감성사전 정보가 결합된 특징통합벡터를 활용한 감성분석

Sentiment Analysis Using Mixed Feature Vector combined with the Sentiment Dictionary Information

저자
(Authors)

김호승, 이지형
Ho-Seung Kim, Jee-Hyong Lee

출처
(Source)

한국지능시스템학회 논문지 30(6), 2020.12, 494-499 (6 pages)
Journal of Korean Institute of Intelligent Systems 30(6), 2020.12, 494-499 (6 pages)

발행처
(Publisher)

한국지능시스템학회
Korean Institute of Intelligent Systems

URL

http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE10507420

APA Style

김호승, 이지형 (2020). 감성사전 정보가 결합된 특징통합벡터를 활용한 감성분석. 한국지능시스템학회 논문지, 30(6), 494-499.

Journal of Korean Institute of Intelligent Systems
Vol. 30, No. 6, December 2020, pp. 494-499
http://dx.doi.org/10.5391/JKIIS.2020.30.6.494

ISSN(Print) 1976-9172
ISSN(Online) 2288-2324

JKIIS

감성사전 정보가 결합된 특징통합벡터를 활용한 감성분석

Sentiment Analysis Using Mixed Feature Vector combined with the Sentiment Dictionary Information

김호승*, 이지형**,†
Ho-Seung Kim and Jee-Hyong Lee†

*성균관대학교 대학원 인공지능학과 석사과정, **성균관대학교 정보통신과학부 교수
*ME Course, Department of Artificial Intelligence, Graduate School, Sungkyunkwan University,
**Professor, School of Information and Communication, Sungkyunkwan University

요 약

기존의 감성분석은 사전 학습된 정보를 이용하는 것보다 단어, 문장 또는 문맥을 인공신경망 모델에서 학습하고 이를 이용하여 감성분석을 시도하고 있다. 본 논문에서는 사전 학습된 정보와 인공신경망을 같이 사용하기 위한 방법으로 감성사전을 선택하였다. 먼저 사전 구축되어 있는 감성사전이 갖고 있는 단어들의 기본 감성극성과 감성사전으로 학습된 모델을 통한 문장의 일반적인 감성극성, 인공신경망을 통한 문맥의 감성극성을 추출한다. 이렇게 얻어진 단어, 문장 감성극성과 문맥 감성극성을 결합하여 특징통합벡터를 만드는 것을 제안한다. 그리고 이를 감성분석 모델에 적용하여 문장이 가지고 있는 감성극성을 분류하는 실험을 하였고 우수한 성능을 나타내는 것을 확인하였다.

키워드 : 감성사전, 감성분석, 극성추출

Received : Jul. 31, 2020
Revised : Aug. 21, 2020
Accepted : Sep. 15, 2020
† Corresponding author
(john@skku.edu)

Abstract

Previous sentiment analysis attempts to analyze sentiment by learning words, sentences or contexts from models using artificial neural networks rather than using pre-trained information. In this paper, sentiment dictionary is selected as a method to use pre-trained information and neural networks together. It extracts both the word, sentence sentiment polarity based on the sentiment dictionary and context sentiment polarity through the neural networks. We suggest to create a mixed feature vector by combining word, sentence and context sentiment polarity information. In addition, it is confirmed that it shows excellent performance through an experiment to classify the sentiment polarity of sentences by using it in the sentiment analysis model.

Key Words : Sentiment Dictionary, Sentiment Analysis, Polarity Extraction

1. 서 론

자연어 처리(Natural Language Processing)에서 가장 중요한 문제이면서 우리가 항상 도전하고 있는 과제는 언어의 이해(Natural Language Understanding)라고 할 수 있다. 그 중에서도 극성(Polarity)을 분석하는 감성분석(Sentiment Analysis)은 텍스트에 담겨 있는 주관적인 정보를 분석하여 언어의 이해를 돕는 자연어처리 기법 중 하나이다. 감성분석은 이전부터 소셜 네트워크 서비스(Social Network Service), 인터넷 쇼핑몰, 뉴스기사, 영화 감상평 등 인터넷 환경에서 제공하는 다양한 서비스에 대한 피드백 텍스트를 분석하는 것부터 사람의 음성에 담겨 있는 감정을 분석하기까지 다양하게 연구되어 이용되고 있으면서 실생활과 매우 밀접한 관련이 있어 많은 관심을 받고 있다.

