

---

# 소셜 빅 데이터를 이용한 이슈 감지 사례분석

송은지\*, 강민석\*\*

\*남서울대학교 컴퓨터학과

\*\*남서울대학교 산업경영공학과

## A Case Study of the Issue detected Analysis on Social Media Big Data

Eun-Jee Song\* Min-Shik Kang\*\* ,

\*Dept. of Computer Science , Namseoul University

\*\*Dept. of Industrial and Management Engineering , Namseoul University

sej@nsu.ac.kr, mskang@nsu.ac.kr

### 요 약

최근 IT업체들은 온라인 상에서 소비자들이 평소에 쏟아내는 의견들을 수집, 축적해서, 원하는 키워드를 중심으로 내용을 분석함으로써, 특정 주제에 대해 어떤 여론이 형성되고 있으며, 여론이 어떻게 전파되고 있는지 경로를 파악할 수 있는 소셜 빅데이터 분석 툴을 경쟁적으로 개발하고 있다.

본 논문에서는 소셜 빅 데이터를 분석함에 있어 이슈를 감지하고 예측하는 기술을 실제 사례에 적용하여 분석한 결과를 고찰해 보고자 한다. 소셜 미디어 데이터 패턴을 비교 분석하고 부정이슈 감지를 위해 부정 여론을 확산시키는데 영향을 미치는 내용과 작성자를 독립변수로 하고, 평균 이슈 도달 시간 및 속도를 종속변수로 정의한다. 부정 여론 형성의 영향력은 트윗수, 리트윗 수를 기준으로 이슈 감지한다. 분석결과 전체 트윗 중 리트윗 메시지가 큰 비중 차지하고 이슈에 대한 버즈가 증가할수록 리트윗 비중이 증가하였으며 크게 확산될 때는 리트윗량이 크게 증가하여 짧은 시간 안에 넓게 확산하였다.

### 키워드

Social media, Big data ,Issue Detected Analysis, Case Study

### 1. 서 론

IT 시대에 인터넷이 핵심 자원이었지만 최근 모바일 스마트 시대를 맞아 빅 데이터가 핵심 자원 역할을 하고 있다. 특히, 기업의 효율적인 경영을 위한 고객피드백을 얻기 위해서 빅 데이터는 꼭 필요한 자원이다. 빅 데이터는 단지 그 막대한 정보의 양 뿐만이 아니라 실시간으로 업데이트 되는 데이터의 생성 및 유통 속도와 매우 다양한 비정형, 비구조 데이터가 상호 융합되어 있다는 복합적인 특성을 가지므로 기존 데이터분석 기술로는 충분하지 않다.

빅 데이터의 대부분이 텍스트와 이미지 등 비정형 데이터로 구성되어 있다는 것을 볼 때 비정형 빅 데이터에의 접근과 분석을 위해 기존의 분석 기술과는 다른 비정형 데이터 분석기술이 필요하다. 온라인상에서 고객들이 쏟아내는 말을 실시간으로 분석, 조사하는 방법으로 버즈 모니터링이라는 방식이 있다. 이것은 웹상의 다양한 정보를 자동으로 검색하고 수집하는 시스템이다. 최근 버즈모니터링 시스템을 이용하여 이슈를 감지하고 분석하고 예측하는 방법을 제안하고 있다.

본 논문에서는 이슈를 감지하고 예측하는 기술을 실제 사례에 적용하여 분석한 결과를 고찰해 보고

자 한다. 분석사례로서 강남구청을 대상으로 부정 여론형성에 대해 살펴보고 시사점을 고찰한다. 부정 여론 형성의 영향력은 트윗수, 리트윗 수를 기준으로 이슈 감지한다.

## II. 이슈감지 방법 및 사례분석

### 1. 이슈감지 방법

이슈 감지 방법은 먼저 예측 모델 개발 도메인 선정한다. 예측모델의 도메인 선정을 위해 다음과 같은 3가지 요인을 고려한다. ①타 도메인에 확장 적용 가능한 표준성, 범용성 ②소셜 미디어 상의 수집 데이터 확보 여부 점검 ③ 이슈의 지속성 이것을 기반으로 이슈 감지를 위한 알고리즘은 다음과 같다.

1)대상 도메인 소셜 데이터 수집하고 분석 도메인과 관련한 하위 담론 유형 수립한다.:하위 담론 유형 정의하고 유형별 핵심 키워드 매핑하고 담론 유형별 패턴 추출한다.

2)도메인 특화 감성 사전의 구축한다.: 도메인 내 주제 지향적 감성 사전 구축하고 주요 감성 카테고리 정의하여 감성어휘 매핑한다. 감성 카테고리별 소셜 데이터의 시계열적 패턴 추출한다.

