

오피니언 마이닝을 이용한 한글 트윗 감정분석 시스템

어문선*, 박두순*

*순천향대학교 컴퓨터소프트웨어공학과

e-mail : nina515@naver.com

The Hanguk Tweet Sentiment Analysis System using Opinion Mining

*Mun-Seon Eo, *Doo-Soon Park

*Dept. of Computer Software Engineering, SoonChunHyang University

요 약

인터넷과 스마트폰의 발달로 SNS서비스의 사용자와 데이터가 활발하게 증가하고 있다. 이로 인하여 SNS 데이터의 가치와 신뢰성이 점점 증가하고 있으며, 이러한 추세에 따라 여러 연구와 실험을 통하여 데이터를 분석하고 분석 결과를 제공하는 서비스가 증가하고 있다. 본 논문에서는 이러한 배경을 바탕으로 특정 키워드를 포함하고 있는 한글 트윗을 검색하여 해당 트윗에 대한 연관 키워드와 감정 키워드를 분석해서 출력해주는 시스템을 개발한다.

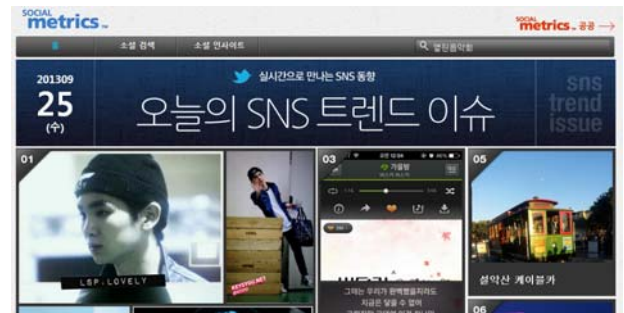
1. 서론

인터넷이 발달하면서 사람들이 소통하기 위하여 SNS 서비스가 생기게 되었고, 스마트 폰이 전 세계적으로 보급되면서 SNS의 사용자와 데이터가 증가하기 시작하였다. 이로 인하여 자연스럽게 사람들은 SNS에 자신들의 여러 의견과 이야기들을 게시하고 공유하게 되었다. 요즘은 이러한 SNS 데이터의 가치를 인지하고 여러 기업과 단체에서 이를 분석하고 자신들이 원하는 형태로 가공하여 사용할 수 있는 시스템을 많이 구축하고 있다.

본 논문에서는 특정 키워드를 포함하고 있는 한글 트윗을 분석하여 연관 키워드와 감정 키워드를 분석해서 출력해주는 시스템을 개발한다.

2. 개발배경 및 다른 분석 시스템 사례

현재 국내에서는 SNS의 발전과 함께 SNS 데이터를 분석하고 제공하는 시스템 시장이 성장해 나가고 있다. 그러나 이러한 시스템은 기업을 대상으로 하는 시스템이 많기 때문에 대부분 영리적으로 운영되고 있다. 그러다보니 일반 사용자들이 SNS 데이터 분석결과를 이용하고 싶어도, 돈을 지불해야 하거나 제한된 분석결과만 이용할 수 있는 실정이다. <그림 1>은 다음소프트에서 운영하는 웹기반 시스템 소셜 메트릭스 인사이트이고, 무료 사용자는 제한된 분석 결과만을 이용할 수 있다.



<그림 1> 다음소프트의 소셜 메트릭스 인사이트

<그림 2>는 트윗 트렌드라는 웹기반 시스템이며, 무료 사용자는 한 가지 키워드만 검색이 가능하다.

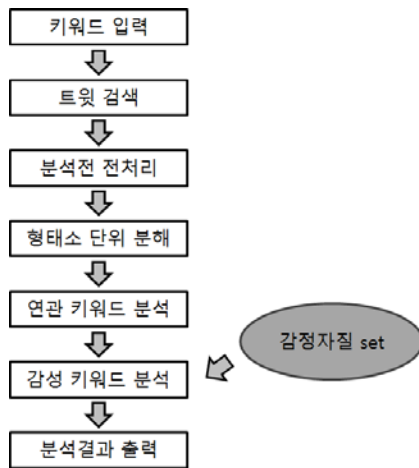


<그림 2> 트윗 트렌드

3. 트윗 분석 시스템의 구성

본 시스템은 분석을 원하는 키워드를 입력하면 해당 키워드에 대한 트윗들을 검색하여 분석한 뒤 연관 키워드와 감정 키워드를 표시 해줌으로써 트윗을 작성한 사람들이 해당 키워드에 대하여 어떻게 생각하는지를 볼 수 있는 지표를 제공한다.

