

Sentiment Analysis of Foot-and-Mouth Disease Using Tweet Text-Mining Technique

Heechan Chae[†] · Jonguk Lee^{**} · Yoona Choi^{***} · Daihee Park^{****} · Yongwha Chung^{*****}

ABSTRACT

Due to the FMD(foot-and-mouth disease), the domestic animal husbandry and related industries suffer enormous damage every year. Although various academic researches related to FMD are ongoing, engineering studies on the social effects of FMD are very limited. In this study, we propose a systematic methodology to analyze emotional responses of regular citizens on FMD using text mining techniques. The proposed system first collects data related to FMD from the tweets posted on Twitter, and then performs a polarity classification process using a deep-learning technique. Second, keywords are extracted from the tweet using LDA, which is one of the typical techniques of topic modeling, and a keyword network is constructed from the extracted keywords. Finally, we analyze the various social effects of regular citizens on FMD through keyword network. As a case study, we performed the emotional analysis experiment of regular citizens about FMD from July 2010 to December 2011 in Korea.

Keywords : Text Mining, Sentiment Analysis, FMD, Twitter, Deep Learning

트윗 텍스트 마이닝 기법을 이용한 구제역의 감성분석

채 희 찬[†] · 이 종 옥^{**} · 최 윤 아^{***} · 박 대 희^{****} · 정 옹 화^{*****}

요 약

구제역으로 인하여 국내 축산업계 및 관련 산업분야는 매년 막대한 피해를 입고 있다. 구제역과 관련한 다양한 학술적 연구들이 현재 진행되고는 있으나, 구제역의 발병에 따른 사회적 파급효과에 관한 공학적 분석 연구는 매우 제한적이다. 본 연구에서는 구제역에 관한 일반 시민들의 감성적 반응을 텍스트 마이닝 방법론을 사용하여 분석하는 체계적인 방법론을 제안한다. 제안하는 시스템은 먼저, 트위터에 게시된 트윗 중 구제역과 관련된 데이터를 수집한 후, 딥러닝 기법을 사용하여 극성 분류 과정을 거친다. 둘째, 토픽 모델링의 대표적인 기법 중 하나인 LDA를 활용하여 트윗으로 부터 키워드들을 추출하고, 추출된 키워드들로부터 극성별 동시출현 키워드 네트워크를 구성한다. 셋째, 키워드 네트워크를 통해 구제역의 위기단계 구간별 사회적 파급효과를 분석한다. 사례 분석으로써, 2010년 7월부터 2011년 12월까지 국내에서 발생한 구제역에 관한 일반 시민들의 감성적 변화를 분석하였다.

키워드 : 텍스트 마이닝, 감성분석, 구제역, 트위터, 딥러닝

1. 서 론

지속적으로 국내에서 발병하는 구제역은 해당 축산업계뿐

만 아니라 일반 소비자들 및 사회 전반에 큰 피해를 야기한다. 특히, 지난 2010년 발생한 구제역은 전국적으로 확산되어 총 피해액이 3조원이 넘는 막대한 피해가 발생하였다[1]. 구제역(Foot-and-Mouth Disease, FMD)은 소, 돼지, 염소, 사슴, 낙타 등 우제류 동물(발굽이 2개인 동물)의 입과 발굽 주변에 물집이 생기는 질병으로, 치사율이 5~55%에 달하는 바이러스성 가축 전염병이다. 전염성이 매우 강한 구제역 바이러스는 가축의 호흡기를 통해 감염되기 때문에 무리에서 한 마리가 감염되면 축사의 나머지 가축에게로 급속하게 감염된다.

최근 정부의 주도하에 구제역을 비롯한 가축 질병에 관한 공공 데이터의 수집과 동시에 가축전염병의 확산 방지를 위한 학술적 연구들[2-5]이 활발하게 진행 중이다. 또한 공공

※ 본 연구는 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2015RID1A3A01018731).

※ 이 논문은 2018년도 한국정보처리학회 춘계학술발표대회에서 '트윗 키워드 네트워크를 이용한 구제역의 감성분석'의 제목으로 발표된 논문을 확장한 것이다.

† 준 회 원 : 고려대학교 컴퓨터정보학과 석사과정

** 정 회 원 : 고려대학교 컴퓨터정보학과 조빙교수

*** 비 회 원 : 고려대학교 컴퓨터정보학과 학사과정

**** 정 회 원 : 고려대학교 컴퓨터정보학과 교수

***** 중신회원 : 고려대학교 컴퓨터정보학과 교수

Manuscript Received : July 6, 2018

First Revision : July 24, 2018

Accepted : August 3, 2018

* Corresponding Author : Daihee Park(dhpark@korea.ac.kr)

데이터뿐만 아니라 온라인 뉴스 데이터를 활용한 네트워크 분석[6] 등의 연구도 보고되었다. 그러나 이러한 공공 데이터와 같은 정형적 데이터나 온라인 뉴스 데이터만으로는 구제역으로 인한 피해 농가 및 일반 시민들이 느끼는 진솔한 감성적 문제를 파악하기에는 한계가 있다. 반면, Social Network Service(SNS)와 같은 온라인 매체는 정형화된 공공 데이터 베이스에서 다루지 않는 다양한 사회적 이슈들이 신속하게 전파될 뿐만 아니라 해당 문제에 관한 일반 시민들의 진솔한 감성 정보를 포함하고 있다[7-9].

