

다중 클래스 감성사전 구축에 관한 연구

황현석

한림대학교 경영학과

e-mail:hshwang@hallym.a

A Study of Building a Mult-Class Sentimental Lexicons

Hyun-Seok Hwang

Dept. of Business Administration, Hallym University

요약

감성분석은 텍스트에 포함된 작성자의 감정을 분석하는 텍스트 마이닝의 한 분야이다. 감성분석의 분석대상인 텍스트는 작성을 위해 사용된 언어의 특징을 반영하고 있어 생활상과 비어와 속어 등의 비표준어에 많은 영향을 받는다. 또한 텍스트에 표현된 감성을 정확하게 분석하기 위해서는 감정을 표현하는 단어를 풍부하고 정확하게 파악하여야 한다. 이러한 목적으로 감성단어 사전을 구축하는 연구가 다수 수행되었고 영어로 표현된 감성사전을 활용한 감성분석 연구가 진행되었다. 그러나 한글 감성사전에 대한 구축과 공유는 상대적으로 활발하게 이루지고 있지 않은 실정이다. 본 연구에서는 영문 감성사전을 기반으로 한글 감성사전을 구축하고자 한다. 감성 단어를 긍정/부정/중립과 같은 극성을 파악하는 선행 연구의 한계를 극복하고자 감성분류에 대한 선행연구를 활용하여 8개 감성단어 분류에 따라 8개 감성사전을 구축하였다. 이를 통해 텍스트에 나타난 감정을 보다 정밀하게 분석할 수 있을 것으로 기대된다.

1. 서론

네트워크 및 스마트 장치의 발전과 확산으로 인터넷 공간에서 제품이나 서비스에 대한 리뷰를 표현하는 것이 용이해지고 있다. 온라인 리뷰는 게시자의 의견과 생각을 전달하기 때문에 오피니언 마이닝이나 감성분석을 위한 데이터로 널리 활용되고 있다. 오피니언 마이닝은 텍스트에 나타난 대중의 견해를 긍정/부정/중립 등의 극성으로 나타내는 분석법을 말하며 감성분석은 텍스트에 사용된 감성 단어를 활용하여 게시자의 감정을 읽어내는 분석법을 말한다. 이 두가지 분석법은 모두 텍스트를 분석하는 텍스트 마이닝의 대표적인 분석 방법이다.

본 연구에서는 감성분석에서 기초가 되는 감성사전(Sentimental lexicon)을 구축하는 것을 목표로 한다. 선행연구에서 영어로 작성된 감성사전은 다수 있으나 한글 감성어 사전에 대한 연구를 소수만이 존재하며 대중이 이를 구축하는 체계적인 방법도 부족한 실정이다.

2. 선행연구

2.1 감정 분류

감성사전을 구축하기 위해서는 감정에 대한 이해가 필수적이다. 인간의 감정을 탐구하기 위해 많은 인간 언어 연구가 수

행되었다. 감정 언어를 찾는 한 가지 접근 방식은 소수의 기본 감정을 기초로 기본 감정이 혼합된 다수의 이차 감정을 정의 하는 것이다 (Plutchik, 2004).

<그림 1>은 Plutchik의 Wheel of Emotion (Plutchik, 1980)을 보여준다.

Plutchik의 Wheel of Emotion 모델에 나타나는 8 차원 감정은 슬픔, 놀라움, 공포, 감탄, 엑스터시, 경계, 분노 및 혐오이며 반대편에 위치한 감정은 반대되는 감정을 나타낸다.



[그림 1] Plutchik의 Wheel of Emotion

2.1 감정사전 구축

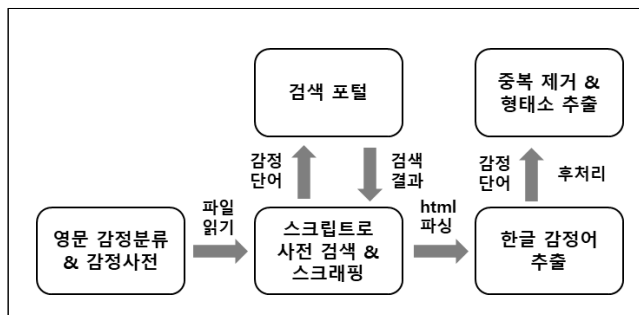
감성분석의 정확성은 감성사전의 정확성과 풍부한 어휘에 비례하며 비속어, 이모티콘과 같은 비표준어를 포함하는 정도에 따라 텍스트 사용자의 일상 언어를 정확히 분석하는 능력이 결정된다. 감성사전은 감성어를 분류하여 범주별로 정리한 어휘집을 말한다. 가장 널리 사용되는 감성사전은 긍정, 부정, 중립의 3가지 범주로 분류된 사전이며 이는 문장의 극성을 파악하는데 사용된다. 예를 들면 특정 주제에 대한 신문 기사를 분석하여 특정 기사가 해당 주제에 긍정/부정/중립 중 어떤 논조로 작성되었는지 파악하는 연구에 사용될 수 있다. 또한 제품이나 서비스의 사용후기를 분석하여 고객의 만족도를 분석하는데도 사용될 수 있는 접근법이다 (Hwang shim and Choi, 2015).

감정의 다양한 분류에 비해 긍정/부정/중립의 극성은 텍스트의 감정을 분석하기에 다소 부족하며 이를 극복하기 위해 다양한 감정 분류별로 감성사전을 구축하는 것이 필요하다.

3. 연구내용

본 연구에서는 영어로 구축된 감성분류와 감성사전을 활용하여 한국어 감성사전을 구축하는 것을 수행하였다. 이를 위해 다음과 같은 일을 단계별로 수행하였다.

- 1단계 : Plutchik의 Wheel of Emotion 모델에 나타난 8차원 감정차원에 해당하는 단어를 수집
- 2단계 : R 스크립트를 통해 해당 단어를 네이버 영한사전에 검색 & 스크래핑
- 3단계 : html 파싱 및 한글 단어 추출
- 4단계 : 사전별 후처리(사전별 중복제거, 형태소 추출) 시행



[그림 2] 연구 프레임워크

4. 결론

본 연구에서는 한국어 다중 클래스 사전을 구축하기 위해 영문 감성사전과 R을 활용하여 사전을 구축하는 연구를 진행하였다. 구축된 감성사전은 과거 긍정/부정/중립으로 파악되는 텍스트의 극성분석의 한계를 넘어서 다양한 감정의 차원으로

텍스트를 분석하는 연구에 활용될 수 있을 것으로 기대된다. 연구의 한계점으로는 영문 감성어의 번역으로 표현되기 힘든 한국어의 고유한 감정어나 비어, 속어 등에 대한 보완으로 정밀한 사전 구축이 필요하다는 점이다.

참고문헌

[1] Hwang, Hae J., Hye R. Shim, and Choi, Jun H. , “Exploration of User Experience Research Method with Big Data Analysis: Focusing on the Online Review analysis of Echo”, Journal of the Korea Contents Association, 제 16권 8호, pp. 517-528, 2016년.
 [2] Plutchik, Robert, “Emotions and Life: Perspectives From Psychology, biology, and Evolution,” Korea: hakjisa, 2004년.