

---

저자 (Authors)	석지민, 윤지현, 정옥란 Ji-Min Seok, Ji-Hyeon Yoon, Ok-Ran Jeong
출처 (Source)	<a href="#">한국정보과학회 학술발표논문집</a> , 2018.12, 1510-1512(3 pages)
발행처 (Publisher)	<a href="#">한국정보과학회</a> The Korean Institute of Information Scientists and Engineers
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07613999">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07613999</a>
APA Style	석지민, 윤지현, 정옥란 (2018). 소셜 미디어 데이터 기반 감정 분석 시스템. 한국정보과학회 학술발표논문집, 1510-1512
이용정보 (Accessed)	가천대학교 203.249.***.201 2019/09/29 19:00 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

# 소셜 미디어 데이터 기반 감정 분석 시스템

석지민<sup>○</sup> 윤지현 정옥란

가천대학교 소프트웨어학과

jmandy@naver.com, jihyeonyoon0120@gmail.com, orjeong@gachon.ac.kr

## Emotion Analysis System based on Social Media Data

Ji-Min Seok<sup>○</sup> Ji-Hyeon Yoon Ok-Ran Jeong

Department of Software, Gachon University

### 요 약

페이스북, 트위터와 같은 소셜 미디어를 통해 자신의 생각과 감정을 표현하는 대중들의 데이터가 방대해지고 있다. 이 데이터를 토대로 특정 대상에 대한 대중의 감정을 분석하여 음악 추천, 마케팅, 선거 결과 예측 등에 사용하고 있다. 본 논문에서는 가장 기본적이고 보편적으로 쓰이는 긍정·부정 감정 분석에서 나아가 좀 더 세분화된 감정 분석 기법을 제안하고자 한다. 이를 위해 기존의 ANEW와 Thayer 감정 모델을 이용하여 감정 영역별 수치화 작업을 수행하여 감정 모델 및 사전을 제작하고 감정 분석 시스템을 설계 및 구현하였다. 또한 실험을 통해 그 유효성을 검증하였다.

### 1. 서 론

2012년에 포브스가 발표한 ‘10년 전에는 존재하지 않았던 유망 직업 10선’에 CLO(Chief Listening Officer)[1]가 자리했다. CLO는 소셜 미디어에 대중들이 올린 자사 브랜드, 제품, 서비스에 관한 글에서 유익한 정보를 추출하여 마케팅에 이용한다.

대중들은 소셜 미디어를 통해 자발적으로 자유롭게 자신의 생각과 감정을 표현한다. 설문 조사나 인터뷰에 비해 더 솔직한 대중들의 생각을 알 수 있기에, 소셜 미디어의 데이터가 대중을 파악할 수 있는 지표로 주목받고 있다. 실제로 특정 대상에 대한 대중들의 감정을 분석하여 마케팅, 선거 결과 예측에 사용하고 있다.

가장 기본적이고 보편적으로 긍정·부정 감정 분석이 쓰이는데, 본 논문에서는 좀 더 세분화된 감정 분석을 했다. 감정을 세분화하기 위해서는 그 과정에서 수치화, 정량화할 수 있는 이

론적 근거가 필요한데, 이에 가장 적합한 것이 ANEW와 Thayer 감정 모델이었다. 또한, 한국어 기반 텍스트를 영어 기반의 모델들에 적용하여 감정을 분석하기에 그 절차를 상세히 나타냈다.

### 2. 관련 연구

본 논문과 관련이 있는 연구로 감정 모델과 ANEW가 있다.

#### 2.1 감정 모델

대표적인 감정보호에는 Thayer's model과 Plutchik's wheel of emotions가 있다.

##### 2.1.1 Thayer's model

J. A. Russe과 R. E. Thayer[3]은 사람의 감정을 valence (긍정성/부정성의 정도)와 arousal (감정의 에너지의 정도) 값으로 정량화 하여 2차원 평면에 나타냈다. 그림 1은 가로축이 valence, 세로축이 arousal 값인 좌표 평면이다. 중점의 값은 arousal, valence 각 5의 값을 나타낸다. 축의 최댓값은 10이다. 축을 기준으로 arousal 값과 valence 값이 모두 5 이

본 논문은 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 기초연구사업지원과 과학기술정보통신부 및 정보통신기술진흥센터의 SW중심대학지원사업의 연구결과로 수행되었음(NRF-2015R1C1A2A01051729, 2015-0-00932-003).

