

신조어 감성사전 기반의 소셜미디어 감성분석 시스템

신판섭*, 오한민^o

^o대진대학교 컴퓨터소프트웨어융합전공

e-mail: psshin@daejin.ac.kr*, ohanmin9@gmail.com^o

Emotion Analysis System for Social Media using Sentiment Dictionary including newly created word

Panseop Shin*, Hanmin Oh^o

^oDivision of Human IT Convergence Engineering, Daejin University

● 요약 ●

오피니언 마이닝은 온라인 문서의 감성을 추출하여 분석하는 기법이다. 별도의 여론조사 없이 감성을 분석 가능하므로, 최근 활발한 연구 분야이다. 그러나 소셜미디어에는 신조어 등이 많이 포함되어 있어 기존 감성분석 시스템으로는 정확한 분석이 어려울 뿐만 아니라, 복합적인 감성에 대한 분석을 내리기에 불리하다. 이에 본 연구에서는 직관적인 감성모형을 제안하고 SNS에서 주목받는 다양한 신조어를 수용한 감성단어사전을 구축한 후, 이를 적용하여 소셜미디어에 나타나는 복합적인 감성을 분석하는 감성분석시스템을 설계한다.

키워드: 감성사전(sentiment dictionary), 감성모형(emotion model), 감성분석(emotion analysis), 오피니언마이닝(opinion mining), 신조어(newly created word)

I. 서론

감성분석은 오피니언 마이닝이라고도 하며, 문서에 포함된 의견, 평가, 태도 및 감성 등을 추출하고 분석하는 분야이며, 별도의 여론조사 없이 감성을 분석할 수 있다[1]. 최근 들어, 소셜미디어의 영향력 증대로 SNS 등에 대한 감성분석이 중요해지고 있으나 SNS에는 신조어, 축약어, 속어 등이 많아 기존 시스템으로 분석하기가 어렵다. 본 연구에서는 이러한 문제점을 해결하기 위해 신조어를 포함하는 감성사전을 구축하고, 이를 활용한 감성분석시스템을 설계한다.

2. 감성사전 구축

본 연구에서 구축한 감성사전은 다양한 감성단어 데이터베이스로써, 감성분석에 최적화된 기본단어 504개, 신조어 143개 등, 총 647개가 등록되어 있다. 감성사전의 기본감성단어는 각종 국어사전에 등재되어 있는 한국어 표준단어를 사용하며, 신조어 감성단어는, SNS와 대중적으로 인기 있는 사이트에서 다양한 신조어를 직접 수집하여 필터링과 정제과정을 거쳐 최종적으로 선별한 143개를 사용하였다. 그 결과는 Table.1과 같다.

II. 연구내용

1. 감성모델 정의

감성모델은 Russell와 Thayer 등의 연구가 대표적이다[2]. 본 연구에서는 Russell, Thayer모델의 기존 X축을 '긍정(Positive, +1)에서 부정(Negative, -1)', Y축은 '긴장(Tension, +1)에서 이완(Relax, -1)' 까지 40개의 격자구간으로 재 정의한다. 다음으로, 기존 18개의 영문 감성단어를 한국어단어로 변환하고, 이를 2차원 감성모델 공간에 배치하여 그 좌표 값을 구한다. 마지막으로 연구[3]에서 제시한 감성범주 11개중 10개(기타 제외)를 추가로 감성모델 공간에 배치한다. 재 정의된 감성모델은 Fig.1와 같다.

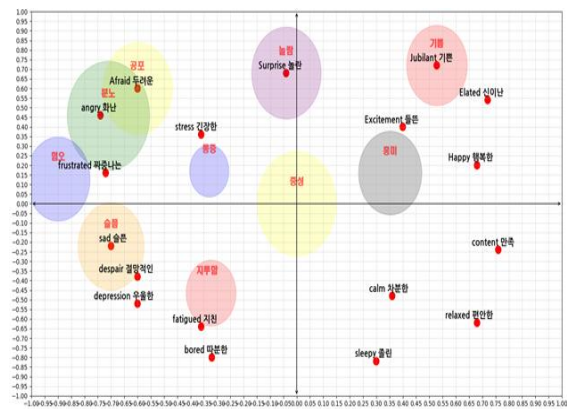


Fig. 1. 제안 감성모델

Table 1. 감성사전 최종선별 결과

수집대상	1차 선별	최종 선별
연구[3]가 정의한 감성단어	504	504
SNS(페이스북, 인스타그램)	32,031	39
국립국어원 신조어자료집	312	59
위키백과	354	45
합계		647

3. 감성분석시스템 설계

감성분석시스템은 소셜미디어의 댓글 따위를 수집하여 감성단어를 추출하고 이를 감성사전과 감성모델에 매핑하여 분석결과를 얻는다. 얻어진 결과는 시각화 모듈을 통해 3가지 결과를 생성한다. Fig.2는 감성분석시스템 동작흐름을 보인 것이다.