본 논문에서는 대표적 SNS인 트위터(Twitter)를 대상으로 딥러닝을 통한 문서 분류와 같은 양적 연구와 키워드 네트워크 분석과 같은 질적 연구의 결합을 통하여 구제역으로 인한 일반 시민들의 감성적 정보를 공학적으로 채굴하여 분석하고자 한다. 먼저, 트위터에 게시된 트윗(Tweet) 중에서 구제역과 관련된 데이터를 수집하고, 전처리 과정을 거친다. 둘째, 최근 문서 분류 분야에서 탁월한 성능이 입증된 Long Short Term Memory(LSTM) 모델[10-11]과 Convolution Neural Network(CNN) 모델[12]의 결합 형태인 LSTM-CNN 모델[13-14]을 통한 트윗 극성(긍정 또는 부정) 분류 과정을 거친다. 셋째, 토픽 모델링의 대표적인 기법 중 하나인 Latent Dirichlet Allocation(LDA)를 활용하여 키워드들을 추출[15]하고, 동시출현 키워드 네트워크를 구성한다. 마지막으로 구성된 네트워크를 통해 극성별, 구제역의 위험구간별 사회적 파급효과를 분석한다. 사례 분석으로써, 국내에서 가장 큰 피해가 발생한 2010년 7월부터 2011년 12월까지 트윗 데이터를 대상으로 구제역으로 인한 일반 시민들의 감성적 변화를 분석한다.

본 논문의 구성은 다음과 같다. 2장에서는 가축질병 및 공학적 감성 분류에 관한 기존의 연구들을 간단하게 소개한다. 3장에서는 본 연구에서 제안하는 SNS의 키워드 네트워크 분석 시스템에 대해 자세히 설명한다. 4장에서는 제안 시스템을 활용한 사례 분석을 통해 시민들의 감성 분석 결과를 살펴보고, 마지막으로 5장에서는 본 연구의 결론을 맺는다.

2. 관련 연구

가축질병 중에서도 구제역 및 조류인플루엔자 등은 한 번 발생하면 전국적인 피해를 야기하는 매우 심각한 사회적 문제로 대두되고 있다. 이러한 문제점을 해결하기 위하여 정부 주도하에 질병 예방 및 위기 대응 방안 분석, 확산 경로 예측 등과 같은 다양한 학술적·실천적 연구들이 현재 활발히 진행되고 있다. 대표적인 연구로써, Kyung 등[4]은 Spatial On-Line Analytical Processing(SOLAP) 기술을 이용하여 방역관리의 의사결정을 지원하는 시스템을 제안하였으며, Jo 등[5]은 GPS를 이용하여 축산분뇨차량의 이동경로를 분석하는 연구를 진행하였다. 한편, Noh 등[6]은 다양한 사회적 이슈들을 신속하게 다루는 온라인 뉴스를 대상으로 구제역의 사회적 파급효과를 분석하였다. 그러나 이러한 공공 데이터와 같은 정형적 데이터나 온라인 뉴스 데이터만으로는 구제역으로 인

한 피해 농가 및 일반 시민들이 느끼는 진솔한 감성적 문제를 파악하기에는 한계가 있다.

최근 디지털 기술의 발달과 다양한 스마트 기기의 보급으로 인하여 소셜 미디어에 대한 접근성이 향상되면서 실시간으로 막대한 양의 데이터들이 생성·전파되고 있다. 또한, 빅데이터의 관리 및 분석을 위한 다양한 도구 및 방법론들이 충분히 성숙되면서, 소셜 미디어를 통한 사용자의 감성 분석에 관한 공학적 연구도 활발하게 진행되고 있다. 예로써, Rosenthal 등[16]은 최근 가장 인기 있는 이슈(예: 도널드 트럼프, iPhone 등)에 관한 다양한 입장에서의 감성분석을 수행했다. Yun 등[17]은 소셜 미디어 데이터 분석을 활용해서 ‘빅데이터’에 관한 사람들의 인식 변화를 비교·분석했으며, Seo 등[18]은 재난 관련 SNS 데이터를 대상으로 감성도 분석을 수행하였다. 한편, Oh 등[19]은 트위터의 감성 분석을 이용하여 실시간으로 사용자가 원하는 장소를 추천하는 연구를 진행하였다. 그러나 Oh 등[19]의 연구(분류 정확도 70%대)에서 보듯이, 사전 기반 방식의 감성 분류 방법은 매우 낮은 분류 정확도라는 한계점을 갖는다. 반면, Cliche[13]은 딥러닝의 CNN과 LSTM의 앙상블 기법을 사용하여 트위터의 감성 분류 시스템을 새롭게 구축하였으며, Baziotis 등[11]은 Bidirectional-LSTM 구조에 두 가지 주목 기법(Attention Mechanisms)을 적용하여 보다 높은 정확도의 감성 분류 시스템으로 발전시켰다.

본 연구에서는, 구제역과 관련된 트윗 데이터를 대상으로 극성 분류를 위한 LSTM-CNN 결합 모델을 포함한 일련의 텍스트 마이닝 기법을 통해 구제역에 대한 일반 시민들의 감성 정보를 분석하고자 한다.

3. 제안 방법

본 논문에서 제안하는 키워드 네트워크 기반의 구제역 감성 분석 시스템은 크게 데이터 수집 및 전처리 모듈, 감성 분류 모듈, 그리고 네트워크 분석 모듈로 구성되며 시스템 구조는 Fig. 1과 같다.

