

소셜 미디어 빅 데이터 분석을 통한 이슈 감지 및 예측에 관한 연구

강민식*, 송은지**

*남서울대학교 산업경영공학과

**남서울대학교 컴퓨터학과

A Study on the Issue detected and Forecast by Analysis of Social Media Big Data

Min-Sik Kang* , Eun-Jee Song**

*Dept. of Industrial and Management Engineering , Namseoul University

**Dept. of Computer Science , Namseoul University

mksang@nsu.ac.kr, sej@nsu.ac.kr

요 약

서비스 산업에 있어 기업 간의 경쟁이 날로 심화되어 가고 있는 가운데 효율적인 경영을 위해서는 시시각각으로 변하는 고객의 니즈를 파악하기 위해 그 어느 때 보다도 고객피드백이 필요한 시대이다. 최근 기업에서는 다양한 고객의 목소리가 담겨 있는 소셜 미디어상의 빅 데이터를 이용하여 고객의 피드백을 파악하려는 노력을 하고 있다. 따라서 모바일 스마트 혁명의 핵심 자원인 빅 데이터를 어떻게 분석, 활용 할 것인지 많은 기업들의 관심이 집중되고 있다.

본 연구에서는 이러한 소셜 빅 데이터를 분석하는 기술로서 최근 이슈를 감지하고 예측하는 방법을 제안한다. 이것은 기관이나 기업 등 분석대상과 관련된 소셜 데이터 자체를 분석하거나 그 외 관련 데이터와 연관 관계 분석 등 여러 가지 방법을 조합하여 부정적 이슈 등의 탐지가 가능하다.

키워드

Social media, Big data ,Issue detected, Forecast

1. 서 론

블로그나 SNS등에는 기업이 상품이나 서비스를 팔고자 하는 소비자가 가득 모여 있기 때문에 실시간으로 소비자의 니즈를 파악할 수 있는 방법으로 SNS 등과 같은 빅 데이터의 이용이 매우 효율적이다. IT 시대에 인터넷이 핵심 자원이었지만 모바일 스마트 시대에는 빅 데이터가 핵심 자원 역할을 할 것이다. 모바일 스마트 현시대엔 기업의 효율적인 경영을 위해 고객피드백을 얻기 위해서 빅 데이터는 꼭 필요한 자원이다. 빅 데이터는 단지 그 막대한 정보의 양 뿐만이 아니라 실시간으로 업데이트 되는 데이터의 생성 및 유

통 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 정보 검색과 텍스트 마이닝 기술과 같은 비정형 데이터 분석기술이 필요하다. 온라인상에서 고객들이 쏟아내는 말을 실시간으로 분석, 조사하는 방법으로 버즈 모니터링이라는 방식이 있다. 이것은 웹상의 다양한 정보를 자동으로 검색하고 수집하는 시스템이다.

본 연구에서는 빅 데이터를 분석하는 버즈모니터링 시스템을 개발함에 있어 최근 이슈를 분석하고 예측하는 방법을 제안하고자 한다. 분석 대상 (기관, 기업 등)과 관련된 소셜 데이터 자체를 분석하거나 그 외 관련 데이터와 연관 관계 분석 등 여러 가지 방법을 조합하여 부정적 이슈의 탐지가 가능하다.

II. 이슈감지 및 예측 방법

분석 대상 (기관, 기업 등)과 관련된 소셜 데이터 자체를 분석하거나 그 외 관련 데이터와 연관 관계 분석을 다음과 같은 여러 가지 방법을 조합하여 이슈의 탐지와 예측이 가능하도록 한다.

(1)Anomaly Detection:정상적인 경우와 비교하면 기준에 알려지지 않은 이례적인 변칙 패턴을 발견하며 기초 분석된 결과나 수집된 원천 데이터를 관심 주제어와 속성 등 조건에 따라 분류, 패턴화한 후, 일반적 경향성에 벗어나는 이상 패턴 발생시 감지한다. 또한 관심 주제어, 대상, 지역, 기간 등 다양한 관점의 통계 결과 가능하다. (2)Predictive Models:가설기반 검증, 예측평가등 다양한 예측 기법을 통해, 알려지지 않은 복잡한 부정 이슈 발생의 패턴 발견하여 이슈 정보 예측 분석한다. 이슈의 실제 발생 추이와 소셜 미디어 데이터 추이 간의 ‘그랜저 인과관계 (Granger Causality)’ 분석 등을 토대로 향후 추이를 예측한다.(3)Network Analysis:소셜 네트워크 연결망, 위치정보 등의 데이터를 활용한 연관 관계 분석을 통해 알려지지 않은 부정적 이슈를 발견한다.(4)Text Mining : 소셜미디어의 비정형 텍스트의 의미를 분석하여 부정적 이슈를 탐지한다.

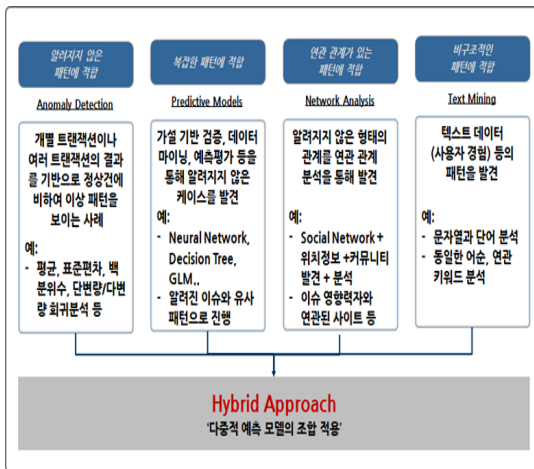


그림 1. 이슈감지 및 예측 모델

이러한 다양한 방법을 조합하여 이슈를 감지하고 예측하는 모델은 그림1과 같다. 제안한 방법에 의해 예를 들어 그림2와 같이 청소년자살수와 자살이라는 키워드의 빈도수를 분석하여 자살률을 예측한다.

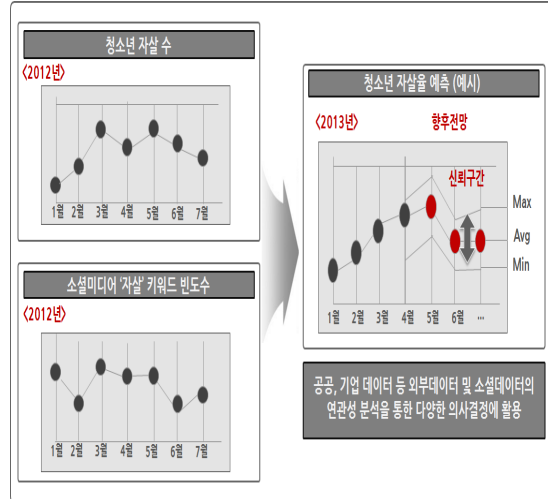


그림 2. 이슈정보 예측 분석기술 예시

III. 결론 및 향후과제

최근 IT업체들은 온라인 상에서 소비자들이 평소 애용에 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 빅데이터를 분석함에 있어 이슈를 감지하고 예측하는 방법을 제안하였다. 이슈 예측 대상 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 감지하도록 한다. 향후 제안한 방법에 의해 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] E.J. Song, S.J. Hong and M.S. Kang, "A Study on the System for Data Collection and Analysis on Social Network" Proceedings of International Workshop on Networking and Communication, 2013.