

오피니언 마이닝을 이용한 여론분석 시스템

김영아*, 김성권*, Fei Hao*, 박두순*
 순천향대학교 컴퓨터소프트웨어공학과
 e-mail: novaray@naver.com

Public opinion analysis system using opinion mining

Young-Ah Kim, Sung-Kwon Kim, Fei Hao, Doo-soon Park*
 Soonchunhyang University, Dept. of Computer Software Engineering*
 Soonchunhyang University,

요 약

최근 스마트폰 사용자와 SNS를 이용하는 사용자들이 늘어나고 있다. 또 다양한 SNS가 등장하면서 SNS데이터의 양이 방대해지고 SNS데이터의 가치와 신뢰성도 점점 높아지고 있다. 이러한 SNS 데이터를 사용하여 특정 키워드의 여론을 분석하고 사용자들의 반응을 얻는 것은 좋은 정보로 여러 분야에 사용될 수 있을 것이다.

본 논문에서는 SNS를 기반으로 오피니언 마이닝을 사용해 특정 키워드에 대한 SNS사용자들의 여론을 분석하였다. 그 결과 실시간으로 올라오는 글들에 대하여 해당 키워드가 어떤 여론을 가지고 있는지 분석 결과를 얻었다.

1. 서론

최근 스마트 기기 사용자의 증가와 함께 트위터(Twitter)와 페이스북(Facebook) 등으로 대변되는 SNS(Social Network Services:소셜 네트워크 서비스)의 이용이 폭발적으로 증가하고 있다[1].

또 10억 명 이상이 사용하고 있는 소셜 네트워크의 활성화로 SNS에서 다양한 의견과 선호도에 관한 중요한 정보를 추출하여 정치, 경제, 서비스, 연애 등 다양한 분야에 이용하고 있다. 따라서 소셜 네트워크에서 발생하는 대량의 데이터로부터 의미 있는 정보를 신속하게 추출하고 이를 처리하는 기술이 필요하게 되었다[2].

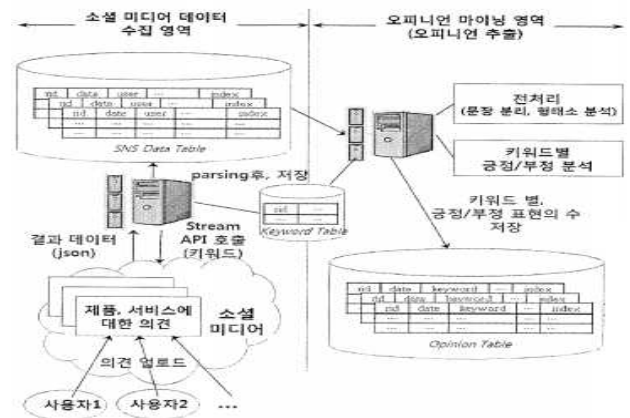
SNS에 올라오는 글들을 분석하기 위해서는 비정형 텍스트들을 분석할 수 있어야 한다. 이러한 텍스트 분석 기술을 활용한 분야 중 대표적인 것으로 오피니언 마이닝(Opinion Mining)을 들 수 있다. 오피니언 마이닝은 감성 분석(Sentiment Analysis)으로 불리기도 하며, 제품, 서비스, 조직, 개인, 이슈, 사건, 토픽, 그리고 이들의 여러 속성에 대한 사람들의 의견, 감성, 평가, 태도, 감정 등을 분석하는 일련의 과정을 의미한다[3].

오피니언 마이닝을 위한 감성분석은 SO - PMI(Semantic Orientation from Point-wise Mutual Information)를 이용하여 어휘의 의미 극성을 판단한다.

본 논문에서는 SNS로부터 얻어지는 사용자의 의견이 담긴 글을 가져와 형태소 분석기를 사용하여 일정한 단위의 형태소로 나눈 뒤, 나뉜 형태소와 감정단어를 통해 극성을 판별해 해당 키워드에 대한 여론분석 결과를 나타내주는 시스템을 개발한다.