상이면 1사분면으로, 행복과 즐거운 감정을 나타낸다. valence 값이 5 미만이고 arousal 값이 5 이상이면 2사분면으로, 걱정과 화난 감정을 표현한다. valence 값과 arousal 값이 모두 5 이하면, 슬프고 우울한 감정을 나타낸다. valence 값이 5 이상, arousal 값이 5 이하면 4사분면으로, 평온하고 안정된 감정을 나타낸다.

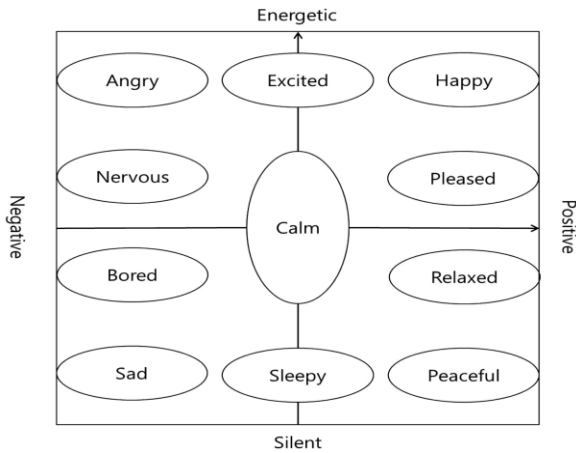


그림 1 Thayer's model

### 2.1.2 Plutchik's wheel of emotions

Plutchik's wheel of emotions[4]는 8가지 기본 감정(기쁨, 신뢰, 공포, 놀람, 슬픔, 혐오, 분노, 기대)을 기준으로 분류하였다. 3차원 원형 모델을 이용하여 세로축은 강도를 나타내고, 원은 감정 사이의 유사성 정도를 보여준다. 그림 2에서 보듯 8개 영역에 8개의 기본 감정을 표시하였으며, 서로 마주하는 구역의 짝은 서로 반대 개념을 표현하였다. 분류된 기준으로 감정의 세기에 따라 다양한 감정을 표현할 수 있는 대표적인 감정 모델이다.

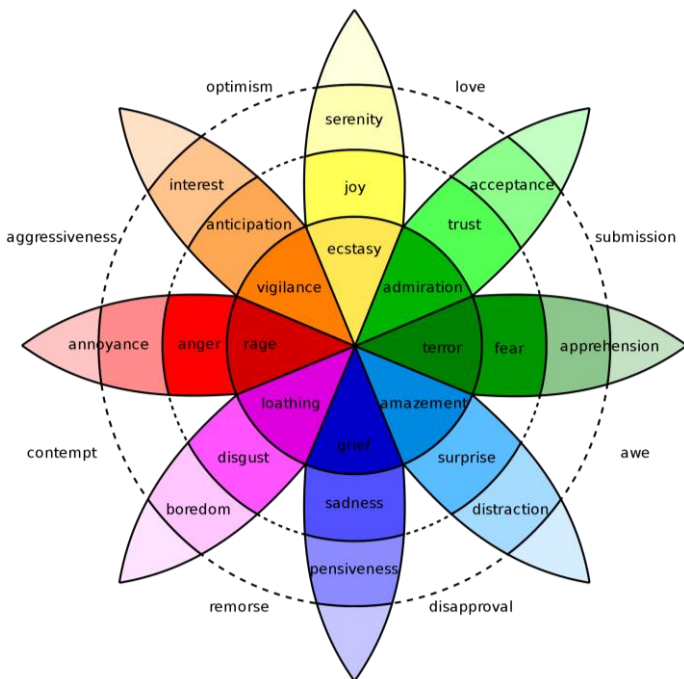


그림 2 Plutchik's wheel of emotions

### 2.2 ANEW

ANEW(Affective Norms for English Words) [5]는 설문 조사를 통해 영어 단어에 대한 감정을 정리한 일종의 감정 단어 사전이다. ANEW는 총 1,034개의 단어에 대한 감정을 valence, arousal, dominance 값을 기준으로 수치로 표현했다.

하나의 단어에 대해 valence, arousal, dominance 각각은 1~10점 사이의 값을 가진다. valence는 긍정적일수록 큰 값을, 부정적일수록 낮은 값을 갖는다. arousal은 활동적일수록 높은 값을, 정적일수록 낮은 값을 가진다. dominance는 우세일수록 높은 값을, 열세일수록 낮은 값을 갖는다.

