

신조어 감성사전 기반의 소셜미디어 감성분석 시스템

Emotional analysis system for social media using sentiment dictionary with newly-created words

저자 (Authors)	신판섭 Pan-Seop Shin
출처 (Source)	한국컴퓨터정보학회논문지 25(4) , 2020.4, 133-140(8 pages) Journal of the Korea Society of Computer and Information 25(4) , 2020.4, 133-140(8 pages)
발행처 (Publisher)	한국컴퓨터정보학회 The Korean Society Of Computer And Information
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09329827
APA Style	신판섭 (2020). 신조어 감성사전 기반의 소셜미디어 감성분석 시스템. 한국컴퓨터정보학회논문지 , 25(4), 133-140
이용정보 (Accessed)	통계청 125.128.71.*** 2020/07/23 15:12 (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.

Emotional analysis system for social media using sentiment dictionary with newly-created words

Pan-Seop Shin*

*Professor, Dept. of Computer Engineering, Daejin University, Pocheon, Korea

[Abstract]

Emotional analysis is an application of opinion mining that analyzes opinions and tendencies of people appearing in unstructured text. Recently, emotional analysis of social media has attracted attention, but social media contains newly-created words and slang, so it is not easy to analyze with existing emotional analysis. In this study, I design a new emotional analysis system to solve these problems. The proposed system is possible to analyze various emotions as well as positive and negative in social media including newly-created words and slang. First, I collect newly-created words and slang related to emotions that appear in social media. Then, expand the existing emotional model and use it to quantify the degree of sentiment in emotional words. Also, a new sentiment dictionary is constructed by reflecting the degree of sentiment. Finally, I design an emotional analysis system that applies a sentiment dictionary that includes newly-created words and an extended emotional model.

▶ **Key words:** Sentiment dictionary, Emotion model, Emotion analysis, Opinion mining, Newly-created words

[요 약]

감성분석은 비정형 텍스트에 나타나는 사람들의 의견이나 성향 등을 분석하는 오피니언 마이닝의 응용 분야이다. 최근에는 소셜미디어에 대한 감성분석이 주목받고 있으나 소셜미디어에는 신조어, 속어 등이 포함되어있어 기존 감성분석으로는 분석이 쉽지 않다. 본 연구에서는, 이러한 문제점을 해결하기 위해, 새로운 감성분석 시스템을 설계한다. 제안 시스템은 신조어, 속어 등이 포함된 소셜미디어에서도 긍/부정 뿐만아니라 다양한 감성분석이 가능하다. 먼저, 현재 소셜미디어에서 많이 나타나는 감성관련 신조어와 속어 등을 수집한다. 그리고 나서, 기존의 감성모델을 확장하고 이를 활용하여 감성단어에 감성정도를 수치화 한다. 또한 감성정도를 반영하여 새로운 감성단어 사전을 구축한다. 최종적으로, 신조어가 포함된 감성사전과 확장된 감성모델을 적용한 감성분석시스템을 설계한다.

▶ **주제어:** 감성사전, 감성모델, 감성분석, 오피니언 마이닝, 신조어

-
- First Author: Pan-Seop Shin, Corresponding Author: Pan-Seop Shin
*Pan-Seop Shin (psshin@daejin.ac.kr), Dept. of Computer Engineering, Daejin University
 - Received: 2019. 09. 24, Revised: 2020. 04. 04, Accepted: 2020. 04. 05.

I. Introduction

감성분석은 오피니언 마이닝의 한 응용 분야로, 텍스트에 나타나는 사람들의 의견이나 성향 등의 주관적 데이터를 분석하는 자연어 처리 기술이다. 오피니언 마이닝은 빅데이터 중 주로 비정형 데이터를 분석하게 되는데, 사람들이 특정 제품이나 서비스에 대해 긍정적인지, 부정적인지를 판별하기도 하고, 정치, 경제, 사회적 문제에 대해 여론의 향방이 어떻게 실시간으로 변화해 가는지 등을 추적하기도 한다. 지금 이 순간에도 누군가가 생산하고 있는 소셜 데이터 속에 존재하는 주요 키워드에 대한 감성을 수치화하여 별도의 여론 조사 없이 감성의 정도를 분석할 수 있고, 감성분석 결과의 시각화를 통해 직관적인 감성을 판단하여 의사결정에 반영이 가능하기 때문에, 자연어처리에서 최근 활발히 연구되고 있는 분야중 하나이다[1]. 특히, 감성분석은 SNS를 주요 대상으로 활용하고 있다. 한국인터넷진흥원의 인터넷이용실태조사(2017)에 따르면 우리나라의 인터넷 이용률 90.3%중, SNS 비율이 68.2%로[2], 기존에는 웹사이트나 카페, 블로그가 여론의 영향을 미쳤다면, 현재는 트위터, 페이스북 같은 SNS의 영향력이 증대되고 있어 SNS의 텍스트데이터에 대한 감성분석이 주목받고 있다. 그러나 SNS에는 신조어, 축약어, 속어 등이 많이 포함되어 있어 기존 감성분석 시스템이나 인공지능기반 학습 시스템으로 분석했을 경우, 그 결과가 부정확하다는 문제점이 발생한다. 또한 기존 오피니언 마이닝은 소셜 미디어에 담긴 사람들의 감성을 단순히 긍정/부정의 비율로만 나타내는 것이 일반적이므로, 복합적인 감성에 대한 자연스러운 예측이나 판단을 내리기에 부족하다. 본 연구는 이와 같은 문제점을 해결하기 위해, 신조어, 속어 등이 많이 포함된 소셜 미디어의 텍스트 데이터를 분석 가능하고, 소셜 미디어에 포함된 긍정/부정의 비율뿐만 아니라 좀 더 구체적인 감성정도까지 표현 가능한 감성분석 시스템을 제안한다. 이를 위해, 현재 소셜 미디어에서 사용빈도가 높은 다양한 신조어와 속어 등을 수집하고, 수집된 결과에서 직접적인 감성과 연관된 단어를 추출하여 기존 감성단어를 포함하는 새로운 신조어 감성사전을 구축한다. 또한 선행연구[3]에서 제안한 감성모델을 좀 더 직관적인 감성표현이 가능하도록 확장하고, 이를 활용하여 감성단어에 감성정도를 수치화한다. 마지막으로, 신조어를 포함한 감성사전과 확장 감성모델을 활용한 감성분석시스템을 설계한다.