본 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단-차세대 정보컴퓨팅기술개발사의 지원을 받아 수행된 연구임(No. NRF-2017M3C4A7069440). 또한 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2019-0-00421, 인공지능대학원지원(성균관대학교))

기존의 감성분석은 특정 도메인이나 맥락 내에서 단어, 문장 또는 문맥을 인공신경망을 통한 학습 또는 기계학습[1]을 이용하여 분석하려고 시도를 하고 있다. 다양한 시도 덕분에 우수한 성능을 나타내고는 있지만 이 과정에서 단어 자체가 가지고 있는 일반적인 정보를 이용하는 방법은 효율성이 떨어지고 별개의 것으로 판단하여 잘 이용되지 않는다. 단어 자체가 가지고 있는 일반적인 정보에 집중하는 연구도 진행되고 있으나 도메인 또는 문맥의 정보를 많이 반영하지 않고 학습이 진행되기 때문에 같은 단어라도 여러 뜻을 가지고, 여러 형태로 변형된다는 점에서 한계점이 있어서 연구하려는 노력은 미비한 상태이다.

감성사전은 단어 자체가 가지고 있는 정보를 종합한 결과물이라고 할 수 있다. 이는 감성 어휘에 대한 사전으로 다양한 언어모델에 있어서 감성분석 수행을 위한 기본 자료로 활용 된다. 감성사전에 포함되어 있는 다양한 감성 어휘는 특정 도메인에 따라서 감성의 종류나 정도가 달라지기 때문에 다양한 방법으로 연구가 되고 있는 분야이다. 특화된 감성사전을 제작하는 연구, 일반적인 감성사전을 제작하는 연구, 특정 도메인에 특화된 감성사전을 다른 도메인에 전이 학습시켜 적용하려는 연구 등 많은 연구들이 그 예이다. 영어로 구성된 데이터 분석을 하기 위해서는 SentiWordNet을 이용한 방법[2]이 많이 사용되는데, 본 논문에서는 단어의 감성극성과 문장이 가지고 있는 일반적인 감성극성을 추출하기 위하여 단어들이 갖고 있는 긍부정어에 대한 수치를 인간의 보편적인 기본 감정 표현을 바탕으로 점수를 부여한 KNU 한국어 감성사전[3]을 사용한다.

본 논문에서는 앞서 소개한 단순히 어휘기반의 접근방식 또는 문맥파악을 위한 기계학습 기반의 접근방식을 별개로 사용하는 것이 아니라, 2개의 접근방식을 결합한 혼합방식을 제안한다. 감성사전과 감성사전을 통한 극성 추출 모델을 통하여 단어의 감성극성과 문장이 가지고 있는 일반적인 감성극성을 추출하였으며, 인공신경망을 이용하여 특정 도메인의 문맥 내 감성극성 또한 추출하여 3가지 정보를 결합하였다. 3가지 정보를 모두 포함한 벡터가 이번 논문에서 제안한 특징통합벡터이며 이를 감성분석에 활용한 실험을 통하여 우수한 성능을 증명하였다.

2. 관련 연구

감성분석에 대한 관심이 높아지는 만큼, 다양한 방법으로 감성분석을 하기 위한 시도가 이루어지고 있다. 먼저 긍정(positive), 중립(neutral), 부정(negative)과 같은 극성을 분석하는 Fine-grained Sentiment Analysis[4] 방법이 있으며, 기쁨(hippiness), 분노(anger), 두려움(fear), 슬픔(sadness), 놀람(surprise)과 같은 감정을 분석하는 Emotion Detection[5], 어떤 단어나 문장이 감성분석에 있어서 가장 큰 영향을 미쳤는지 판단하는 Aspect-based Sentiment Analysis[6] 등이 있다.

