

도서 호감도 분석을 위한 감성어 사전구축 방안

이태석, 이수명, 강승식
한국과학기술정보연구원, 아르고넷, 국민대학교
tsyi@kisti.re.kr, smlee0818@argonet.co.kr, sskang@kookmin.ac.kr

Building Emotional Dictionary to Analysis a Good Feeling of a Book

요 약

감성은 개인적인 생활경험을 통해 표현되며 동일한 감정상태와 정보자극을 주더라도 다른 감성이 발생할 뿐만 아니라 개인, 사회, 문화 요인에 따라서 크게 변한다. 따라서 다른 영역의 감성과 도서에 대한 감성이 같지 않기 때문에 별도의 감성 사전 구축이 필요하다. 구축된 감성사전은 비슷한 성향의 도서와 사람을 묶어 추천해 주는데 활용할 수 있다. 감성 사전 구축을 위한 원천 정보로 네티즌이 책을 읽고 호감도와 함께 짧은 문장으로 쓴 소감을 활용하였다. 감성분석에서 가장 기본이 되는 분류는 긍정과 부정으로 나누는 것이다. 하지만, 실제로 도서를 추천하기 위해서 긍정과 부정으로만 구분하는 것은 충분하지 않다. 따라서 본 연구에서는 도서에 대해서 감성을 긍정과 부정의 호감정도와 감성의 활성도를 조합한 8개의 감성으로 분류하고 각각의 지수를 함께 산출하여 감성어 사전을 구축하고 활용하는 방안을 제시하였다.

주제어: 감성어 분석, 도서 100자평, 긍정부정 호감지수, 각성이완 활성지수

1. 서론

감성은 감각 및 정보자극을 받은 사람이 표출되는 것이다. 감성의 표현은 개인적인 생활경험을 통해 각기 다르게 표현되며, 동일한 감각 및 정보자극을 주더라도 서로 다른 감성이 발생할 수 있다[1].

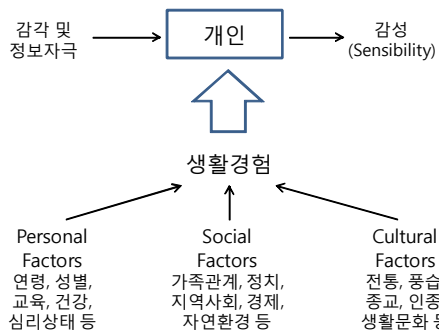


그림 1. 개인의 감성발생과 요인

감성을 다룰 때, 고려할 사항은 객관적인 감성과 개인적 상태 및 변수에 따른 주관적 감성이 함께 섞이는 것이다. 개인적인 요인으로서는 연령, 성별, 교육, 건강, 심리상태 등이 있다. 그러므로 감성은 개인마다 조금씩 다르게 표현될 수 있음을 충분히 감안해야 한다. 공통된 요소를 중심으로 감성을 처리하고 이해하기 위해 사회적 요인과 문화적인 요인을 이용한다. 사회적, 문화적 요인을 통해 공통된 개인 감성을 식별할 수 있기 때문이다.

감성을 분석 활용하는 분야가 최근에 매우 다양하게

늘어가고 있고 새로운 가치로 조명되고 있다. 하지만, 아직 정확도나 활용가치에 대한 기준은 명확하지 않다. 최근, 영화에 대한 감성과 여러 종류의 온라인 콘텐츠에 대한 감성정보를 추출하여 추천에 활용하는 사례가 있지만, 감성정보는 그 영역에 따라 조금씩 다른 감성을 보여주기 때문에 영화나 온라인 콘텐츠의 감성어 사전을 그대로 다른 영역에서 활용할 수 없다. 결국 개별 영역에 맞는 감성어 정보를 구축하여야 한다. 영화의 경우에는 온라인에서 수집할 수 있는 영화평을 이용한다. 마찬가지로 도서에 대한 감성을 분석하기 위해서 네티즌이 책을 읽고 호감도와 함께 짧은 문장으로 쓴 소감을 수집하는 것이 필요하다. 본 연구에서는 알라딘 사이트에 올라온 네티즌의 100자평을 사용하였다. 100자평은 도서를 읽고 짧은 문장으로 솔직한 느낌을 별점과 함께 작성된 정보이다.