II. Preliminaries

1. Opinion Mining

빅데이터의 시대의 소셜 미디어 사용자들은 누군가가 생산한 데이터를 단순히 소비하는데 그치지 않고 적극적인 생산자의 역할을 수행한다. 소셜 데이터 분석은 크게 두 가지를 대상으로 하는데, 데이터베이스에 일정한 형태로 저장할 수 있는 정형 데이터와 문서, 이미지, 영상 등 데이터의 필드가 정해져 있지 않은 비정형 데이터로 구분된다. 빅데이터의 85% 이상이 비정형 데이터이며, 오피니언 마이닝은 주로 비정형 데이터를 분석한다. 오피니언 마이닝은 흔히 감성분석이라고도 불리며, 소셜 미디어 기반의 오피니언 마이닝은 트렌드의 파악, 제품/서비스 평가, 미래 예측 등에 활용된다[4]. 정치, 사회적 이슈뿐만 아니라, 연예나 스포츠 등의 분야에서의 트렌드의 파악은 매우 중요한 이슈이며, 기업의 매출과 직결되는 제품/서비스 평가는 보다 직접적인 활용도를 지닌다. 오피니언 마이닝은 의사결정의 보조 수단이며, 주가나 국가경제위기 예측 등에 다양하게 활용되고 있다[5]. 그러나 기존의 감성분석은 대상 문서에 포함된 감성을 긍정/부정이나 또는 긍정/부정의 비율로만 나타내므로 그 결과가 단순하다. 또한 감성분석에는 감성단어사전의 역할이 중요하나 소셜 미디어 분석에 범용으로 사용가능한 감성사전이 매우 제한적이거나 영단어 위주로 만들어진 것이 대부분이다. 따라서 본 연구에서는 신조어, 속어 등도 포함하고 다양한 감성정도를 수치화한 감성사전을 만들고 이를 적용한 감성분석 시스템을 제안한다.

2. Emotion Model

감성모델이란 주관적인 감성정보를 객관화, 정량화하기 위한 연구의 결과물로, Russell 모델[6]과 Thayer 모델[7] 그리고 Plutchik 모델[8] 등이 감성연구 분야에서 주로 활용되는 대표적 감성모델이다[8]. 그 중, 가장 대표적인 Russell의 감성모델[6]은 심리학을 기반으로 인간의 감성을 2차원 공간에 배치하여 감성의 객관화를 연구하였다. Fig. 1과 같이 x축이 유인성(誘引性) 또는 호감도(好感度)를 나타내며, y축은 각성(覺醒)도를 나타낸다. 두 개의 축을 바탕으로 2차원 평면을 정의하고 여기에 인간의 기본 감성단어 18개를 정의하고 배치하였다. 하지만 Russell의 감성모델에 표현된 기본 감성단어는 한국인의 정서와는 그 중요성 측면에서 약간의 의미적 차이를 보이며 이를 단순히 번역하여 해당 위치에 배치할 경우 한국어 감성단어와는 미묘한 감성표현의 차이 역시 존재함을 알 수 있다. 또한 심리학 기반의 유인성(호감도), 각성도와 같은 표현

은 일반적인 응용분야에 적용하기에 그 의미가 다소 모호하므로 본 연구에서는 기존 감성모델을 변형한 새로운 감성모델을 제안한다. 한편, 연구[10]에서는 한국인의 표준적이고 대표적인 감성 단어를 추출하고 각각의 감성단어의 감성 범주를 규명하였다. 본 연구에서는 연구[10]에서 제시한 감성 범주와 감성표현강도(Intensity Range)를 활용하여 제안한 감성모델에 각 대표감성을 재 정의하고, 이를 추가 배치하여 제안 감성모델을 확장한다.