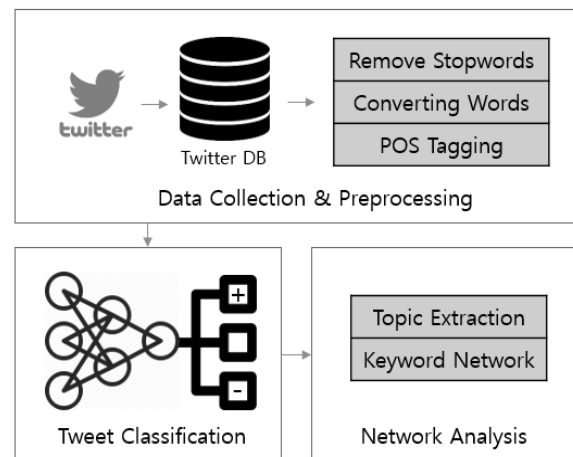


Fig. 1. System Architecture

3.1 데이터 수집 및 전처리 모듈

데이터 수집 단계에서는 크롤러(Crawler)를 이용하여 트위터에서 '구제역'이 포함된 트윗만을 수집한다. 데이터 수집 시, 시간 변화에 따른 분석 등 다양한 분석을 위하여 트윗의 게재 시간 및 내용을 함께 수집한다. 전처리 모듈에서는 트윗의 불용어 제거, 단어 변환, 형태소 분석(Parts-Of-Speech Tagging, POS Tagging)을 수행한다. 불용어 제거 과정에서는 선택된 트윗에 포함된 불용어와 @태그, #태그, URL, 광고 등을 제거한다. 본 연구에서 사용한 단어의 필터링 및 변환 규칙의 예는 Table 1과 같다.

Table 1. Word Filtering and Converting Rules

	Filtering and Converting Rules	Example of Target Words	Result Words	
1	Remove @, # tags, and URLs	<ul style="list-style-type: none"> @meonson http://bit.ly/ab #KOREA #농가 	(Remove)	
2	Word Converting	Area	충청북도	충북
		Emoticon	<ul style="list-style-type: none"> ㅠㅕㅠ ㄷ.ㄷ 	ㅠㅠ
		Other Words	가카	대통령
			각하	
때때	때문에			
멤시				

3.2 트윗 극성 분류 모듈

트윗의 극성 분류를 위해 Fig. 2와 같은 구조의 LSTM-CNN 모델[13]을 사용한다. LSTM은 시계열 데이터(Time-Series Data)와 같이 시간의 흐름에 따라 변하는 데이터를 학습하기 위한 딥러닝 모델로써, 일종의 시계열 데이터인 트윗의 텍스트 데이터를 처리하기에 적합하다[12]. 또한, CNN구조는 Convolution 연산을 통하여 데이터의 지역적 정보를 탐지하는 신경망 구조로써, 최근 텍스트 마이닝에도 적극적으로 적용되고 있다[9]. 트윗의 극성 분류를 위해 본 연구에서 사용하는 LSTM-CNN 모델은 트윗의 각 토큰(Token)에 대한 단어 임베딩(Word-Embedding)을 입력으로 LSTM layer와 CNN구조가 결합된 형태로 구성되며 최종적으로 '긍정'('+'), '중립'('0'), '부정'('-')의 label을 출력한다.

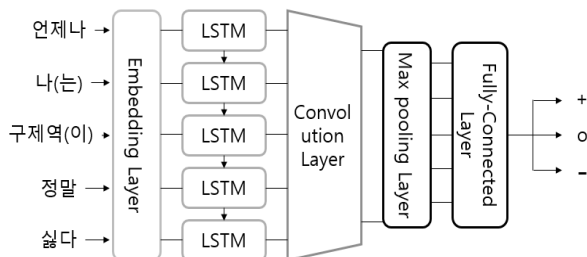


Fig. 2. LSTM-CNN Model

한편, 학습 데이터 부족으로 인한 분류기의 정확도 저하 문제를 해결하기 위하여 본 연구에서는 영화 데이터를 활용한 전이학습(Transfer Learning)[20-21]과 동의어 대체 방식[22]의 데이터 확대(Data Augmentation)기술을 사용하여 LSTM-CNN 모델의 분류 정확도를 향상한다.

3.3 키워드 네트워크 분석 모듈

키워드 네트워크 분석 모듈에서는 먼저, 극성탐지 모듈에서 결정된 '긍정', '부정'에 해당하는 트윗을 대상으로 토픽 모델링 방법 중 하나인 LDA를 사용하여 극성별 트윗과 관련된 다양한 키워드를 추출한다. 추출된 키워드들은 각 구간 및 극성에서 가장 많이 언급된 상위 3가지 주제(Topic)와 연관 깊은 키워드들로써, 추출된 키워드를 이용하여 동시출현 키워드 네트워크를 구성하고, 구제역으로 인한 시기별, 극성별 파급효과들을 분석한다. 동시출현 키워드 네트워크는 키워드가 해당 트윗에 출현한 횟수를 나타내는 문서-키워드 행렬(Document-Keyword Matrix)을 키워드간의 인접 행렬(Adjacency Matrix)로 변환하고, 이를 동시출현 키워드 네트워크로 구성한다. 이때, 네트워크 노드의 크기는 키워드의 출현 빈도를 나타내며, 엣지(Edge)의 두께는 키워드간의 동시출현 빈도를 의미한다. 즉, 두 키워드의 동시출현 빈도가 높을수록 두 키워드의 연관성이 높음을 뜻한다.

