한국컴퓨터정보학회 하계학술대회 논문집 제23권 제2호 (2015. 7)

소셜미디어와 빅 데이터 마이닝 기술을 이용한 청소년 관련문제 분석시스템

서지애[O], 김창기[*], 서정민[**]
[O]상주상지여자고등학교, [*]한국교통대학교 사회복지학과
[**](주)디커뮤니케이션즈 기술연구소
e-mail:cgkim@ut.ac.kr[*], jmseo@web-d.co.kr[**]

An Youth-related Issues Analysis System Using Social Media and Big-data Mining Techniques

Ji Ea Seo[O], Chgan Gi Kim[*], Jeong Min Seo[**]
[O]Sangju Sangji Girl's High School,
[*]Dept. of Social Welfare, Korea Univ. of Transportation
[**]Research Center of DCommunications Co., Ltd.

● Abstract ●

본 논문에서는 학교 교육환경에서 청소년들에게 발생 할 수 있는소 셜미디어의 역기능을 빅 데이터 처리를 통하여 분석 할 수 있는 방법을 제시하고, 특히 악성 댓글을 위주로 한 청소년들 간의 소셜미디어를 중심으로 빅 데이터의 마이닝 기술을 활용하여 대표적인 청소년 문제의 확산을 방지 할 수 있는 시스템 제안한다.

키워드: 빅 데이터(Big-data), 마이닝(Mining), 소셜미디어(Social Media), 청소년(Youth)

I. Introduction

스마트 폰을 이용하여 소셜 미디어 등의 영역에서 가장 활발한 활동성을 보이고 있는 청소년 세대는 아직까지 자아형성이 완전히 이루어지지 않아 주위의 환경에 민감하게 반응하여 자아형성의 파괴형상을 일으킬 수도 있다. 특히 성적 압박, 학교폭력, 집단 따돌림 등의 스트레스가 원인이 되어 많은 청소년들이 방황을 하거나 충동적인 사건을 일으키기도 한다. 그러나 청소년들은 주로 인터넷이나 SNS 등의 소셜 미디어를 통해 자기의 현재 감정적 정보를 표현하는 경우가 많다. 이에 본 논문에서는 청소년들의 소셜 미디어 게시 글을 이용하여 그들의 현재 감정 상태를 분석하는 시스템을 제안한다. 시스템을 실험하기 위해 경상북도 상주시에 위치한 여자고등학교 1학년 학생들의 도움을 받아 페이스북 계정과 친구 맺기를 하여 청소년들의 글을 수집하여 분석하였다.

II. Related Works

1. 오피니언 마이닝

오피니언 마이닝(Opinion Mining)은 모바일 환경과 소셜 미디어가 급속하게 확산되면서 각종 여론 조사나 트랜드의 향방 조사 등에서 이용하는 기술로 주목받고 있다. 기존의 텍스트 마이닝 연구가 정보를 추출하는 것을 목적으로 하는 것과 달리 오피니언 마이닝에서는 텍스트에 내재된 감성을 추출하는 것을 목적으로 한다. 텍스트에서 감성 정보를 추출하는 방법은 정보검색 기술을 이용하는 자동화 방법[1]과 심리학적 지식을 이용하는 방법, 그리고 컴퓨터 시스템과 사람의 인지 능력을 이용하는 반자동 추출 방법[2]이 있다.

2. 오피니언 마이닝 감성 추출

의견 혹은 감성은 주관적인 표현, 즉 개인적인 느낌(feelings), 관점(views), 감정(emotions), 신념(beliefs) 들로 나타난다. 이것에 대한 평가는 일반적으로 긍정과 부정의 강도로 표현되며, 이를 표현하기 위한 요소로 감성값과 함께 상황, 인물, 시간 등의 정보가 사용된다.

가장 중요한 정보는 감성값인데 감성 사전이나 감성 의미망과 같은 리소스로부터 가져오게 된다. 영어권에 서 구축된 감성 관련 자원은 SentiWordNet[3]이 있으며 이들은 기존 단어 시소러스인 WordNet에 감성 정보를 추가한 것이다. 그러나, 영어와 한국어 단어가 가지는 감성도의 차이에 의한 한계가 존재한다.