2. 키워드 여론분석 시스템 설계 및 구현

오피니언 마이닝의 실행과정은 [그림 1]과 같다.



[그림 1] 데이터 수집과 오피니언 마이닝 실행과정

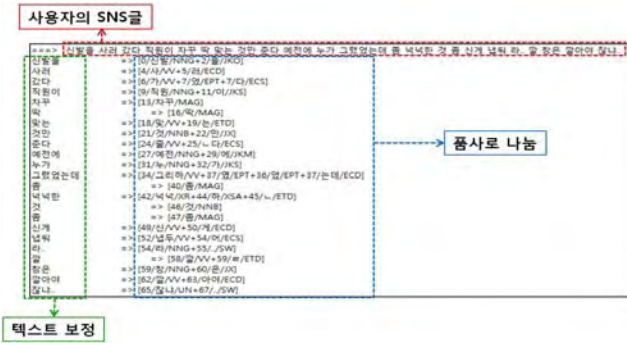
키워드 여론분석 시스템은 [그림 2]과 같다.

※ 이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2014R1A1A4A01007190)



[그림 2] 키워드 여론분석 시스템 구성도

[그림 2]을 보면 페이스북 API를 사용하여 타임라인에 실시간으로 올라오는 SNS글들을 데이터베이스에 저장시켜 주고, 사용자가 키워드를 입력했을 때 키워드를 사용해 분석에 들어간다. 사용자가 입력한 키워드를 받아 키워드에 맞는 SNS 사용자들의 글을 찾고 추출한다. 그 후 추출한 글을 형태소 분석기를 통해 형태소 단위로 자르고 명사를 뽑아낸다. [그림 3]은 가져온 타임라인을 텍스트 보정해 품사로 나눈 모습이다.



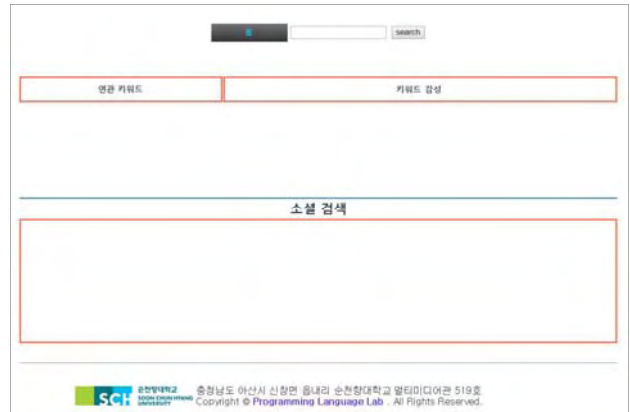
[그림 3] 형태소분석기를 통한 형태소 나눔

뽑아진 명사를 데이터베이스에 저장하고 카운트를 세어 해당 명사가 얼마나 그 키워드와 연관이 있는지 추출한다. 감정단어가 저장되어있는 데이터베이스와 PMI를 응용한 SO-PMI 연관도를 사용해 입력한 단어의 극성을 구하고자 하는 키워드와 기준 용어(seed term)와의 PMI를 계산하여 SO-PMI(word) 값이 양수라면 긍정으로 음수라면 부정으로 분류하는 방법이다[4]. 키워드의 긍정, 부정을 판단하고, 입력한 키워드가 포함된 SNS글과 연관검색어, 긍정, 부정에 대한 정보를 사용자에게 제공해주게 된다. [표 1]은 키워드 여론분석 시스템의 구현환경을 나타낸 표이다.

[표 1] 키워드 여론분석 시스템 구현 환경

시스템 구성요소	시스템 세부 내용
Server	Window 7 Java JDK 8.0 Tomcat 7.0
Database	Mysql 5.5
Web Language	PHP, JSP
API	꼬꼬마분석기 2.0 Oauth2.0
Develop Tools	Eclipse, Aptana Studio3

[그림 4]는 키워드 여론분석 시스템의 구현 초기화면이다.