4. 실험 및 결과 분석

4.1 실험 데이터 및 실험 설계

본 연구의 실험에서는 국내 트위터에서 '구제역' 키워드를 포함하는 게시글을 수집하였으며, 수집 기간은 구제역으로 가장 큰 피해가 발생한 시기의 전·후 기간인 2010년 7월부터 2011년 12월까지로 설정하였다. 트윗 수집 결과는 Fig. 3과 같으며, 실제 구제역 발생 추이와 구제역 관련 트윗 수가 비슷한 모양새를 보임을 확인할 수 있다. 수집된 트윗 데이터의 형태소 분석은 트위터에서 제공하는 한국어 형태소 분석기를 사용하였고, 불용어 제거 및 토픽 모델링은 통계프로그램 R의 KoNLP 패키지와 topicmodels 패키지를 사용하였다. 또한, LSTM-CNN 모델의 학습 및 분류를 위하여 Python의 tensorflow 패키지를 이용하였으며, 네트워크 시각화 및 분석

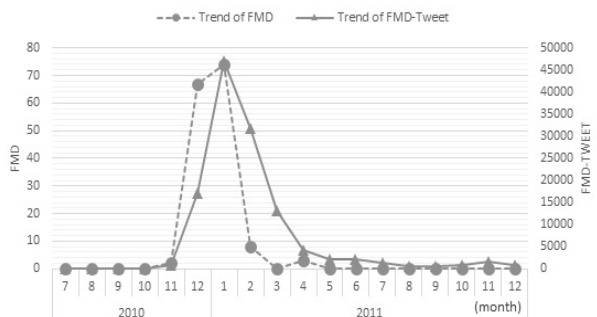


Fig. 3. Trend of FMD and FMD-Tweet

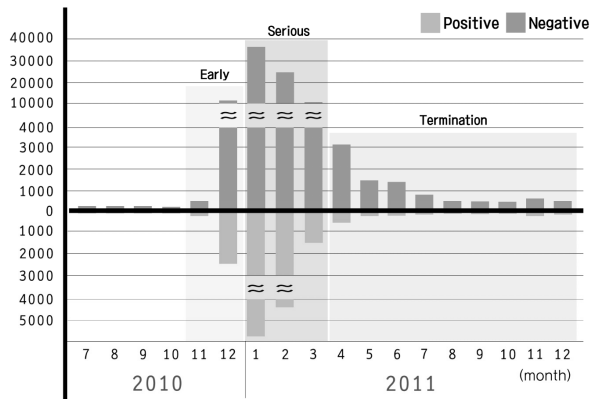


Fig. 4. Trend of FMD-Tweet Polarity

패키지인 igraph 패키지를 사용하여 ‘구제역’ 키워드를 중심으로 성형(star) 구조의 동시출현 키워드 네트워크를 구성하였다. 또한, Fig. 4와 같이 정부의 구제역 위기정보단계를 기준으로 구제역 발생 시기를 세 구간(‘발생 초기(Early)’, ‘심각기(Serious)’, ‘종식 이후(Termination)’)으로 구분하고, 각 구간에서의 긍·부정에 해당하는 일반 시민의 감성적 정보를 키워드 네트워크를 통해 분석하도록 설계하였다.

4.2 극성 분류 실험

실험에 앞서, 총 13만 건의 트윗 데이터 중에서 LSTM-CNN 분류기의 학습을 위해 극성을 레이블링하는 작업을 진행하였다. 극성 분류 기준은 한국어학 연구를 참조하여[23], 총 3명의 연구원들이 레이블링 작업에 참여하였다. 연구원들간의 의견 차이는 다수결 원칙으로 해결하였다. 총 레이블링된 트윗의 텍스트 데이터 수는 약 3천개(긍정: 800개, 중립: 1,000개, 부정: 1,200개)이다. 그러나 3천개의 데이터만으로는 분류기의 학습 정확도를 보장할 수 없기 때문에 데이터 확대 및 전이학습을 실시하였다. 데이터 확대를 위해서는 Natural Language Processing(NLP)분야에서 통상적으로 사용하는 동의어 대체 방식을 사용하였다(3천개의 데이터를 6천개로 증가(Table 2 참조). 또한, 웹 포털 사이트 ‘N’사에서 제공하는 영화 리뷰 데이터 12만 건을 수집하였고, 영화의 평점을 기준으로(‘1~4점: 부정’, ‘5, 6점: 중립’, ‘7~10점: 부정’) LSTM-CNN의 선행학습을 진행한 후, 선행 학습된 분류기의 매개변수 값을 기반으로, 준비된 구제역 트윗 데이터를 학습시키는 전이학습(Transfer Learning)을 실시하였다. 본 연구에서 선행학습 및 전이학습에 사용된 LSTM-CNN 분류기의 하이퍼파라미터(Hyperparameter)는 Table 3과 같으며, 실제

Table 2. Example of Synonyms Substitution

	Target Words	Substitution Words
1	구제역	가축질병
2	안동	경북
3	바라다	원하다
4	돼지	가축

Table 3. Training Parameters

	Pre-training/Transfer-training
Embedding Dimension	100
Epoch	10
Batch Size	64
Filter	32
Kernel Size	3
Pool Size	2
Dropout	0.5
Learning Rate	0.01/0.001

3천개의 트윗 텍스트 데이터를 9대1 비율로 나눠 학습 및 검증 데이터로 사용하여 실험한 결과, 감성 분류 정확도를 86.7%까지 향상시킬 수 있었다.