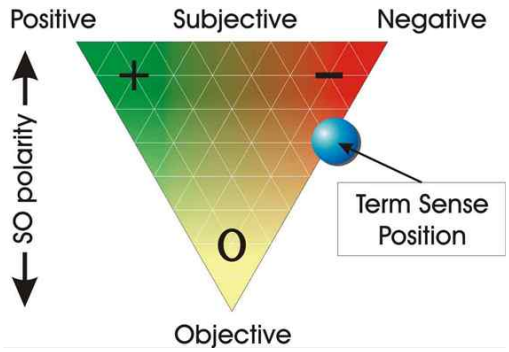

Fig. 1. The Structure of Sentiment Score

III. The Proposed System

제안하는 시스템에서 분석하고자 하는 소셜 미디어의 대상은 페이스북을 기반으로 하였다. 본 논문에서는 소셜 미디어를 분석하기 위해 자연어처리를 이용한 여고생의 오피니언 마이닝 방법을 이용하였다. 따라서 본 연구에 참여하는 여고생들의 페이스북 펠로우를 형성하도록 한 후 자기의 현재 감정에 관한 글을 수시로 올리도록 하였다. 시스템은 크게 5개의 모듈로 나눌 수 있는데, 최상위의 모듈은 펠로우들로 부터 관련 자료를 모으는 Collector 부분이며, 두 번째 모듈은 모은 자료들을 Text-Pre-Processing과 Text-Mining 및 이를 바탕으로 오피니언 마이닝을 하는 부분으로 나눌 수 있다. 세 번째 부분은 하둡의 마스터 노드로 본 논문에서는 4개의 서버 중 1개를 Master로 지정하였다. 그리고 네 번째와 다섯 번째 부분은 Slave Node와 Database로 구성되어 있는데, 우리는 총 3개의 Slave Node를 구축하였다. 그림 2는 자료를 이용하여 실행한 예이다. 불쾌감정에 관한 단어들의 빈도를 조사하고 단어의 의미에 따른 가중치를 주어 전체적인 값을 모두 더하여 단어의 개수로 나누어 설정하였다.

Fig. 2. Analysis Result

그림에서 보면 4월 11일부터 4월17일까지의 불쾌 감정 지수가 아주 높았는데, 이 시기가 바로 중간고사 기간이었다. 그리고 5월초에는 매우 낮은데 그 이유는 5월초가 거의 단기 방학과 같은 휴일의 연속으로 장기간 휴학기간이었기 때문으로 추측이 된다. 그러나 단순히 결과를 보았을 때 우리나라 여학생들의 평소 감성은 매우 불안한 요소를 띄고 있음을 알 수 있는데, 이는 대학 진학과 취업 등의 미래에 대한 스트레스가 평소에도 지속적으로 높음을 알 수 있다.

IV. Conclusions

오피니언 마이닝을 위한 기존 감성 분류는 긍정과 부정의 감성에 대한 감성의 강도를 설정하는 것이 일반적이다. 이 방식은 텍스트의 의견의 방향을 결정하기에는 적합하지만 그 의견에 대한 감성의 종류와 같은 상세한 오피니언 마이닝을 수행하기에는 충분한 정보를 제공하지 못한다. 본 논문에서는 감성을 나타내는 단어별 가중치와 함께 감성의 카테고리를 분류하는 방법을 이용하였다. 제안하는 방법을 실험하기 위해 경상북도 상주시의 S여고 1학년 여학생들의 중간고사 전후의 페이스북 글을 이용하였다.

References

[1] Han-Hoon Kang, et. al., “Automatic Extraction of Korean Opinion Words Using PMI-IR and Performance Improvement Method,” Proceedings of KIIS Spring Conf. 2010, vol.20, no.1, pp.318-321, 2010.

[2] Jae-Seok Myunget. et. al., “A Korean Product Review Analysis System Using a Semi-Automatically Constructed Semantic Dictionary,” Jour. of KIISE vol.35, no.6, pp.392-403, 2008.

[3] Andrea Esuli et. al., “SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining,” Proc. of the 5th Conf. on LREC, pp.417-422, 2006.