본 논문에서는 Fine-grained Sentiment Analysis에 초점을 맞추어 3가지 접근방법에 관심을 가지고 연구를 진행하였다. 적용한 3가지 접근 방법으로는 감성사전이 갖고 있는 단어들의 기본 감성극성을 이용하는 방법, 감성사전으로 학습된 모델을 통한 임베딩을 이용하는 방법, 인공신경망 구조를 변형하는 방법이 있다.

감성사전을 이용하는 방법은 크게 2가지로 나누어지게 되는데 하나는 보편적인 기본 감정 표현을 바탕으로 감성사전을 구축하고, 구축되어 있는 사전을 적절한 데이터셋에 적용하는 방법이다. 이번 연구에 적용한 KNU 한국어 감성사전이 그 예가 되겠다. 다른 하나는 도메인에 특화된 감성 사전을 제작하기 위해, 도메인에서 사용된 문장이나 글을 이용하여 모델을 학습시킨다. 이렇게 만들어진 감성사전[7]은 해당 도메인에서 사용할 때는 높은 효율성을 보이나, 기타 도메인에서 사용할 경우 효율성이 매우 떨어져서 일반적으로 사용이 제한되는 방법이다.

임베딩이란 고차원 벡터의 변환을 통해 생성할 수 있는 저차원 벡터를 가리키는 것이다. 임베딩하는 방법은 여러 방법이 있으며, 자연어 처리에서는 주로 워드 임베딩을 많이 사용한다. 단어가 구성하고 있는 고차원의 벡터를 저차원으로 변형시켜주는 워드 임베딩은 아주 많은데 그 중 잘 알려진 기법으로는 NNLM[8], Word2Vec[9]이 있다. 임베딩 방식의 차이에 따라 성능의 차이도 나타나기 때문에 최근에 이런 임베딩 기법을 다양하게 적용함으로써 성능을 높이는 연구가 진행 중 이다.

마지막으로 인공신경망 구조를 변형하는 방법인데, 현재 국내에서도 데이터의 형태나 특징에 따라서 다양한 구조들이 연구되고 있다. Sequential 정보를 분석해야 하는 자연어처리에 자주 사용되는 구조로는 RNN (Recurrent Neural Networks)[10], 그리고 본 연구에서 이용한 LSTM (Long Short-Term Memory models), Bi-LSTM (Bi -Directional LSTM) 등이 있으며, 최근에는 CNN (Convolution Neural Net works)[11]을 활용한 자연어처리 연구도 시도되고 있다.

3. 제안 기법

3.1 모델 구조

본 연구에서는 감성사전을 이용하여 고차원의 정보를 가지고 있는 문장 벡터를 단어와 문장 감성극성을 포함하고 있는 저차원의 벡터들로 전환시켜주는 방법을 제안한다. 감성사전에 의해 전환된 단어극성벡터(Word Polarity Vector), 문장극성벡터(Sentence Polarity Vector)를 문맥의 감성극성 정보를 포함한 문맥극성벡터(Context Polarity Vector)와 결합하여 감성분석을 하는 것이 최종 모델의 메커니즘이다.

3.2.1 단어극성벡터 추출 방법

단어극성벡터는 감성사전을 이용하여 문장 내 단어가

기본적으로 가지고 있는 감성극성을 특징으로 추출하였다. 문장 내에 감성사전에 있는 용어와 일치하는 단어가 있는지 여부를 판단하고, 단어를 카운트하는 방식을 이용하였다. 단어가 있다면 감성사전에 있는 감성극성을 바탕으로 점수를 측정하여 벡터형태로 표현했다. 이때 중립을 뜻하는 단어는 긍정, 부정을 분류하는데 큰 영향이 없을 것이라 판단하였고 긍정 또는 부정으로 임의 조정을 하기가 제한되어, 감성사전 내 없는 단어와 마찬가지로 단어 점수를 0점으로 부과하였다. 다양한 극성을 가지고 있는 단어가 한 문장 내에 많이 있다면 표현된 단어 점수를 가감하여 점수를 부과하였다. 결과 점수를 '-2', '-1', '+1', '+2'를 기준으로 총 5개의 레이블로 구분하였다.