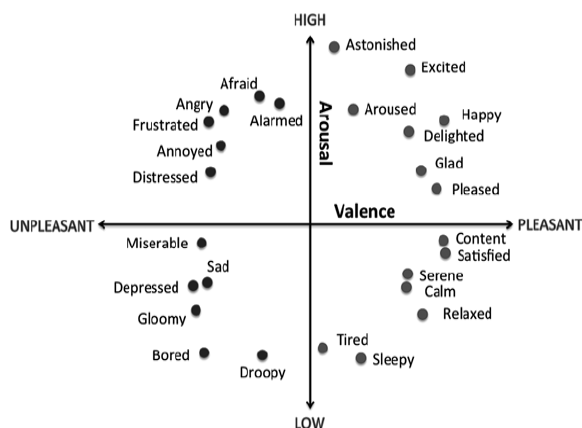


Fig. 1. Russell. Valence-Arousal

III. The Proposed Scheme

1. Extended Emotion Model

본 연구에서는, 기존 Russell, Thayer모델의 X, Y축을 X축은 '긍정(Positive)에서 부정(Negative)까지 +1에서 -1을 부여하였으며, Y축은 '긴장(Tension)에서 이완(Relax)까지 +1에서 -1을 각각 40개의 격자구간으로 재 정의한다. 다음으로, 러셀 모델에서 제시되었던 기존 18개의 영문 감성단어를, 의미에 가장 부합하는 한국어 단어로 변환하고, 이를 한국어의 의미로 재해석하여 2차원 감성모델 공간에 재배치한다. 마지막으로, 연구[8]에서 제시한 대표 감성범주 11개중 10개(기타 감성 제외)를 추출하여 감성모델 공간에 추가 배치하고 집단지성을 활용한 여러 단계의 피드백 과정을 통해 감성단어의 위치를 수정한 후, 최종적으로 고정된 감성단어의 좌표 값을 구한다. 재 정의된 감성모델은 Fig. 2와 같다.

2. Sentiment Dictionary

감성사전은 다양한 감성단어를 포함하는 데이터베이스이며 소셜 미디어 분석의 핵심적 역할을 수행한다. 본 연구를 통해 구축된 감성사전에는 대표감성단어 504개, 사용빈

도가 높은 최신 신조어 143개 등, 총 647개의 감성단어가 등록되어 있다. 감성사전의 대표감성단어는 각종 국어사전에 등록되어 있는 한국어 표준 단어를 기반으로 선별된 단어를 사용하며, 신조어 감성단어는, SNS나 대중적으로 인기 있는 사이트에서 출현빈도가 높고 최신 트렌드를 가급적 반영할 수 있는 다양한 신조어를 직접 수집하여 필터링과 정제과정을 거쳐 최종적으로 선별한 143개를 포함한다.

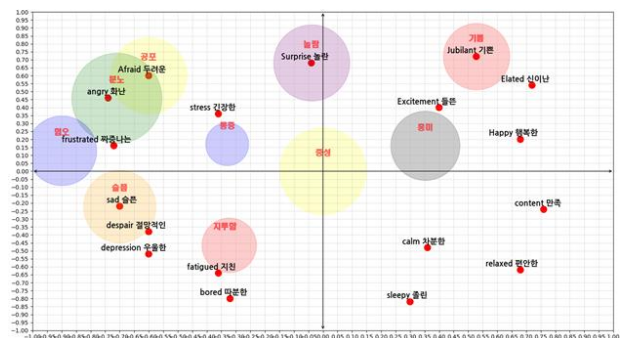


Fig. 2. Proposed Emotion Model

2.1 Basic Emotional Words Selection

기본 감성단어는 연구[9]에서 제시한 감성단어를 기반으로, 한 가지 감성범주를 나타내는 단어 426개, 두 감성범주를 나타내는 단어 72개, 세 감성 범주를 나타내는 단어 6개 등, 총 504개의 감성단어를 선별하였다.

2.2 Additional Word Collection in SNS

신조어 감성단어의 수집을 위해 본 연구에서는 SNS로부터 감성단어를 추출 및 선별하였다. 그 결과는 Table.1과 같다.

Table 1. Collected Words in SNS

SNS	Collection	Post	period	morpheme
Facebook	SNUBamboo	4,468	2017.1.26.~ 2017.7.26	2,676,609
Instagram	Post related Music	51,000	2017.1.1.~ 2017.10.26	2,324,151
total		55,468		5,000,760

다음으로, 페이스북과 인스타그램에서 추출된 형태소 5,000,760개에 대한 분석 및 정제 과정을 거쳐 감성단어를 선별하였다. 먼저, 형태소분석기[11]를 통해 명사, 동사, 형용사, 감탄사 등을 추출한다. 이렇게 추출된 단어는 총 2,324,151개이다. 다음으로, 기존 감성단어와 중복된 단어를 제거하기 위한 필터링 과정을 거친다. 필터링 과정에서는 단어를 기본형으로 변환하여 처리하는 표제어추출

(Lemmatization) 과정이 필요하다. 예를 들어, ‘가소로운’, ‘가소로울’ 이라는 단어는 ‘가소롭다’로 처리한다. 이렇게 필터링을 거쳐 중복된 단어를 제거한 후에, 추출된 단어를 사용 빈도에 따라 워드카운팅을 수행하여 분류하였으며, 그 결과로 32,031개의 단어를 얻었다. 이 단어들 중, 상위에 랭크된 주요 단어와, 비록 빈도수는 적지만 ‘땡큐’, ‘귀르가즘’, ‘땡곡’ 같이 감성단어로서 유의미하다고 판단되는 단어들을 수작업으로 수집한다. 이와 같은 과정을 거쳐 총 39개의 신조어 감성단어를 수집하였다.