이 트윗 분석 시스템의 동작 순서는 웹페이지에서 사용자에게 분석할 키워드를 입력 받고, 트위터 API를 이용하여 키워드를 포함하고 있는 한글 트윗을 검색한다. 검색된 트윗들은 인터넷 용어들을 제거하는 전처리 과정을 거치게 되며, 형태소 단위로 분해하여, 명사를 가지고 연관 키워드를 분석하고, 명사, 동사, 형용사로 감정 키워드를 분석한다[1]. 모든 분석을 마치면 사용자가 보기 편하도록 결과를 출력해주면 시스템의 동작이 완료되게 된다. 트윗 분석 시스템의 구성은 <그림 3>과 같다[2].



<그림 3> 트윗 분석 시스템의 구성

이 시스템에서 형태소 분석 부분과 감정 키워드를 분석하는 부분이 중요한 부분이다. 연관, 감정 키워드 분석을 위해서는 필수적으로 형태소 단위로 문장을 분해하고, 명사, 동사, 형용사 등을 분류해야 하기 때문이다. 형태소 분석을 하기 위해서 무료로 제공되는 JAVA용 형태소 분석기인 KOMORAN[3]을 이용했다.

그리고 감정 키워드 분석의 경우는 형태소 분석을 하면서 분해된 명사와 형용사들을 카운팅 하여, 상위 키워드를 다시 검색 한 뒤 검색된 문장에 포함되어 있는 “좋다”, “나쁘다”, “신난다”, “우울하다”와 같은 감정자질을 긍정과 부정으로 분류하여 수치화 한 다음 해당 상위 연관 키워드가 어떠한 감정을 가지고 있는지 보고, 긍정과 부정의 속성을 가지게 되면 감정 키워드로 분류하여 결과를 보여 줄 수 있도록 하였다.

<그림 4>는 트윗 분석 시스템으로 “연구”를 검색 하여 분석 했을 때 결과를 보여준다.

“연구” 트위터 검색 <540건>		13-09-24(11:20) ~ 13-09-24(13:00)	
연관 키워드		감정 키워드	
1 결과	236 회	1 표절	96 회
2 한국	174 회	2 긍정적	88 회
3 대한	154 회	3 성공	80 회
4 미국	140 회	4 도움	68 회
5 효과	136 회	5 위협	65 회
6 교수	127 회	6 치명적	47 회
7 개발	112 회	7 행복	37 회
8 기술	94 회	8 욕	36 회
9 대상	87 회	9 끝투	20 회
10 팀	43 회	10 공격적	14 회

<그림 4> “연구” 키워드를 분석한 결과

4. 결론

SNS 데이터의 양과 다양성이 증가하면서 데이터의 가치가 증가하고 있지만 실질적으로 일반 사용자들이 이를 효율적으로 분석할 수 있는 툴이나 시스템이 부족한 실정이다. 본 논문에서는 분석을 원하는 키워드를 입력하면 해당 키워드에 대한 트윗들을 검색하여 분석한 뒤 연관 키워드와 감정 키워드를 표시 해줌으로써 트윗을 작성한 사람들이 해당 키워드에 대하여 어떻게 생각하는지를 볼 수 있는 지표를 제공하는 시스템을 개발하였다.

앞으로는 일반 사용자들이 편하게 사용할 수 있는 분석 툴과 시스템들이 더욱더 많이 개발 되어야 할 것이다.

참고문헌

- [1] 황재원, 고영중, “감정 분류를 위한 한국어 감정 자질 추출 기법과 감정 자질의 유용성 평가”, 인지과학 제19권, 제4호, 499~517, 2008
- [2] 박경미, 박호건, 김형근, 고희동, “SNS에서 오픈이닝 마이닝 연구”, 정보과학회지 제11호, 54-60, 2011
- [3] <http://shineware.tistory.com/>