도서의 감성정보를 분석하여 감성사전을 구축하면, 비슷한 성향의 도서와 사람에 대해 짝을 지어 추천하는 서비스에 활용할 수 있다. 하지만, 감성사전을 단순히 긍정과 부정으로만 구분하는 것은 추천서비스에서 활용도가 떨어진다. 따라서 본 연구에서는 도서에 대해서 감성을 긍정과 부정의 호감정도를 식별하고 감성의 각성과 이완의 활성정도를 조합한 8개의 감성으로 분류하는 감성 표현 및 인지 모델을 정의하였다. 이 분류를 기반으로 감성어 사전을 구축하고 활용하는 방안을 보였다.

2. 관련 연구

감성분류와 관련한 연구는 대부분 연구 목적으로 만들어졌다. 고감도 오감 센싱을 통한 감성변화를 직접 입력받아 정량화하는 것으로 이것은 생체신호기반 감성변화를 측정하는 것이다. 감성변화 대상, 환경 요인은 사진이나 소리를 이용하였다. 이러한 초기 연구를 기반으로 자연어처리기술을 이용하여 문장의 감성을 추출하고, 색인어로 이용하는 마이닝으로 발전하고 있다.

대표적 감성요소로서 “유쾌-불쾌, 각성-이완, 타인중심-자기중심” 이용하여 분류하였다. “유쾌-불쾌”는 좋은 느낌과 나쁜 느낌으로 긍정과 부정으로 볼 수 있으며, 표정이나 언어로 표현한다. “각성-이완”은 감정의 정도를 표현하는 것으로 강한 감정분출과 약한 감정분출의 정도로 표시되며 주로 뇌파의 활성도로 측정한다. “타인중심-자기중심”은 남을 배려하거나 자신의 욕구 충족에 따라 나타나는 감정을 분류하는 것이다. 표 1.에 감성 분류연구에 대한 내용을 정리하였다 [2][4][5][6][7].

표 1. 감성 분류 기준 연구

연구자	연구명	연구 내용
Plutchik (1980)	Emotion: Theory, research, and experience	- 감정 휠(Wheel of Emotions) - 8개의 기본 감정 - 감정의 조합과 대립 - 2차, 3차 복합 감정 표현
Russell (1980)	A Circumplex Model of Affect	- 정서심리학자 James A. Russell의 정서모델 - 2차원 모델 - 유쾌/불쾌, 활성화/비활성화
Watson & Tellegen (1985)	Toward a consensual structure of mood	- Russell의 정서모델에 기반한 확장된 모델 제시 - 2차원 모델 - Positive Affect, Negative Affect
박인조 (2005)	한국어 감정단어의 목록 작성과 차원 탐색	- 감정단어 목록 작성 - 현대 한국어 어휘빈도 자료집 (1998) 활용 - 원형성/친숙성/쾌 - 불쾌/활성화 분석 및 유사성 평가(불쾌, 고탈성 > 유쾌, 저활성)
이준웅 (2008)	정서 단어 분류를 통한 정서의 차원 및 위계적 범주에 관한 연구	- 정서의 위계적 구조, 정서의 구성 차원 - 3차원 구조 제시 (유쾌/불쾌, 각성/이완, 타인중심/자기중심)

3. 감성 인지 모델

다양한 개인감성 인지 모델중 단순성, 실용성, 공통성, 표현력을 고려하여 도서에 대한 감성분류를 어떻게 할지 검토하였다. 단순성은 맞춤형 도서정보 서비스에 적용될 때 사용자들이 쉽게 이해할 수 있어야 하며, 전

문가가 아닌 일반인이 보기에 직관적이어야 하는 것을 의미한다. 실용성은 감정의 정도를 쉽게 측정하고 인식할 수 있어야 하며, 사람의 피드백(feed-back)을 받기에 적절해야하는 것을 의미한다. 공통성은 기존의 감성 연구들에서 공통적으로 언급하는 기본 요소를 포함해야하는 것을 의미한다. 표현력은 모든 감성 어휘를 모델에 맵핑할 수 있어야 하는 것을 의미한다.

단순성, 실용성, 공통성, 표현력 등 4가지 요소를 고려할 때 그림 2. James A. Russell의 감성 모델이 적절하다고 판단하였다. “유쾌-불쾌”와 “각성-이완”의 두 축으로 구분하여 8가지의 감성을 표현할 수 있다. 각 축으로 감성 점수를 부여함으로써 점수화 하여 쉽게 활용할 수 있는 장점이 있다. 마주보는 각 감성은 서로 상반되는 감성이 위치해있는 특성이 있다.



그림 2. James A. Russell 감성 모델

도서를 읽은 후 감성표현을 James A. Russell의 감성 모델에 기반으로 하여 표 2.와 같이 정리하였다.