2.3 Resource of National Institute of Korean Language

국립국어원은 매년 신조어의 지속적인 조사와 관리를 통해 우리말 자료를 주기적으로 수집, 기록하여 신조어 자료집을 발간한다. 해당 자료는 각종 일간지, 인터넷 매체 등에서 국립국어원이 수집, 분류, 정리한 단어목록이다. 2008년부터 2016년까지 9년간의 신조어자료집에서 총 3,623개의 신조어를 2단계의 선별과정을 거쳐 수집한다. 1차 선별과정에서는 현재 사용빈도가 적거나 잘 사용하지 않는 단어를 제외한다. 예를 들어, ‘나똥’, ‘기모치렌스’, ‘고렘이’ 같은 단어들이 현재 사용되지 않는 대표적인 단어들이며 이러한 것들을 제외하여 총 312개의 단어를 선별하였다. 2차 선별과정에서는 1차에서 걸러진 단어 중, 감성과 직접적인 연관이 있는 단어를 선별한다.

Table 2. Collected Words from Resource of National Institute of Korean Language

Year	# of Words	1 st extraction	2 nd extraction
2016	625	20	8
2015	277	34	19
2014	334	56	15
2013	476	14	4
2012	500	17	4
2010	351	72	3
2009	588	76	6
2008	472	23	3
total	3 623	312	59

이 과정을 거쳐 Table.2와 같이 총 59개의 신조어 감성 단어를 최종적으로 수집하였다.

2.4 List of Internet New Words in Korea from Wikipedia

위키백과의 인터넷 신조어 목록은 대한민국의 네티즌들이 많이 사용하는 단어이나 한국어 사전에 공식적으로 등재되지 않은 신조어들을 담고 있다. 여기에 나열된 354개의 신조어 중, 감성과 직접적인 관련성이 있는 단어를 선별하여 총 45개의 신조어 감성단어를 수집하였다. Table.

3 은 여러 단계를 거쳐 최종적으로 수집된 신조어 감성단어의 개수를 보인 것이다.

Table 3. Final result of the collected words

Collection	1 st extraction	2 nd extraction
Basic Emotion Word	504	504
SNS  	32,031	39
Resource of NIKL	312	59
Wikipedia	354	45
Total	33,201	647

2.5 Quantification of Emotion Word

앞선 과정을 통해 최종적으로 수집된 감성단어 647개는 정량화 작업을 위해 제안한 감성모델에 배치하여 좌표 값을 구한다. 정량화 과정에는 1차로, 7명의 연구원이 감성단어 배치에 참여하였으며, 이 결과를 3번에 걸쳐 연구원 간 상호 대조하는 방식으로 조정한다. 이렇게 생성된 1차 결과는 34명의 일반 학생들의 집단지성을 이용하여 2차 정량화 조정 과정을 거쳐서 감성모델 평면의 최종적인 감성단어 위치를 결정한다. 감성단어의 배치 과정은 다음과 같다. 첫 번째, 표준어 감성단어를 먼저 배치하고 2차적으로 신조어 감성단어를 배치한다. 두 번째, 표준어 감성단어를 배치할 때는 단어의 감성범주가 1개인 것을 우선 배치하고 감성범주가 2개 이상인 것은 이후에 배치한다. 예를 들어, ‘골치 아프다’를 배치할 때, ‘골치가 아프다’라는 것은 보통 단순한 통증이 아니라 어떤 이유나 상황 때문에 골치가 아픈 것이다. 따라서 격양정도는 중성에 가깝지만 상당한 통증이 존재한다. 그러면 통증을 기준으로 봤을 때 단순한 통증이 아니라 통증보다 훨씬 더 네거티브에 가까운 통증이라 할 수 있으므로, 그 위치는 통증보다 네거티브에 인접한 지점(-0.43, 0.16)에 위치하는 것이 합당할 것이다. 이와 같은 단어배치 및 조정과정을 거쳐 최종적으로 배치가 완료된 감성모델의 상태는 Fig. 3과 같다.

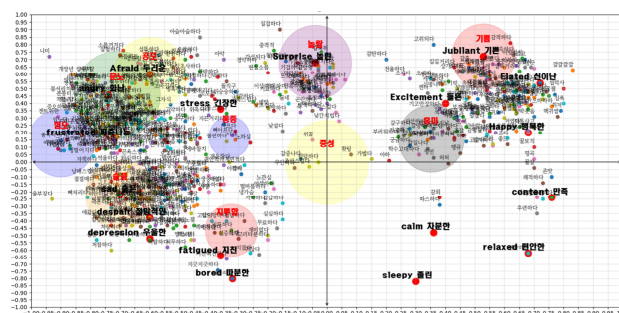


Fig. 3. Emotion words(647) Allocation on Proposed Emotion Model

3. Designing Emotion Analysis System

본 연구에서는, 특정 주제나 이슈에 대한 SNS 등의 분석대상 소셜 미디어의 댓글들을 수집하여 그 안에 표현된 감성의 흐름을 확장된 감성모델과 제안한 감성단어사전을 활용하여 효과적으로 분석할 수 있는 감성분석시스템을 설계한다.