LSTM-CNN 분류기의 최종 분류로써 약 13만 건의 트윗 데이터 중에서 중립성을 띠는 데이터를 제외한 10만 여건의 데이터의 극성을 분류하였다. 분류 결과는 Fig. 4와 같으며, 구제역이라는 가축질병의 특성상 부정적인 트윗이 긍정적인 트윗보다 확연하게 많은 것을 확인할 수 있었다.

4.3 키워드 네트워크 분석

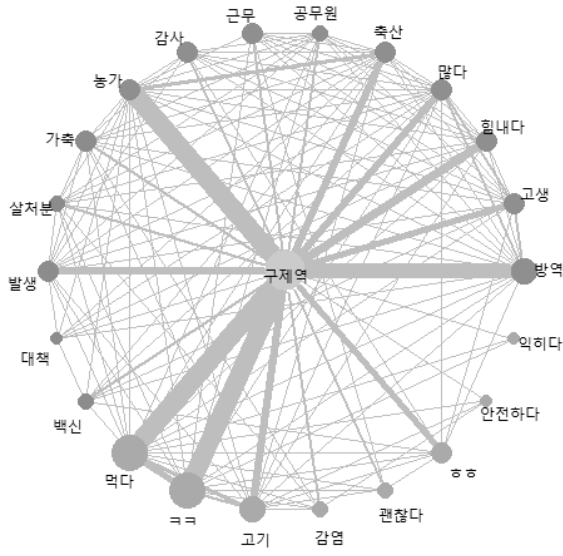
키워드 네트워크를 구성한 결과는 Fig. 5-7과 같으며, 극성별 기간별 트윗 중 같은 주제(Topic)로 언급된 키워드는 같은 색의 노드(Node)로 표현하였다.

1) 구제역 초기 구간

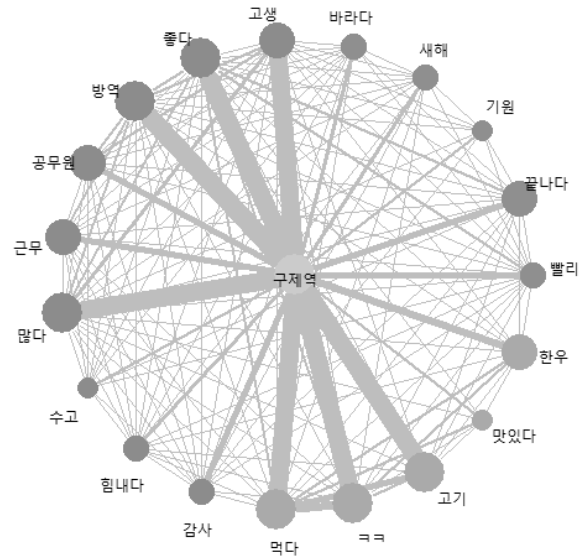
구제역 발생 초기 구간의 키워드 네트워크 분석 결과는 Fig. 5와 같다. 먼저, Fig. 5A에 의하면 초기 긍정 네트워크에서는 ‘공무원’, ‘근무’, ‘감사’, ‘힘내다’와 같은 키워드들이 등장하는 것으로 보아 방역 업무로 고생하는 공무원들을 위한 응원의 메시지로 해석된다. 또한, ‘고기’와 관련해서 ‘떡다’, ‘안전하다’, ‘익히다’와 같은 키워드들은 구제역이 발생했음에도 고기를 익혀먹으면 괜찮다는 뉴스 기사들이 일반 시민들에게 긍정적으로 작용된 것으로 보인다. 반면, 초기 부정 네트워크는 Fig. 5B와 같으며, 구제역 발생 및 확산에 관한 부정적인 시선뿐만 아니라 ‘살처분’, ‘가축’, ‘불쌍하다’와 같은 키워드들이 강하게 나타나는 것으로 보아 살처분으로 죽어가는 돼지 및 관련 농가에 대한 연민의 감정이 구제역 초기 구간의 부정적인 의견 중에서 상당수를 차지하고 있는 것을 확인할 수 있다. 한편, ‘세차’, ‘도로’, ‘통제’, ‘방역’과 같은 키워드들의 등장은 현실적인 측면에서, 구제역의 발생으로 인한 도로 통제 및 자동차 방역으로 인한 세차의 필요성에 대한 일반 시민들의 불만으로 해석된다.