표 2. 도서 감성표현

호감	기운	감성 표현	상세 감성표현
중립	강함	힘나는	벽찬, 감동적인
긍정	강함	신나는	활기차고 기분이 좋은
긍정	중립	즐거움	미소 짓게 하는, 가벼운 기쁨
긍정	약함	평온한	잔잔하게 위로가 되는
	중립	평범한	별 느낌 없음 (너무 쉬운... 보통인)
부정	강함	화나는	격하게 놀람, 두려움, 반감, 불만, 짜증나는
부정	중립	슬픈	눈물이 나는
부정	약함	우울한	기분이 나쁜
중립	약함	나른한	따분하고 졸림

4. 감성사전 구축을 위한 감성 지수 계산

감성 사전 구축에 대한 대표적 해외 사례는 SentiWordNet이다. SentiWordNet은 긍정 감성지수, 부정 감성지수, 중립성 지수로 구분하여 0 ~ 1의 값을 부여하여 구축한 감성 사전이다. 실제로는 단어의 긍정과 부정을 주로 표현하는 방법이다[3]. 여러 가지 뜻을 가지는 단어는 번호를 붙여서 각 의미마다 감성 점수를 부여하고 있다.

도서 감성지수의 계산은 식 1. 과 같이 호감도가 높은 문서와 호감도가 낮은 문서에서 상대적으로 출현한 단어의 빈도로 표현하며 -1 ~ 1의 값을 가진다. 활성 지수는 식 2. 와 같이 문서의 각성 이완의 정도에 따라 상대적으로 출현한 단어의 빈도로 표현하면 -1 ~ 1의 값을 가진다. 감성 지수의 중립상태에 대한 임계값은 -0.3 ~ 0.3 으로 설정하여 그림 3. 과 같이 도서 감성 인지 모델을 만들었다.

$$Sentiment\ Score_{호감}(t) = \frac{Count_{긍정}(t) - Count_{부정}(t)}{\max[Count_{긍정}(t), Count_{부정}(t)]}$$

식 1. 호감 지수 계산식

$$Sentiment\ Score_{활성}(t) = \frac{Count_{각성}(t) - Count_{이완}(t)}{\max[Count_{각성}(t), Count_{이완}(t)]}$$

식 2. 활성 지수 계산식

t = 단어
 Count_{긍정}(t) = 단어 t를 포함한 긍정 문서의 수
 Count_{부정}(t) = 단어 t를 포함한 부정 문서의 수
 Count_{각성}(t) = 단어 t를 포함한 각성 문서의 수
 Count_{이완}(t) = 단어 t를 포함한 이완 문서의 수

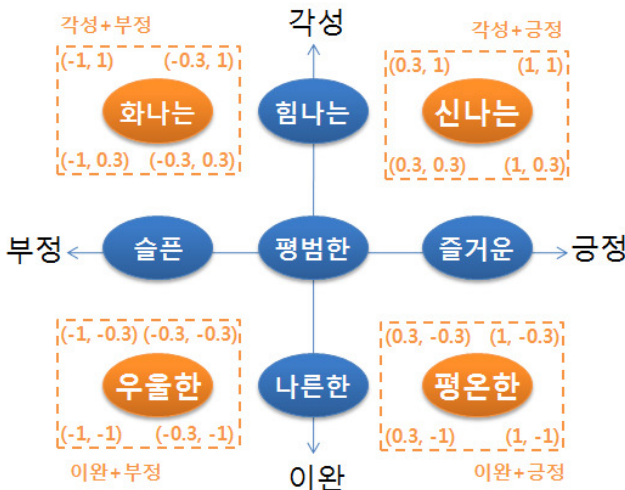


그림 3. 도서 감성 인지 모델

감성지수의 활용은 도서평에 포함된 용어의 감성지수의 함으로 어떤 글의 감성 극성을 판단할 수 있다. 양수이면, 긍정(신나는, 즐거운, 평온한) 느낌과 각성(화나는, 힘나는, 신나는) 느낌으로 해석한다. 음수이면, 부정(화나는, 슬픈, 우울한) 느낌과 이완(우울한, 나른한, 평온한) 느낌으로 해석한다. 두 지수를 조합하여 8가지

감성 분류에 따른 독자의 성향이나 도서의 감성을 예측하거나 비교하고 추천할 수 있다.