3.1 System Diagram

사용자는 분석을 원하는 주제나 이슈의 사이트 혹은 SNS의 URL을 제공하거나 또는 이미 관련 댓글 등에 대해 수집한 데이터가 있다면 이를 본 감성분석시스템에 업로드하거나 분석을 원하는 목표 키워드 등을 감성분석시스템에 요청할 수 있다. 수집된 데이터를 업로드 한 경우, 감성분석시스템은 해당 데이터를 분석하고, 그 분석결과를 시각화하여 사용자에게 반환한다. URL이나 분석 대상 키워드를 요청한 경우, 감성분석시스템은 크롤러(crawler)를 통해 SNS에서 목표 키워드와 연관된 데이터를 수집하고, 이를 분석하여 시각화한 결과를 반환한다. Fig. 4는 전체 시스템의 동작 흐름을 나타낸 것이다.

3.2 Analysis Process

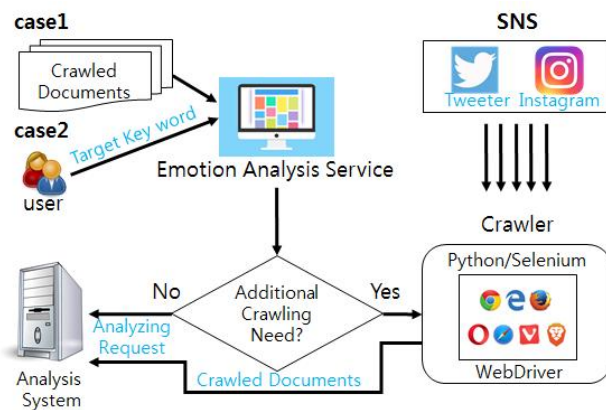


Fig. 4. System Diagram

감성분석시스템은 수집과정을 거쳐 시스템에 전달된 수집결과를 자연어처리 과정을 통해 분류하는데, 이 과정에서, 기본적인 형태소분석을 수행하게 된다. 먼저, 수집결과에서 감성을 표현하는 명사, 동사, 형용사, 감탄사를 추출한다. 다음으로, 추출한 단어의 원형을 찾아주는 표어추출(lemmatization)을 수행하며, 그 결과로 추출된 단어들의 기본형과 감성사전을 비교하여 감성단어만 선별한 후, 그 단어들의 빈도수를 측정한다. 최종적으로 선별된 감성단어들은 그 빈도수를 가중치로 감성모델에 배치되는

데, 빈도수가 높은 감성단어 중 동일하거나 유사한 감성범
주에 속한 단어들은 상대적으로 큰 원(circle)의 형태로 군
집화 되어 감성모델의 2차원 평면에 표현된다.



Fig. 5. Word Cloud of Initial System

일반적인 오피니언 마이닝 시스템들은 분석된 단어들을 Fig.5와 같이 단어 클라우드 형태로 시각화 하여 표현한다. 그러나 이러한 시각화 형태는 단순히 빈도수가 높은 단어들의 나열이므로 그 안에 담긴 의미와 감성은 최종 사용자가 개별적으로 판단하여야 하는 단점을 지닌다. 그러나 제안 시스템은 분석 대상으로부터 감성과 직접적인 연관이 있는 감성단어만을 추출하여 이를 분석한 후, 2차원 감성모델 평면에 단어들의 군집과 분포를 표현함으로써 단순한 단어의 나열이나 긍/부정 등의 결과가 아닌 직관적이면서 댓글에 담긴 다양한 감성적 의미를 표현 가능하다. Fig.5에 사용된 동일한 데이터에 대한 제안 시스템의 분석 결과는 Fig.6과 같다. 2차원 감성모델 평면상에 단어들은 그 빈도수에 따라 폰트의 크기와 군집 형태가 다르게 표현되며, 표현된 감성단어의 분포를 통해, 슬프고 안타까운 감성들이 주를 이루며 나머지 감성들이 2사분면의 좌측부분과 3사분면과의 경계면에 집중적으로 나타나므로 슬픔 감성이 강하고 약간의 침통함은 있으나 강한 우울감은 수반되지 않음을 직관적으로 알 수 있다.

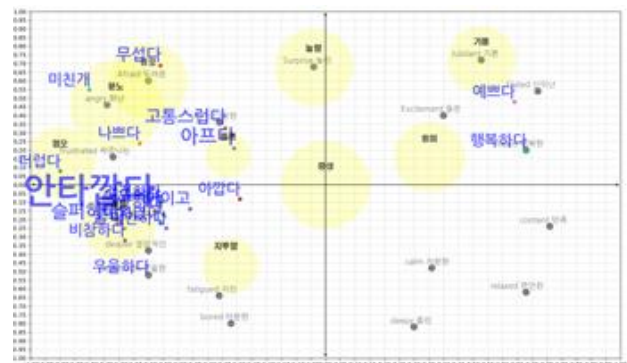


Fig. 6. Word Mapping on Emotion Model of Initial System

마지막으로, 본 감성분석시스템은 시각화 모듈을 통해 3가지의 결과를 생성하는데, 첫 번째, 감성모델 기반의 2차원 평면을 생성하고, 두 번째, 긍정/부정의 비율을 나타내는 원 그래프(Pie Chart)를 생성하며, 세 번째, 각 감성범주의 감성 점유율을 나타내는 방사형 그래프(Radar Chart)를 생성한다. Fig.7은 감성분석의 전체적인 과정을 보인 것이다.