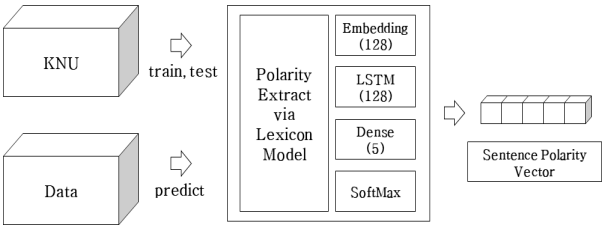

그림 1. 감성사전을 통한 극성 추출모델
Figure 1. Polarity Extraction Model via Sentiment Dictionary

3.2.2 문장극성벡터 추출 방법

문장극성벡터는 감성사전을 이용한 사전 훈련모델을 통하여 일반적으로 문장이 가지고 있는 감성극성을 추출한다. 감성사전을 통한 극성 추출모델(Polarity Extraction Model via Sentiment Dictionary) 구조는 그림 1과 같다. 감성사전에 있는 단어를 이용하여 말뭉치(Corpus)를 구성하였으며, 이를 학습 데이터와 테스트 데이터와 이용하여 극성 확률을 추측하는 모델이다. 인공신경망은 LSTM을 사용하였고, 데이터셋이 순차적으로 처리되는 텍스트이기 때문에 방향성을 가지고 분석하여 좋은 성능을 보였다. 레이블은 매우 부정, 부정, 중립, 긍정, 매우 긍정을 의미하는 총 5개의 벡터로 이루어져 있으며, 이 모델을 통하여 일반적으로 문장이 가지고 있는 극성을 추출하였다.

3.2.3 문맥극성벡터 추출 방법

문맥극성벡터는 영화라는 도메인의 특징을 살려 도메인에 특화된 문맥상의 감성극성을 추출한다. 문장극성벡터와 다르게 학습용 데이터에 있는 단어를 이용하여 말뭉치를 구성하였으며, 이를 학습 데이터와 테스트 데이터와 이용하여 문맥의 특징을 추출하는 모델(Context Feature Extract Model)을 학습하였다. 모델의 모습은 그림 2와 같으며 Bi-LSTM을 사용하여 문장의 길이가 늘어나서 생기는 정보 손실을 최소화하고자 하였다. 레이블은 모두 부정(0), 긍정(1) 총 2개로 구성되어 있으며, 이 모델을 통하여 문맥이 가지고 있는 문맥극성벡터를 추출하였다.

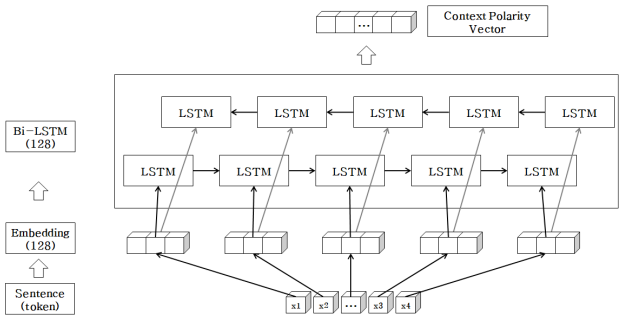

그림 2. 문맥 특징 추출 모델
Figure 2. Context Feature Extract Model

3.3 감성분석 모델

이렇게 단어, 문장, 문맥 3가지 측면의 감성극성 특징을 나타내는 벡터들을 만들고 케라스의 concatenate layer를 추가하여 모든 벡터가 결합(Concatenate)된 특징통합벡터(Mixed Feature Vector)를 완성하였다. 최종적으로 완성된 특징통합벡터를 그림 3과 같이 입력으로 사용하여 Dense NN 인공신경망으로 분류하는 감성분석 모델(Sentiment Analysis Model)을 제안한다.