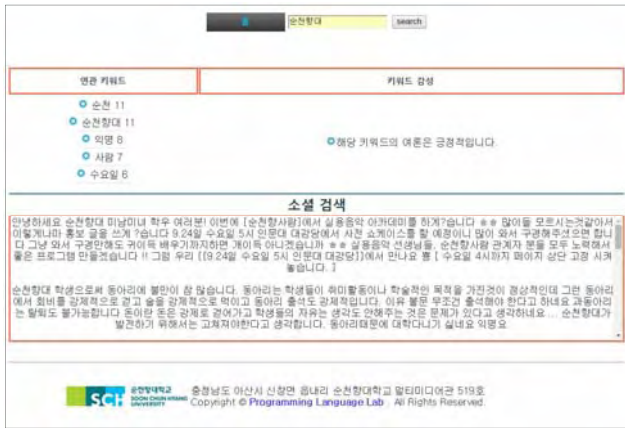
[그림 4] 키워드 여론분석 시스템 구현 화면

홈을 누르면 초기화면으로 돌아가게 되고, 옆에 있는 입력창에 여론을 알고 싶은 키워드를 입력하여 해당 키워드의 여론을 알 수 있다. 키워드를 입력하면 연관 키워드와 키워드의 긍정, 부정 여부, 그리고 해당 키워드가 속해있는 타임라인을 출력 받아볼 수 있다.

[그림 5]는 키워드로 한 예시로 '순천향대'를 입력했을 때의 결과화면이다. 연관키워드를 보면 형태소 분석을 거쳐 텍스트 보정된 단어를 걸러 추출된 명사들 중 키워드와 가장 많이 나타난 연관 키워드를 연관 키워드로 출력하였다. 연관 키워드의 오른쪽에는 SO-PMI를 통해 추출된 긍정, 부정 값으로 긍정은 양수, 부정은 음수, 0이면 중립으로 판단된 키워드의 여론을 출력해주게 된다. 그 아래쪽으로 보이는 소셜 검색 부분은 해당 키워드가 있는 타임라인을 출력해주게 된다.

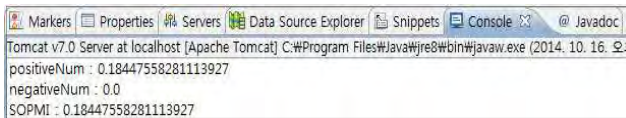
참고문헌

- [1] 지성훈, SNS 상의 기업계정에 관한 지속적 사용의도에 관한 연구 : 감정적 지향 요인을 중심으로. 高麗大學 校 大學院 : 經營學科, 2013. 2
- [2] 백봉현, SNS상의 비정형 빅 데이터로부터 감성정보 추출방법 영남대학교 대학원 : 컴퓨터공학 컴퓨터응용 및멀티미디어전공. 2014. 8
- [3] 김승우, 오피니언 분류의 감성사전 활용효과에 대한 연구 국민대학교 비즈니스IT전문대학원 : 비즈니스IT전공. 2014. 2
- [4] 자바 서버 페이지 - <http://www.wikipedia.org>



[그림 5] 키워드 ‘순천향대’ 결과화면

[그림 6]은 키워드로 ‘순천향대’를 입력했을 때 긍정PMI 값의 결과와 부정PMI의 결과를 계산한 SO-PMI의 계산 값이다.



[그림 6] 키워드 ‘순천향대’ SO-PMI값

3. 결론

최근 스마트폰의 발달로 인한 SNS사용자의 증가로 사용자들의 서로 정보를 공유하고 소통하면서 SNS데이터가 증가하고 그 데이터 가치가 높아지고 있다.

따라서 본 논문에서는 일반 사용자들도 특정한 키워드에 대한 여론을 알 수 있도록 하여 제약 없이 여론분석할 수 있는 시스템을 제작하게 되었으며, 연관검색어와 극성 판별을 통해 해당 키워드가 어떠한 여론을 가졌는지 분석할 수 있는 자료를 제공한다.