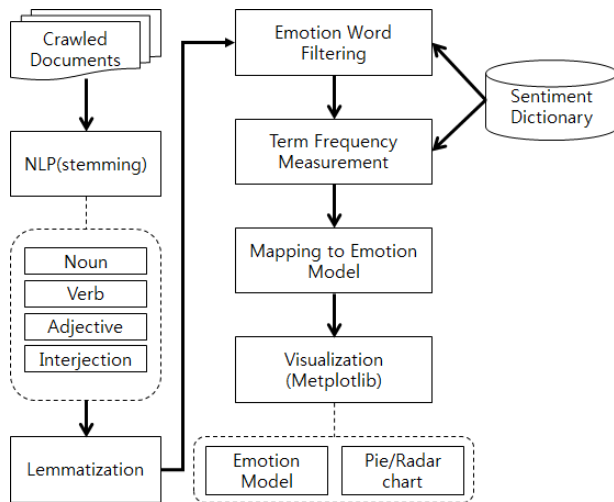


Fig. 7. Process Flow of Sentiment Analysis System

IV. Experiment Result

제안한 감성분석시스템의 특징과 성능을 평가하기 위해 다음과 같이 데이터를 수집하여 이를 분석하고 그 결과의 유효성과 정확성을 평가하였다. 평가 대상으로는 소셜 미디어에서 크게 이슈가 되었던 기사 세 가지를 선정하여 해당 기사와 관련된 소셜 미디어의 댓글을 수집하여 분석하였다. 기사는 부정적인 기사, 중립적인 기사, 긍정적인 기사를 각 1개씩 선정하였으며, 그 내용은 Table. 4와 같다. 최종적인 분석결과는 Matplotlib[12]을 이용해 시각화하였다.

첫 번째 뉴스는 “택진이형 밤새웠어요? 우린 밤새웠어요” (<http://naver.me/GqOKtoAl>) 라는 제목의 2019년 11월 27일자의 기사로, 댓글 544개가 달렸다. 기사의 내용은 엔씨소프트의 기대작 ‘리니지2M’가 높은 매출을 기록하며 국내 게임 시장의 위기를 극복하는 기폭제 역할을 할 것 이라는 내용으로 네이버 뉴스에서 IT 섹션으로 분류되었다. 해당 기사에 달린 댓글들의 내용은 대체적으로 부정적이며, 대체로 게임과 개발자를 비난하는 내용들로 기사를 읽은 사람들의 감정 표현도 ‘화나요’가 압도적으로 많은 기사이다. 두 번째 뉴스는 안 먹히려 맹독성 독사 행세한 두꺼비 위장술 첫 확인 (<http://bitly.kr/ZXSNNRpgV>) 이라는 제목의 2019년 10월

Table 4. News for Analysis

No.	Title	comments	Original article
1	택진이형 밤새웠어요? 우린 밤새웠어요	554	
	안 먹히려 맹독성 독사 행세한 두꺼비 위장술 첫 확인	313	
3	강다니엘, '두근:대결구도' 이상형월드컵서 10번째 우승... 왕좌 등극	7,578	
total		8445	

22일의 기사로, 댓글 313개가 달렸다. 기사의 내용은 생태계에서 살아남기 위하여 맹독성 독사로 위장한 두꺼비의 위장술이 처음으로 확인되었다는 내용의 정보성 기사로, ‘IT와 세계’ 섹션으로 분류된 기사이다. 이 기사의 댓글은 대체로 흥미롭다는 내용으로 이루어져있고, 기사를 읽은 사람들의 감정표현도 ‘좋아요’가 많은 긍정적 중립에 성향을 띤다. 세 번째 뉴스는 강다니엘, ‘두근:대결구도’ 이상형월드컵서 10번째 우승... 왕좌 등극 (<http://bitly.kr/ERtmpnTZ>) 이라는 제목의 2019년 4월 2일자의 기사로, 댓글 7,578개가 달렸다. 기사의 내용은 VS놀이 사이트인 ‘두근:대결구도’에서 실시하는 이벤트에 아이돌 강다니엘이 우승을 차지했다는 내용으로, 해당 기사에 달린 댓글은 대체로 긍정적이며, 기사를 읽은 사람들의 감정 표현도 ‘좋아요’가 압도적으로 많은 기사이다.