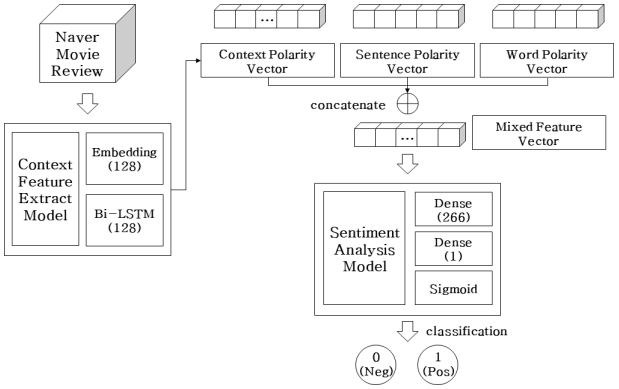

그림 3. 감성분석 모델
Figure 3. Sentiment Analysis Model

4. 실험 및 결과

4.1 실험 데이터

실험 데이터 셋[12]은 네이버 영화평과 영화 리뷰를 긍정과 부정으로 구분하는 레이블을 지닌 데이터 셋(약 20,000개)을 이용하였다. 데이터를 임의로 나누어서, 학습용 데이터 셋 약 15,000개(학습용 13,000개, 검증용 2,000개), 테스트용 데이터 셋 약 5,000개로 구분하였고 연구를 진행하였다.

KNU 한국어 감성사전에는 단어가 총 약 15,000개 포함되어 있다. 단어는 매우긍정, 긍정, 중립, 부정, 매우부정으로 구분되는 총 5개의 레이블은 지니고 있으며, 학습 데이터로 14,000개를 사용하였고 테스트용 데이터로 1,000개를 사용하였다.

4.2 단어 감성극성 추출 실험 및 결과

단어 감성극성을 추출하기 위하여 본 연구에서는 감성사전을 이용하였다. 전처리 과정에서 감성사전 내 단어는 별도의 전처리를 실시하지 않은 상태에서 단어의 모양을 유지하고 진행하였다. 학습용 데이터와 테스트용 데이터는 문장 내에 한글 외 특수문자, 숫자는 전부 제거하고, KoNLPy 패키지의 Okt를 이용한 형태소별 토큰화와 중복되는 내용을 삭제하는 전처리 과정을 진행하였다. 전처리 후 데이터의 문장 내에 감성사전에 포함되어 있는 단어가 얼마나 있는지 카운트하였고, 포함되어 있는 단어마다 매우 부정으로 분류되어 있는 단어는 '-2', 부정'-1', 중립 '0', 긍정 '+1', 매우 긍정 '+2'로 점수를 부가하였다. 그렇게 부가된 점수를 합산하여 문장에 대한 점수를 측정하였고 점수가 -2미만은 매우 부정, -2이상 -1이하는 부정, -1초과 +1미만은 중립, +1이상 +2이하는 긍정, +2초과는 매우 긍정으로 분류하였다.

실험결과는 아래의 표 1과 같다. 중립으로 분류된 데이터들은 정확도 측정에서 제외하였으며, 그 결과 학습용 데이터 약 12%의 정보와 테스트 데이터 36%의 정보는 제외되었다. 제외한 뒤 목표 데이터의 긍정, 부정과 구분하여 정확도를 측정하였다. 정확도는 약 73%정도로 이 데이터만 가지고 극성을 확실히 판단할 수는 없다. 하지만 충분히 단어로부터 감성극성을 추출하였으며 기타 문장, 문맥의 감성극성과 결합하였을 때 활용 가능성은 충분히 있음을 알 수 있다.