설문자들이 각 단어를 보고 느낀 감정을 valence, arousal, dominance 세 영역에 대해 1~10점 사이의 점수를 매기고, 설문자들이 매긴 점수의 전체 평균이 그 단어의 수치로 정해진다. 이렇게 만들어진 ANEW는 텍스트에서 감정을 추출하는 데 사용될 수 있다.

### 3. 감정 분석 시스템

감정 분석 시스템 구축을 위해 선택한 감정 모델에 대해 알아보고, 감정 사전을 제작한 방법과 감정 분석하는 과정에 대해 단계별 방법은 다음과 같다.

#### 3.1 감정 모델

본 논문에서는 감정 분석을 위해 Thayer 감정 모델을 선정하였다. 감정을 세분화하기 위해서는 수치적 기준이 필요한데, 각 단어에 대해 ANEW에서 수치화한 valence, arousal 값이 Thayer 모델의 X축, Y축에 부합했기 때문이다.

#### 3.2 감정 사전 제작

감정 사전은 ANEW를 기반으로 보충하여 구축하였다. ANEW는 1,034개의 단어에 대한 감정을 valence, arousal, dominance 3개의 값으로 정량화한 것이다. 이 중 Thayer 감정 모델에 적용할 수 있는 valence와 arousal 값을 활용하였다.

감정 사전을 확장하기 위해 WordNet을 활용하여 ANEW 사전의 단어 리스트에 해당하는 동의어를 추가하였다. 또한, SNS 상에서 일상 언어의 사용 빈도가 높으므로 국립국어원에서 선정한 가장 많이 사용되는 단어 6,000개를 영어로 번역하여 추가하였다.

#### 3.3 감정 분석 단계

감정 분석단계는 크게 6단계로 나뉜다.

첫 번째는, 한글 문장을 트위터에서 만든 오픈소스 한국어 처리기(Official Fork of twitter-korean-text)를 이용하여 한글 문장을 정규화한다. 이 과정을 거치면, 다음(입니답ㅋㅋ->입니다 ㅋㅋ, 샤롱해->사랑해)과 같이 문장이 정규화된 다.

두 번째는, 정규화된 문장을 영어로 번역하는 것이다. 번역을 위해 구글에서 제공하는 Cloud Translation을 사용하여 한글 문장을 영어 문장으로 번역하였다.

세 번째는, 번역된 문장을 토큰화하는 과정이다. 번역된 문장을 띄어쓰기 단위로 토큰화했다.

네 번째는, 토큰화된 단어들을 원형으로 변형하는 단계이다.

이 과정에서 명사의 복수형(dogs)은 단수형(dog)이 되고, 원형이 아닌 동사(played, plays, playing)는 동사원형(play) 형태로 변형시킨다. 이렇게 변화시키는 이유는 ANEW 감정 사전에 정리된 단어들이 원형의 형태로 저장되어 있기 때문이다.

다섯 번째는, 원형으로 저장된 각 단어를 자체 제작한 감정 사전에 매칭하는 단계이다. 감정 사전에는 각 단어의 valence와 arousal 값이 기록되어 있다. 토큰화 된 단어가 이 사전에 포함된 경우, 그 단어에 해당하는 valence와 arousal 값을 저장한다.

마지막은 감정 추출 단계이다. 한 문장에 포함된 모든 단어의 감정 사전 매칭이 끝나면, 각 단어가 가진 valence와 arousal 값의 평균값을 추출한다. 이 두 가지 값의 조합으로 4가지 감정을 표현할 수 있다. 이 두 가지의 값으로 Thayer 감정 모델에 적용해 감정을 추출한다.

### 3.4 감정 분석 시스템 구조도

감정 분석 시스템의 전체 구조도는 그림 3과 같다.