2) 구제역 심각기 구간

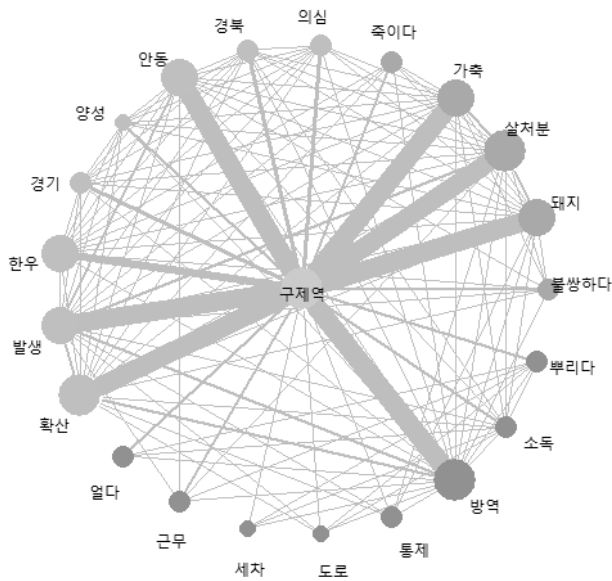
구제역 발생 심각기 구간의 키워드 네트워크 분석 결과는 Fig. 6과 같다. 먼저, 심각기 긍정 네트워크는 Fig. 6A와 같으며, 심각기 초기와 비슷하게 공무원들에 대한 응원과 고기에 대한 긍정적 시선을 여전히 엿볼 수 있다. 한편, 새해와 설날



A. Keyword Network in Positive Tweets

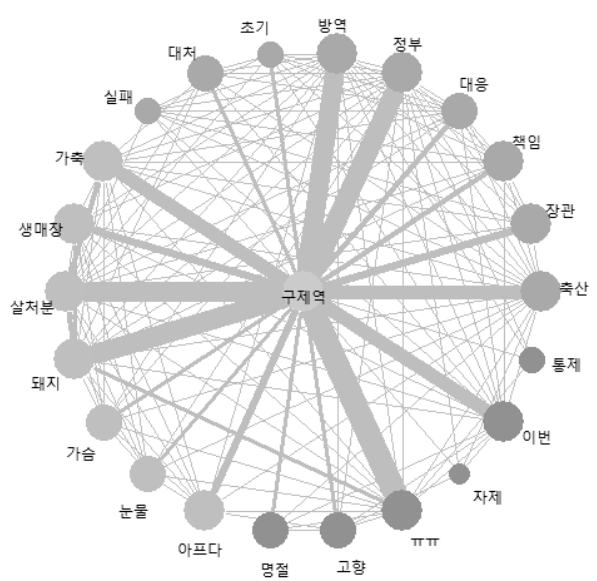


A. Keyword Network in Positive Tweets



B. Keyword Network in Negative Tweets

Fig. 5. Keyword Network in the Early Period of FMD



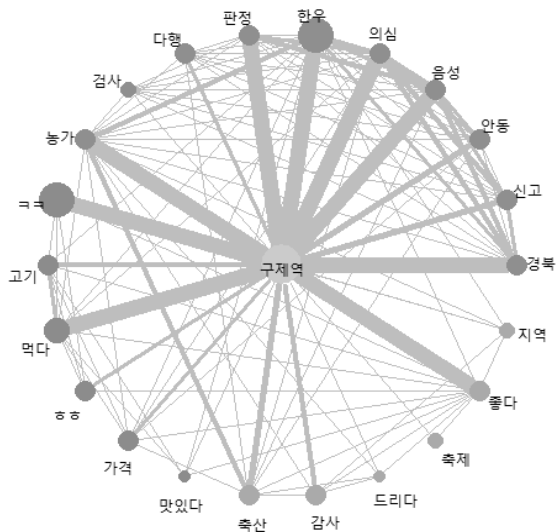
B. Keyword Network in Negative Tweets

Fig. 6. Keyword Network in the Serious Period of FMD

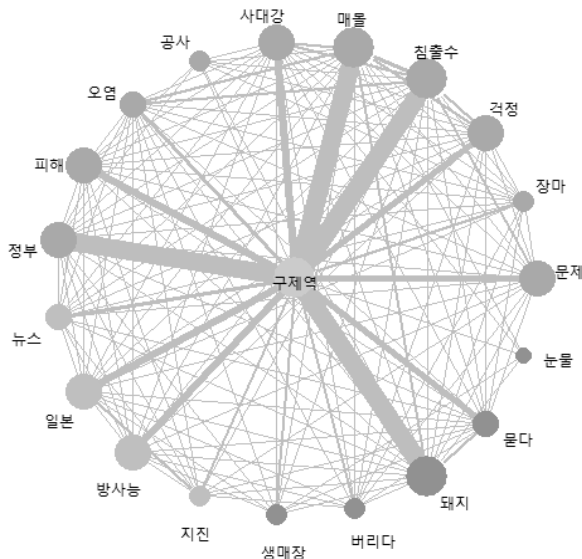
이 시기적으로 구제역 심각기 구간에 해당되어 ‘끝나다’, ‘기원’, ‘새해’, ‘바라다’와 같이 새해 및 설날 소망으로 구제역이 빨리 끝나길 바라는 희망적인 키워드들이 많이 등장하고 있다. 반면, 심각기 부정 네트워크(Fig. 6B)에서는 살처분에 대한 연민의 감정은 초기에 이어 계속 등장하고 있으며, ‘정부’, ‘대응’, ‘실패’와 같은 키워드들이 강하게 연결돼 있는 것을 확인 할 수 있다. 결국, 구제역이 심각해지면서 정부의 구제역 대응에 대해 일반 시민들이 강한 불만을 가지고 있음을 확인 할 수 있다. 한편, ‘명절’, ‘고향’, ‘자제’, ‘이변’, ‘통제’와 같은 키워드들로 구제역으로 인해 명절임에도 고향을 방문하지 못하는 사람들의 현실적인 고충도 엿볼 수 있다.

