통해 각 재난별 생존 특성과 공통적으로 나타나는 유형을 분류하고 예외 유형을 발견하였다. 분석 결과를 통해 재난 이슈의 돌발형 특징과 주기적 특징을 발견하였고, 특히 사회재난의 생존 기간에 영향을 미치는 요인을 추출하고 이를 통해 생존 기간을 예상할 수 있었지만 대형재난으로 정의한 연구범위의 한계로 인하여 도출한 모델 알고리즘을 폭넓게 적용하지 못했다는 한계가 있었다. 따라서 향후 새롭게 발생하는 재난에 대해서 해당 알고리즘이 충분히 적용 가능한지 검증이 필요하며, 이후 연구 대상의 범위를 넓히고 새로운 영향 요인 및 예외사항을 발견하고 세분화하는 연구가 요구된다.

References

- [1] J. L. Lee, "A Study on the Direction of Policing for Effective Disaster Prevention and Management", *Korean Security Science Review*, No. 51, pp. 315-334, Jun. 2017.
- [2] K. H. Lee, W. H. Yi, and W. J. Yang, "A Study on Risk Analysis of Social Disaster", *Journal of Korean Society of Disaster & Security*, Vol. 9, No. 2, pp. 15-21, Dec. 2016.
- [3] C. M. Kim and C. I. Choi, "An Exploration of Issue Attention Cycle of Great Disasters", *Seoul Studies*, Vol. 16, No. 4, pp. 147-162, Dec. 2015.
- [4] H. J. Kim, "A Study on the Role of the Disaster and Crisis-related Government's Social Media as the Crisis Management Communication Tool", *Journal of Practical Research in Advertising and Public Relations*, Vol. 7, No. 4, pp. 60-98, Nov. 2014.
- [5] H. J. Oh, Y. J. Bae, and B. H. Yun, "Temporal and Spatial Continuity Analysis of Issue Events", *Journal of KIIT*, Vol. 11, No 9, pp. 161-167, Sep. 2013.
- [6] H. S. Yoo, H. Myung, and H. J. Oh, "Issue Survival Cycle Analysis of Social Disasters in Social Media", *Journal of KIIT*, Vol. 15, No. 7, pp. 139- 147, Jul. 2017.
- [7] A. Downs, "Up and Down with Ecology-the Issue Attention Cycle", *Public Interests*, No. 28, pp. 38-50, Summer, 1972.
- [8] R. E. Dunlap and R. Scarce, "Environmental problems and protection", *Public Opinion Quarterly*, No. 55, pp. 651-671, Winter, 1991.
- [9] H. Y. Jang, "The Network Analysis of Public Opinion Building in Korea: Shared Editorials among Nine Major Daily Newspapers", *Korean Journal of Journalism & Communication Studies*, No. 41, pp. 108-141, Jul. 1997.
- [10] Y. H. Kim and S. Y. Lee, "A Study on the Priorities of Winning Determinants of Real Estate Development Permit Lawsuit Case Using Decision Tree Analysis Method", *Korea Real Estate Research Institute*, Vol. 27, No. 2, pp. 19-34, Jun. 2017.
- [11] J. H. Lee, D. H. Shin and S. M. Kang, "The Examination of 'Pot Journalism' through Issue-Cycle Analysis - An Application of Spline Function Analysis", *Korean Journal of Broadcasting and Telecommunication Studies*, Vol. 27, No. 4, pp. 206-250, Jul. 2013.

저자소개

유 호 선 (Ho-Suon Yoo)



2016년 2월 : 전북대학교
수학과(이학사)
2016년 3월 ~ 현재 : 전북대학교
기록관리학과 석사과정
관심분야 : 전자기록,
빅데이터정보처리

김 현 진 (Hyeon-Jin Kim)



1997년 : 부산대학교
전자계산학과(석사)
1997년 ~ 현재 : 한국전자
통신연구원 스마트콘텐츠연구실
책임연구원
관심분야 : 콘텐츠분석,
텍스트마이닝, 빅데이터정보처리

오 효 정 (Hyo-Jung Oh)



2008년 : 한국과학기술원
컴퓨터공학과(공학박사)
2000년 ~ 2015년 : 한국전자
통신연구원 지식마이닝연구실
책임연구원
2015년 ~ 현재 : 전북대학교
기록관리학과 조교수

관심분야 : 정보검색, 텍스트마이닝, 빅데이터정보처리