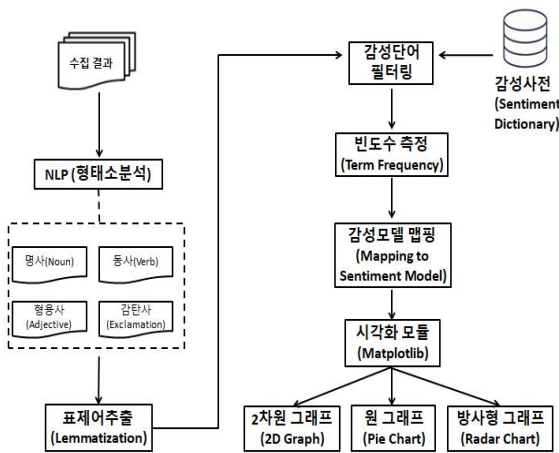


Fig. 2. 감성분석시스템 동작 흐름도

III. 실험결과

분석시스템의 성능 평가를 위해, 네이버 뉴스 기사 중 2016년 3월 13일자 일파고 관련 기사 댓글(11,720건)에 포함된 감성을 분석해 보았다. 분석결과는 Matplotlib[4]을 이용해 시각화하였다.

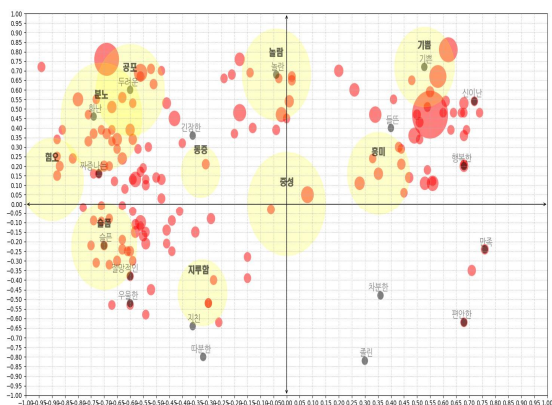


Fig. 3. 기사에 대한 분석결과

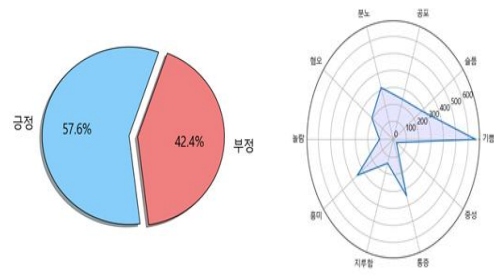


Fig. 4. 기사에 대한 분석결과(기존)

주요 감성단어 군집이 1사분면과 2사분면에 밀집되어 나타나는데, 즉, 군집의 영역이 두 사분면에 양립하므로, 댓글에 인공지능 일파고에 대한 두려움(부정적 감성)과 일파고에게 승리를 거둔 이세돌에 대한 칭찬(긍정적 감성)이 함께 표현되고 있음을 알 수 있다.

IV. 결론

본 논문에서는 143개의 신조어를 등록된 감성단어사전을 구축한 후, 이를 적용한 감성분석시스템을 설계하였다. 본 시스템은 2차원 감성모델에 표현된 감성단어의 분포를 통해 소셜미디어 댓글에 포함된 상세한 감성을 분석할 수 있다. 본 시스템은 선거 전략이나 마케팅 등 다양한 분야에 적용이 가능할 것으로 기대된다.

REFERENCES

- [1] Liu, B., "Sentiment Analysis and Opinion Mining", Morgan and Claypool Publishers, 2012.
- [2] Sunkyung Kim, Panseop Shin, "Emotion Model for Semantic-Based Retrieval of Music Content", Journal of the Korea Entertainment Industry Association, Vol. 9, No. 1, pp. 75~81, Feb. 2015
- [3] Sun Ju Sohn, Mi Sook Park, "Korean Emotion Vocabulary: Extraction and Categorization of Feeling Words", Korea Society for Emotion and Sensibility, Vol. 15, No. 1, pp. 105~120, 2012.
- [4] Matplotlib, <https://matplotlib.org>