1. Negative Articles

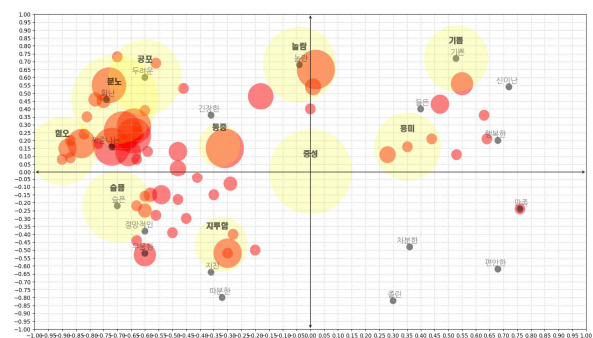


Fig. 8. Emotional analysis of article 1

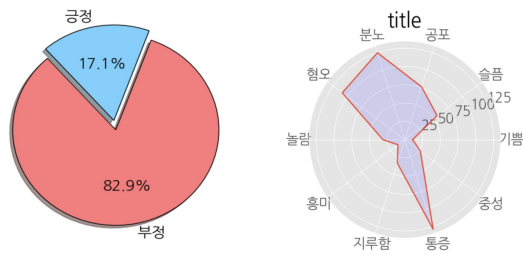


Fig. 9. Polarity & Main Sentiment Words of article 1

군집의 영역이 2사분면에 밀집해 있으므로 대체로 게임과 개발자를 비난하는 내용이 많음을 알 수 있다. 이때 2사분면의 분포가 공포보다는 혐오와 분노에 가까운 것을 볼 수 있다. 또한 1사분면과 2사분면 사이의 놀람 부분에 밀도가 높은 군집이 형성되어 있는 것으로 볼 수 있다. 이를 통해 해당 게임에 대한 놀라움과 부정적인 입장이 동시에 표현되고 있음을 알 수 있다.

본 기사에 대한 분석 수행 시, 신조어를 추가하지 않은 감성사전과 신조어 추가 후의 감성사전 모두를 사용하여 그 차이점을 비교하였다(두 사전 모두 본 연구에서 구축한 사전이며 신조어의 유무 차이만 존재함). 신조어를 배제한 감성단어 사전을 사용했을 때 필터링 된 감성 단어의 개수는 123개, 신조어가 포함된 감성단어 사전을 사용했을 때 필터링 된 감성 단어의 개수는 131개로 신조어가 포함된 감성단어 사전을 사용했을 때의 분석 결과가 그렇지 않은 감성단어 사전을 사용했을 때 보다 6.5% 더 많은 감성 단어 매칭 결과를 얻었다.

2. Neutral Articles

감성 분석 결과, 긍정 35.6% 부정 64.4%로 부정의 비율이 다소 높게 나타나고 있으나, 댓글에 부정적인 단어와 긍정적인 단어들이 적절하게 섞여서 대체적으로 흥미를 나타내는 내용으로 구성되어있다. 또한 2차원 감성 모델 상의 감성 단어들이 낮은 긴장도를 나타내는 4사분면을 제외하고 전체적으로 모든 사분면에 골고루 나타나고 있다는 것을 확인할 수 있다.

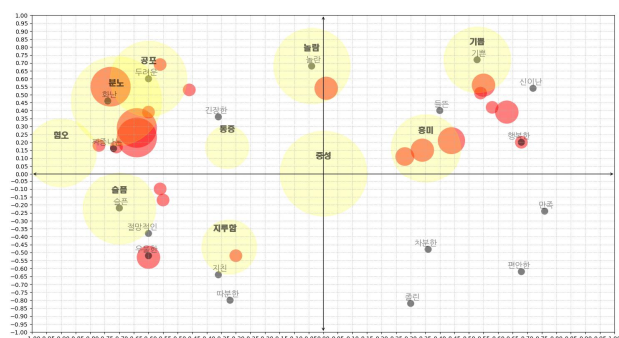


Fig. 10. Emotional Analysis of article 2

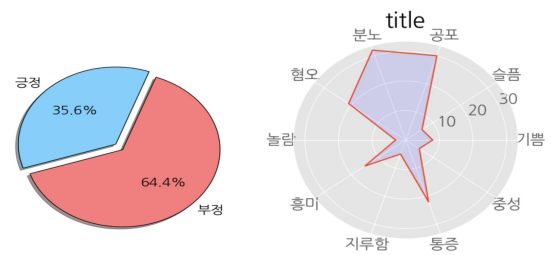


Fig. 11. Polarity & Main Sentiment Words of article 2

부정적인 부분의 분노와 공포가 다소 많이 나타난 이유로는 내용과 무관한 정치적 댓글과 기사의 사진이 징그럽게 느껴진다는 댓글이 많아 부정의 비율이 높았으며 이외에는 주로 흥미나 자연에 대한 놀라움을 표현한 댓글이다. 기사 2는 정보성 기사이며, 단어 군집들의 분포가 넓게 나타나지 않는 것으로 보아 댓글들에 표현된 감정이 어느 정도 공통적인 형태로 표현되고 있음을 알 수 있다.

3. Positive Articles

감성 분석 결과, 댓글들에 쓰인 단어들이 긍정 64.8%, 부정 35.2%로 긍정적인 단어들이 많이 사용된 것을 볼 수 있으며, 2차원 감성모델 상에서 단어들의 분포가 전체적으로 모든 사분면에 나타나지만, 긍정적이고 감성의 세기가 강한 1사분면에 주로 큰 군집이 나타남을 알 수 있다.