5. 감성 사전 구축

감성사전을 구축하기 위해서는 다수의 독자가 책을 읽고 감정을 표현한 문장과 정도를 나타내는 지수가 필요하다. 이를 위해 알라딘 사이트에 네티즌이 올린 100자평을 수집하였다. 이 사이트에는 많은 량의 도서가 있으며, 이 중에서 독자층이 비교적 넓다고 여겨지는 소설/시/희곡, 경제/경영, 인문학 분야의 추천도서에 대한 100자평과 평점을 55,000여건 수집하였다. 100자평의 평점은 2, 4, 6, 8, 10의 다섯 단계로 나뉘지는데, 각 단계별로 2점 1,231건, 4점 1,444건, 6점 3,928건, 8점 13,711건, 10점 34,686건을 수집하였다.

독자가 도서를 구입할 때는 관심이 있는 도서를 구입하는 것이 일반적인 경향이다. 그래서 수집된 서평은 만족도가 높은 8, 10점이 많으며 만족도가 낮은 2, 4점은 적었다. 따라서 각 단계별로 균형을 맞추기 위해 수집된 100자평 중에서 1,000건씩을 선정하여 총 5,000건의 100자평을 사용하여 감성사전을 구축하였다.

감성사전의 구축은 선정된 100자평을 R을 사용하여 회귀분석 기법에 따라 구축하였다. 회귀분석 기법에는 선형 회귀분석, 라쏘 회귀분석, 릿지 회귀분석, 엘라스티넷이 있으며 각 방법을 모두 시도한 후 최종적으로 엘라스티넷에 의한 결과로 감성사전을 구축하였다. 이렇게 하여 구축된 사전은 긍정어 사전과 부정어 사전으로서 호감도 사전이 된다.

본 연구에서는 호감도 사전과 함께 활성도 사전이 구축되어야 한다. 활성지수는 감정의 강도를 표현한다고 할 수 있다. 평점이 2점과 10점은 감정의 강도가 크다고 할 수 있으며 6점은 강도가 작다고 볼 수 있다. 또한 강도는 “매우, 아주, 조금” 과 같이 부사어로 표현하는 경우가 많으므로 부사어만으로 활성도 사전을 구성하였다. 구축 방법은 호감도 사전과 동일하게 R과 회귀분석을 활용하였으며, 2점과 10점 100자평 3,000건, 6점 100자평 3,000건을 사용하였다. 구축된 활성도 사전은 각성어 사전과 이완어 사전으로 구성된다.

6. 결론

도서에 적합한 감성표현을 James A. Russell의 감성 모델에 기반을 두어 정의하고 호감지수와 활성지수를 계산하는 방법과 함께 활용하기 위한 도서 감성 인지 모델을 보였다. 제시한 감성 인지 모델에 따라 사전을 구축하고 도서의 감성을 해석하고 독자와 도서를 추천할 때 보다 세분화된 감성을 이용할 수 있도록 하였다.

그러나 감성어 사전을 구축하기 위해 사용한 100자평은 평점이 낮은 부정 서평이 아주 부족하여 전체적으로 부족한 데이터로 사전을 구축하였다. 또한 활성도 사전은 평점이 활성지수와 완전히 일치한다고 보기에 약간의 무리가 있다

향후 도서에 대해 좀 더 객관적으로 호감도와 각성도

의 감성을 표현한 문장과 지수를 수집하고 충분한 데이터를 기반으로 사전을 구축하여 감성분석의 결과를 평가하는 연구가 필요하다.

참고문헌

- [1] 이구형, "감성과 감정의 이해를 통한 감성의 체계적 측정 평가", 한국감성과학회지, 제1권, 제1호, pp. 113-122, 1998.
- [2] James A. Russell, "A Circumplex Model of Affect", Journal of Personality and Social Psychology, Vol. 39, No. 6, pp. 1161-1178, 1980.
- [3] SentiWordNet, <http://sentiwordnet.isti.cnr.it>
- [4] Robert Plutchik, "Emotion: Theory, Research, and Experience", Published by Academic Press, 1985
- [5] Watson David, Tellegen Auke, "Toward a consensual structure of mood", Psychological Bulletin, Vol. 98(2), pp. 219-235, 1985.
- [6] 박인조, 민경환, "한국어 감정단어의 목록 작성과 차원 탐색", 한국심리학회, 제19권, 제1호, pp. 109-129, 2005.
- [7] 이준용, 송현주, 나은경, 김현석, "정서 단어 분류를 통한 정서의 구성 차원 및 위계적 범주에 관한 연구", 한국언론학보, 제52권, 제1호, pp. 85-116, 2008.