3) 구제역 종식이후 구간

구제역 발생 종식이후 구간의 키워드 네트워크 분석 결과는 Fig. 7과 같다. 먼저, 종식이후 긍정 네트워크(Fig. 7A)에서는 ‘의심’, ‘안동’, ‘음성’, ‘판정’, ‘신고’와 같은 키워드들이 서로 강하게 연결돼 있음을 알 수 있다. 이런 키워드들은 구제역이 결국 종식됨을 다행이라고 생각하는 일반 시민들의 긍정적 반응을 대변한다. 반면, 종식이후 부정 네트워크는 Fig. 7B와 같다. ‘침출수’, ‘매몰’과 같은 키워드들이 ‘구제역’과 강한 연결을 보이고 있으며, ‘사대강’, ‘격정’, ‘오염’ 등의 키워드가 함께 등장하는 것으로 보아, 가축의 매몰로 인한 오염으로 발생하는 침출수의 환경 피해 문제가 구제역 종식 이후 무엇보다



A. Keyword Network in Positive Tweets



B. Keyword Network in Negative Tweets

Fig. 7. Keyword Network in the Termination Period of FMD

보다도 일반 시민들에게 가장 큰 우려의 대상이 되는 것을 확인할 수 있다. 또한, 이와 같은 분석 결과는 기존 Noh 등 [6]이 진행한 온라인 뉴스 데이터를 활용한 구제역의 과급효과 분석에서 ‘침출수’와 같은 환경적 과급효과가 중요한 사회적인 문제로 부각되지 않았으나 트위터를 활용한 결과 시민들이 해당 부분에 상당한 관심이 있음이 드러났다. 이와 같이 트윗 분석을 통해 타 매체로는 분석이 어려운 실제 사람들이 해당 사건을 어떻게 해석하고, 본인과 관련을 짓는지에 대한 실질적 과급 효과에 대한 분석이 가능함을 확인하였다.

5. 결론

본 논문에서는 다양한 사회적 이슈들에 즉각적으로 반응하는 대표적 SNS인 트위터를 대상으로 텍스트 마이닝 방법

론을 활용하여 구제역이 일반 시민들에게 실질적으로 어떤 감정적 영향을 미치는 지를 분석하였다. 구제역 발생 시기를 ‘발생 초기’, ‘심각기’, ‘종식 이후’로 구분하고, 각 구간에서의 극성에 따른 과급효과들을 동시출현 키워드 네트워크를 사용하여 분석하였다. 분석결과를 요약하면, 전 구간에 걸쳐 구제역 예방 및 대처에 대한 키워드들이 많이 등장하면서 구제역의 피해로부터 벗어나고자 하는 사람들의 긍정적인 의견을 엿볼 수 있었다. 또한, 구제역 확산 및 정부 대응에 대한 강한 불만과 가축 매몰로 인한 침출수의 문제가 일반 시민들에게 가장 큰 문제이자 부정적 측면으로 다가움을 확인했다.

본 연구는 구제역의 발생 초기부터 심각기를 거치면서 종식 이후까지 일반 시민들이 느끼는 진솔한 감정적 변화를 트윗을 통하여 공학적으로 추적하는 시도라고 할 수 있다. 구제역에 대한 일반 시민들의 진솔한 감성분석이라는 사회학적 문제를 체계적으로 해결하기 위하여, 새로운 공학적인 분석 시스템을 구축하여 제시했다는 점이 본 연구의 주요한 의미라고 볼 수 있다. 본 연구에서 제안한 분석 시스템은 독립적 혹은 다른 방법론과의 협력적 활용을 통하여 다양한 분석 과정에서 본 연구의 시스템이 적용될 수 있을 것으로 전망되며, 추후 보다 건설적 차원의 후속 연구들이 기대된다.

References

- [1] S. I. Pak and S. H. Bae, “A Space-Time Cluster of Foot-and-Mouth Disease Outbreaks in South Korea, 2010~2011,” *Journal of the Korean Association of Regional Geographers*, Vol.18, No.4, pp.464-472, 2012.
- [2] Z. Xu, J. Lee, D. Park, and Y. Chung, “Multidimensional Analysis Model for Highly Pathogenic Avian Influenza using Data Cube and Data Mining Techniques,” *Biosystems Engineering*, Vol.157, pp.109-121, 2017.
- [3] H. Kim, S. Oh, S. Ahn, and B. Cho, “Real-time Monitoring Method of Cattle’s Temperature for FMD Prevention and Its Case Studies,” *Journal of KIIT*, Vol.15, No.5, pp.141-150, 2017.
- [4] M. J. Kyung and J. H. Yom, “Implementation of Open Source SOLAP Decision-Making System for Livestock Epidemic Surveillance and Prevention,” *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol.30, No.3, pp.287-294, 2012.
- [5] J. M. Jo, J. S. Jo, W. Qasim, B. E. Moon, M. H. Lee, et al., “Analysis of Route Patterns of Vehicles Transporting Livestock Manures using GPS,” *Journal of Agriculture and Life Science*, Vol.52, No.1, pp.83-91, 2018.
- [6] B. Noh, Z. Xu, J. Lee, D. Park, and Y. Chung, “Keyword Network based Repercussion Effect Analysis of Foot-and-Mouth Disease using Online News,” *Journal of KIIT*, Vol.14, No.9, pp.143-152, 2016.
- [7] H. J. Woo and Y. H. Kim, “Spatial Distribution Patterns of Twitter Data with Topic Modeling,” *Journal of the Korean Association of Regional Geographers*, Vol.23, No.2, pp.376-387, 2017.