3)이슈 예측 대상 도메인 데이터 패턴 즉, 소셜 미디어 데이터 패턴 (하위 담론 유형별 패턴, 감성 카테고리별 패턴) 비교 분석한다 : 패턴 비교에 의한 이슈 사전 예측, 이상 패턴 감지에 의한 이슈를 감지한다.

부정이슈 감지를 위해 부정 여론을 확산시키는 데 영향을 미치는 내용과 작성자를 독립변수로 하고, 평균 이슈 도달 시간 및 속도를 종속변수로 정의하여 부정 여론 형성의 영향력은 트윗수, 리트윗 수를 기준으로 이슈 감지하도록 한다.

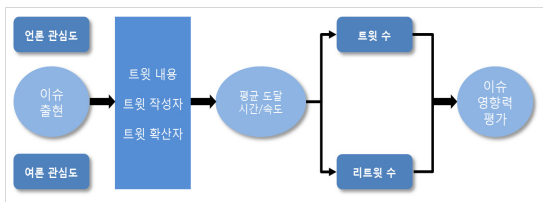


그림 1. 이슈감지 프로세스

### 2. 사례분석

이슈 감지 대상으로 강남구청의 이슈 사례 분석을 시행하였다 강남구청에 대한 버즈 및 연관어 분석 결과, 2013년 5월(박원순 시장고소), 2014년 2월(연예인 홍보대사 위촉)으로 인한 이슈가 증폭하였다. 연관어 분석 결과, 행사 및 사건이 차지

하는 비중이 50% 이상을 차지하였고 ‘exo 홍보대사 위촉’, ‘박원순 시장 고소’, ‘아이파크 헬기 충돌’ 등이 주요 이슈로 나타났다.

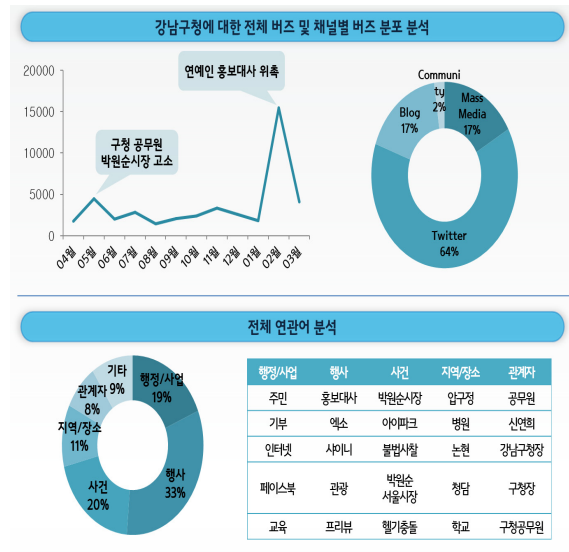


그림 2. 강남구청 이슈 분석결과

## III. 결론 및 향후과제

최근 IT업체들은 온라인 상에서 여론이 어떻게 전파되고 있는지 경로를 파악할 수 있는 소셜 빅 데이터 분석 툴을 경쟁적으로 개발하고 있다. 본 논문에서는 소셜 빅데이터 분석기술을 이용하여 실제 강남구청의 이슈를 감지하고 분석한 적용사례 결과를 제시하였다. 이슈 예측 대상 소셜 미디어 데이터 패턴을 비교 분석하고 부정 여론 형성의 영향력은 트윗수, 리트윗 수를 기준으로 이슈 감지하도록 하였다. 분석결과 전체 트윗 중 리트윗 메시지가 큰 비중 차지하고 이슈에 대한 버즈가 증가할수록 리트윗 비중이 증가하였으며 크게 확산될 때는 리트윗 량이 크게 증가하여 짧은 시간 안에 넓게 확산하였다. 향후 다양한 실제사례에 적용하여 분석할 예정이다.

### 참고문헌

[1] C.H. Lee, "Technology Trends of Issue Detection and Predictive Analysis on Social Big Data", ETRI Electronics and Telecommunications Trends, pp62-71, 2013.

[2] 송은지외, "소셜미디어 빅 데이터 분석을 통한 이슈 감지 및 예측에 관한연구" 한국정보통신학회 2014춘계 학술대회 논문집 제18권 제1호, pp.629-630, 2014.

[3] 류범모외, "심층 언어분석 기반 소셜미디어 이슈 탐지 및 모니터링 기술", 한국정보과학회 논문지 제30권 제6호, pp.47-58, 2012.