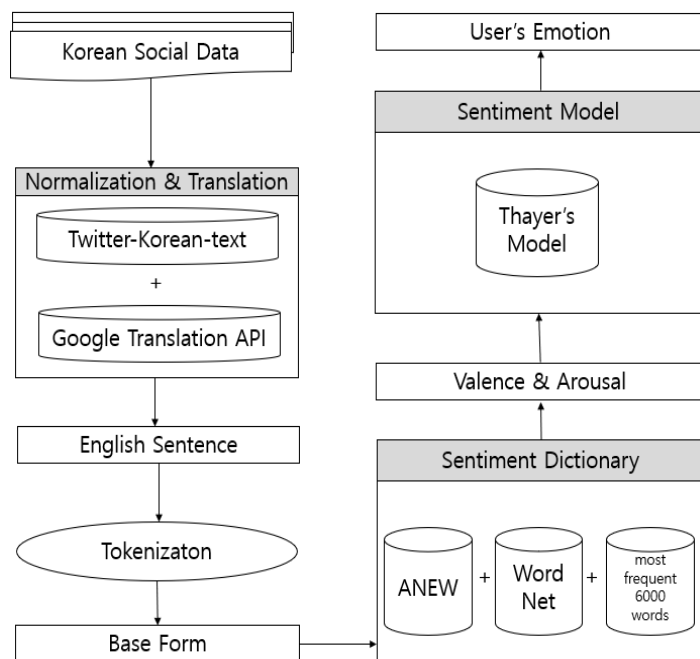


그림 3 감정 분석 시스템 구조도

### 4. 실험

본 논문에서는 트위터 및 페이스북에서 임의로 게시물들 크롤링 한 후, 감정이 드러나는 3,000개의 한글 게시물들 추려냈다. 추려낸 각 게시물들의 예상 감정과 본 논문에서 제시한 감정 분석 기법의 감정 결과 일치 여부를 대조하였다. 본 연구에서 감정 분류의 기준인 4개의 감정(행복, 걱정과 화남, 슬픔, 평온)에 대해 감정 일치 여부를 실험한 결과 각각 78.1%, 75.3%, 79.6%, 75.1%의 정확율을 보여주었다. 본 실험의 평균 정확율은 77%였으며, 다음 그림 4에서 그 결과를 보여준다.

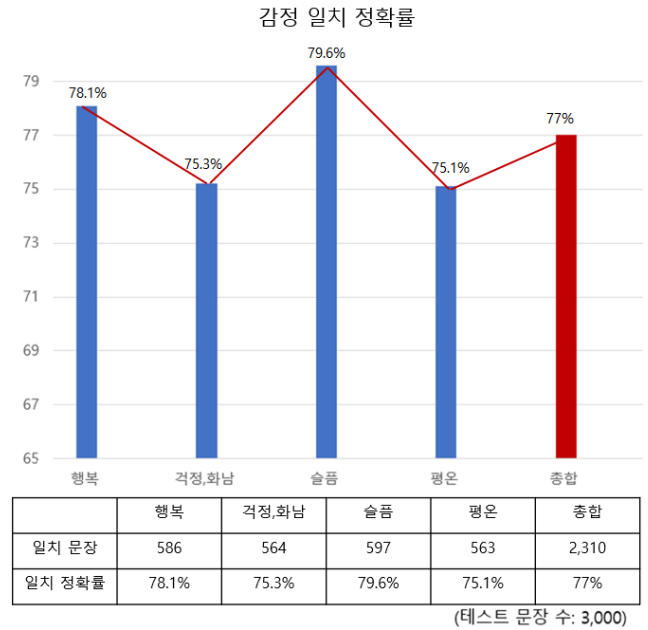


그림 4 실험 결과

### 5. 결론

본 연구를 통해 소셜 데이터의 감정을 분석하는 시스템을 직접 구축해보고, 긍·부정 감정 분석에서 나아가 좀 더 세분화된 감정 분석을 실시하였다. 그 과정에서 한국어 텍스트 분석 시 영어 기반의 감정 모델과 ANEW를 적용하는 것의 간극을 메우기 위하여 번역 횟수를 최소로 하였다. 한국어로 작성된 소셜 데이터를 영어로 번역하여 그 후에는 영어 기반 감정 모델과 감정 사전을 적용한 것이다.

향후 연구에서는 문장 자체의 의미 분석을 하는 것을 목표로 한다. 단어가 어떤 문맥에서 쓰이는지에 따라 그 의미가 달라지기에 이를 감정 분석에 적용한다면 정확도가 더 높아질 것이라 예상된다.

본 연구에서 진행한 감정 분석 시스템을 이용하여 사용자의 감정을 추출한다면, 이를 사용자 감정 기반 음악 추천, 마케팅, 선거 결과 예측 등에 사용할 수 있을 것으로 기대한다.

### 참고문헌

- [1] IDG Tech Report, [goo.gl/av5bmC](http://goo.gl/av5bmC), 2014.06
- [2] R. E. Thayer, "The Biopsychology of Mood and Arousal," New York: Oxford University Press, 1989.
- [3] Melissa Donaldson, Plutchik's Wheel of Emotions, sixseconds, 2017
- [4] Bradley, M. M., & Lang, P. J., "Affective norms for English words (ANEW): Instruction manual and affective ratings," (Tech. Rep. No. C-1)