표 1. 단어 감성극성 추출 정확도
Table 1. Accuracy of word sentiment polarity extraction

	training data	test data
number of data	145,791	48,996
number of neutral	17,681(12%)	17,761(36%)
accuracy	0.7304	0.7254

4.3 문장 감성극성 추출 실험 및 결과

문장 감성극성을 추출하는 단계에서는 인공신경망을 활용하였다. 이 단계에서는 단어 형태를 최대한 맞추기 위하여, 감성사전 단어와 영화 리뷰 데이터를 모두 형태소별 토큰화를 실시하였으며 감성사전 단어를 기준으로 말뭉치를 제작하였다. 레이블을 데이터 전처리 후 학습과 테스트에 입력 데이터로는 감성사전의 단어를 이용하고 타겟 데이터로는 감성사전 내 단어가 가지고 있는 지정된 레이블 극성을 사용하였다. 레이블은 매우 부정(-2), 부정(-1), 중립(0), 긍정(1), 매우 긍정(2)으로 나누어져 있으며, 5개 벡터로 표현하여 이를 구분하기 위해 훈련하였다. 모델과 임베딩 사이즈를 변경해가면서 반복 실험을 하였고 그 결과는 표 2와 같다.

다양한 모델과 임베딩 사이즈의 실험을 통해 최적의 모델과 임베딩 사이즈를 선택하였고, 기타 파라미터를 조정해나가면서, 최적화를 진행하였다. 그 결과 정확도 약 80%의 결과를 얻을 수 있었으며, 이 모델의 가중치를 사용하여 영화 리뷰 데이터에 적용하였다. 그 결과 학습 데이터 문장이 가지고 있는 일반적인 감성극성을 추출하였다.

표 2. 문장 감성극성 추출 정확도
Table 2. Accuracy of sentence sentiment polarity extraction

model	Embedding size	accuracy
LSTM 128	64	0.7610
	128	0.7958
LSTM 256	128	0.7844
	256	0.7730
Bi-LSTM 128	64	0.7867
	128	**0.8054**
Bi-LSTM 256	128	0.7957
	256	0.7802

4.4 문맥 감성극성 추출 실험 및 결과

문맥 감성극성을 추출하는 단계에서는 기존에도 자주 사용되는 인공신경망을 활용하여 감성극성을 추출하였다. 이 때는 전체 데이터에 대하여 토큰화를 진행하고 더불어서 필요 없다고 판단되는 조사 위주의 불용어를 선정하여 제거하였다. 말뭉치는 학습용 데이터의 단어로 구성하였으며, 타겟 데이터는 긍정, 부정으로 구분되어 있는 결과값을 이용하였다. 데이터 준비 이후에 문장 감성극성 추출단계와 마찬가지로 모델과 임베딩 사이즈, 패딩 사이즈, 학습용 데이터 크기를 변경해가면서 반복 실험을 진행 하였고, 표3에서 그 결과를 확인 할 수 있다.

표 3. 문맥 감성극성 추출 정확도
Table 3. Accuracy of context sentiment polarity extraction

parameter	value	accuracy	
		LSTM	Bi-LSTM
training data size	**100%**	0.8429	**0.8454**
	50%	0.8195	0.8328
	10%	0.7900	0.7906
padding size	30	0.8419	0.8439
	20	0.8429	**0.8454**
	10	0.8363	0.8329
number of layer	64	0.8482	0.8433
	128	0.8429	**0.8454**
	256	0.8403	0.8475
embedding size	64	0.8310	0.8355
	128	0.8429	**0.8454**
	256	0.8331	0.8316

학습용 데이터에 대하여 임의로 데이터를 선택하여 크기를 줄여나가며 실험하였다. 실험결과 크기를 줄여가며 비교해 보았을 때, 데이터양이 줄어들수록 정확도가 감소하긴 하나, 데이터양이 10%밖에 되지 않음에도 불구하고 성능을 79%정도로 유지하고 있음을 알 수 있다. 이는 그만큼 모델의 성능이 충분함을 보여주는 결과이다.

그 외 모델과 패딩 사이즈, 임베딩 사이즈를 변경하면서 최적의 모델을 찾고자 노력하였고, 연구 결과에는 포함하지 않은 학습율, 배치 사이즈와 같은 기타 파라미터들도 조정하여 성능향상에 많은 노력을 기울였다. 그 결과 문맥 감성극성을 통한 분류결과 정확도 약 84%의 결과를 얻을 수 있었다. 다른 단어 감성극성, 문장 감성극성의 정확도와 비교하였을 때 성능이 우수함이 확연하게 구분됨을 보여줌으로써 감성분석에 있어서 문맥정보의 중요도를 증명하였다.