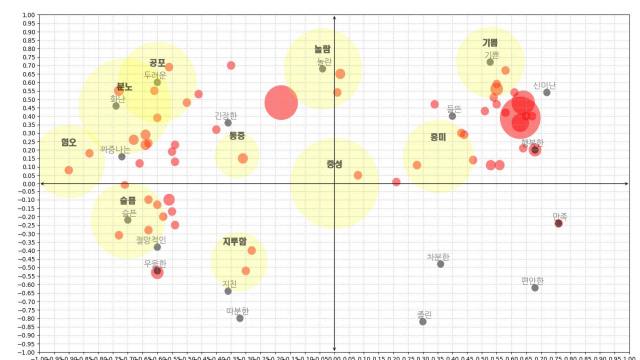


Fig. 12. Emotional Analysis of article 3

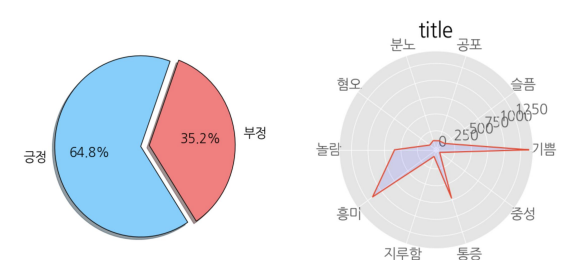


Fig. 13. Polarity & Main Sentiment Words of article 3

군집의 영역이 대부분 1사분면에 존재하며 특히 기쁜 감정에 밀집되어 있다는 것을 확인할 수 있으며, 이는 많은 팬들을 보유한 가수가 이벤트에서 우승을 차지하자 이에 환호하는 댓글들이 전체 댓글에 상당부분을 차지한 것이 해당 기사의 감성적 특징에 반영된 결과라 할 수 있다.

V. Conclusions

본 논문에서는 심리학 분야의 감성모델을 텍스트 마이닝의 감성분석에 적합하도록 변형하여 직관적으로 이해하기 쉬운 감성모델을 정의하고, 소셜 미디어의 감성분석에 필수적인 504개의 기본 감성단어와 다양한 소셜 미디어로부터 수집, 추출한 143개의 신조어를 등록한 감성단어사전을 구축한 후, 이를 적용한 감성분석시스템을 설계하였다. 설계한 시스템의 분석 성능 검증을 위해 3개의 뉴스 기사를 분석하여 그 결과를 제시하였다. 제안 시스템은 단순히 긍정/부정 여부를 넘어서 감성모델의 2차원공간에서 군집을 이루는 감성단어의 분포를 통해 더 상세하게 기사에 포함된 감성을 분석 가능하였다. 따라서 본 시스템은 선거결과 예측이나 최근 소비자의 트렌드 파악과 같은 마케팅 전략 등의 다양한 분야에 폭넓은 적용이 가능할 것으로 기대된다. 향후 연구로, 감성모델에 감성단어를 배치함에 있어 집단지성의 활용성을 높여 각 단어들의 피드백을 통해 세밀한 위치 조정을 수행할 예정이다. 또한 “미친 너무 행복해”와 같이 긍정을 부정형 단어로 강조하여 표현하는 문장의 경우, 문장에 긍정과 부정의 단어가 동시에 출현할 수 있는데, 이러한 문장의 문맥상의 실제 의미를 정확하게 분석 가능한 의미 분석, 구문분석과 같은 자연어처리에 대한 연구도 수행하여 그 결과를 분석시스템에 반영하고자 한다. 또한 다양한 시각화 기법도 추가하고자 한다.

REFERENCES

- [1] Liu, B., "Sentiment Analysis and Opinion Mining", Morgan and Claypool Publishers, 2012. <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>
- [2] "2017 Survey on the Internet Usage", KISA, <https://isis.kisa.or.kr/board/index.jsp?pageId=060100&bbsId=7&itemId=821&pageIndex=1>
- [3] Sunkyoung 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
- [4] Byungun Yoon, "Opinion Mining with Artificial Intelligence-What is Social Opinion Mining?", Samsung SDS insight reports, 2017, https://www.samsungds.com/global/ko/support/insights/1195888_2284.html
- [5] Hannah Kim, Young-Seob Jeong, "Social Issue Analysis Based on Sentiment of Twitter Users", Journal of Convergence for Information Technology, Vol. 9, No. 11, pp. 81-91, 2019
- [6] Russell, J., "A circumplex model of affect". Journal of Personality and Social Psychology Vol. 39, No. 6, pp 1161-1178, 1980
- [7] Thayer, R. E. "The Bio psychology of Mood and Arousal", New York: Oxford University Press, 1989.
- [8] Plutchik, R., "The Nature of Emotions", American Scientist, 2011.
- [9] Soojin Lee, Sunkyoung Kim, Panseop Shin, "Korean-based Emotion Model for Music Content Retrieval", Proceedings of Conference of KOEN 2016, The Korean Entertainment Industry Association(KOEN), pp. 71~74, 2016. 5.
- [10] 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.
- [11] Konlpy, <http://konlpy.org>
- [12] Matplotlib, <https://matplotlib.org>

Authors



Pan-Seop Shin received the B.S., M.S. and Ph.D. degrees in Computer Engineering from Hongik University, Korea, in 1992, 1994 and 2000, respectively. Dr. Shin joined the faculty of the Department of Computer Engineering

at Daejin University, Pocheon-Si, Gyeonggi-Do, Korea, in 2002. He is currently a Professor in the Department of Computer Engineering, Daejin University. He is interested in distributed databases system, data mining and data visualization technology.