- [8] S. Son, D. Kim, S. Lee, M. Gil, and Y. Moon, "Storm-Based Dynamic Tag Cloud for Real-Time SNS Data," *KIPS Transactions on Software and Data Engineering*, Vol.6, No.6, pp.309-314, 2017.
- [9] Y. Kim and H. Kang, "An Analysis of Relationship Between Word Frequency in Social Network Service Data and Crime Occurrences," *KIPS Transactions on Computer and Communication Systems*, Vol.5, No.9, pp.229-236, 2016.
- [10] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, Vol.9, No.8, pp.1735-1780, 1997.
- [11] C. Bziotis, N. Pelekis, and C. Doukeridis, "Datastories at SemEval-2017 task 4: Deep LSTM with Attention for Message-Level and Topic-Based Sentiment Analysis," *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp.747-754, 2017.
- [12] Y. Kim, "Convolutional Neural Networks for Sentence Classification," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp.1746-1751, 2014.
- [13] M. Cliche, "BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs," *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp.573-580, 2017.
- [14] C. Wu, F. Wu, J. Liu, S. Wu, and Y. Huang, "THU_NGN at SemEval-2018 Task 1: Fine-grained Tweet Sentiment Intensity Analysis with Attention CNN-LSTM," *Proceedings of The 12th International Workshop on Semantic Evaluation (SemEval-2018)*, pp.186-192, 2018.
- [15] M. Ko, "Unstructured Data Processing Using Keyword-Based Topic-Oriented Analysis," *KIPS Transactions on Software and Data Engineering*, Vol.6, No.11, pp.521-526, 2017.
- [16] S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 task 4: Sentiment Analysis in Twitter," *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp.502-518, 2017.
- [17] Y. D. Yun, J. C. Jo, Y. A. Hur, and H. S. Lim, "A Comparative Analysis of Cognitive Change about Big Data using Social Media Data Analysis," *KIPS Transactions on Software and Data Engineering*, Vol.6, No.7, pp.371-378, 2017.
- [18] M. S. Seo and H. H. Yoo, "Citizen Sentiment Analysis of the Social Disaster by Using Opinion Mining," *Journal of the Korean Society for Geospatial Information Science*, Vol.25, No.1, pp.37-46, 2017.
- [19] P. Oh and B. Y. Hwang, "Real-Time Spatial Recommendation System Based on Sentiment Analysis of Twitter," *Journal of Society for e-Business Studies*, Vol.21, No.3, pp.15-28, 2017.
- [20] J. Howard and S. Ruder, "Fine-tuned Language Models for Text Classification," *arXiv preprint arXiv:1801.06146*, 2018.
- [21] G. Daval-Frerot, A. Boucekif, and A. Moreau, "Epita at SemEval-2018 Task 1: Sentiment Analysis using Transfer Learning Approach," *Proceedings of The 12th International Workshop on Semantic Evaluation (SemEval-2018)*, pp.151-155, 2018.
- [22] S. Vosoughi, P. Vijayaraghavan, and D. Roy, "Tweet2vec: Learning Tweet Embeddings using Character-Level CNN-LSTM Encoder-Decoder," *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. ACM*, pp.1041-1044, 2016.
- [23] J. Hong and Y. Jeong, "Establishing the Category of Emotion Verb and Classifying Emotion Verbs," *Korean Linguistics*, Vol.45, pp.387-420, 2009.

채 희 찬



<https://orcid.org/0000-0002-6589-7694>

e-mail : chay219@korea.ac.kr

2018년 고려대학교 컴퓨터정보학과(학사)

2018년~현 재 고려대학교

컴퓨터정보학과 석사과정

관심분야: 딥러닝, 텍스트마이닝,

자연어처리, 빅데이터

이 종 욱



<https://orcid.org/0000-0002-2077-4850>

e-mail : eastwest9@korea.ac.kr

2002년 고려대학교 전산학과(학사)

2005년 고려대학교 전산학과(석사)

2014년 고려대학교 전산학과(박사)

2014년~현 재 고려대학교

컴퓨터정보학과 초빙교수

관심분야: 딥러닝, 데이터마이닝, 융합 IT, 음향분석

최 윤 아



<https://orcid.org/0000-0002-8207-9108>

e-mail : cyabc2@korea.ac.kr

2015년~현 재 고려대학교

컴퓨터정보학과 학사과정

관심분야: 빅데이터, 데이터마이닝,

딥러닝



박 대 희

<https://orcid.org/0000-0003-4726-4508>
e-mail : dhpark@korea.ac.kr
1982년 고려대학교 수학과(학사)
1984년 고려대학교 수학과(석사)
1989년 플로리다 주립대학 전산학과(석사)
1992년 플로리다 주립대학 전산학과(박사)

1993년~현 재 고려대학교 컴퓨터정보학과 교수
관심분야: 빅데이터, 데이터마이닝, 인공지능, 융합 IT



정 용 화

<https://orcid.org/0000-0001-6539-167X>
e-mail : ychungy@korea.ac.kr
1984년 한양대학교 전자통신공학과(학사)
1986년 한양대학교 전자통신공학과(석사)
1997년 U. of Southern California(박사)
1986년~2003년 한국전자통신연구원
생체인식기술연구팀(팀장)

2003년~현 재 고려대학교 컴퓨터정보학과 교수
관심분야: 병렬처리, 영상처리, 융합 IT