4.5 특징통합벡터 감성분석 및 결과

단어, 문장, 문맥의 감성극성을 포함한 벡터를 비교실험을 포함하여 3가지 방법으로 결합하여 감성분석을 실시하였다. 표 4는 감성분석 결과이다. 입력 데이터는 추출된 3가지 감성극성을 포함한 벡터이고, 타겟 데이터는 영화 리뷰 데이터의 2개의 레이블(긍정, 부정)로 구분되어 있는 결과값을 이용하였다.

표 4. 감성분석 정확도
Table 4. Accuracy of sentiment analysis

model	accuracy
word polarity vector + sentence polarity vector + context polarity vector	**0.8594**
sentence polarity vector + context polarity vector	0.8578
word polarity vector + context polarity vector	0.8547
context polarity vector	0.8454

본 논문에서 제안한 3가지 벡터를 모두 결합한 특징통합벡터를 사용한 감성분석 모델은 단순하게 문맥 감성극성 추출 결과를 이용한 결과보다 약 1.5%의 성능향상을 보이며 우수한 성능을 보였다.

비교실험으로 진행한 문장 감성극성만을 결합한 경우와 단어 감성극성만을 결합한 경우를 나누어서 실험하였다. 비교해본 결과 2가지 경우의 분류 정확도는 약 0.2%의 차이밖에 보이지 않았으나, 전체적으로 문장 감성극성을 결합한 경우에 조금 더 뛰어난 성능을 보였다. 또한 문맥 감성극성을 이용하여 감성분석한 결과보다는 약 1%정도의 성능향상을 보이면서 2가지 방법 모두 감성분석에 있어서 유의미함을 보였다.

5. 결론 및 향후 연구

본 논문에서는 영화 리뷰 데이터에 대하여 감성사전을 이용하여서 해당 문장에 포함되어 있는 단어의 감성극성, 일반적으로 판단 가능한 문장의 감성극성을 추출하고, 인공신경망을 이용하여서 데이터가 속해있는 특정 도메인에서 확인 가능한 문맥 감성극성 3가지를 추출하였다. 그리고 이 감성극성들을 결합하여 혼합특징을 나타내는 특징통합벡터를 완성하였고, 이를 이용한 감성분석 모델을 제안하였다.

감성사전을 이용하여 추출된 단어와 문장의 감성극성의 정확도는 약 73%, 80% 내외로 충분히 감성분석에 도움을 줄 수 있음을 보여주었다. 이 2가지 정보를 결합하여서 기존의 문맥정보만 가지고 있을 때와 비교하여 성능향상을 약 1.5%정도 이루면서 우수한 성능을 보였다.

비교실험으로 단어극성만 결합하였을 때와 문장극성만 결합하였을 때를 비교함으로써 각 정보가 성능향상에 기여함을 보였다. 또한 단어극성에 비하여 문장극성이 조금 더 성능향상에 더 많은 기여를 함을 보여주었고, 단순히 극성을 가진 단어만을 가지고 문장의 극성을 판단하기 보다는 문장, 문맥을 가지고 극성을 판단하는 것이 더 정확함을 보였다.

앞으로의 연구에서는 이번 연구에서 제외한 특수문자, 영어를 포함한 다양한 실험 데이터를 대상으로 영역이 확장되어야겠다. 그러기 위해서 다양한 도메인에 특화되어 있는 감성사전 또는 모든 도메인을 아우르는 감성사전의 개발과, 분석을 위한 전처리 방법과 모델도 개선하는 방향으로 본 연구를 이어가고자 한다.

Conflict of Interest

저자는 본 논문에 관련된 어떠한 잠재적인 이해상충도 없음을 선언한다.

References

[1] Hong Sola, Yeounoh Chung, Jee-Hyong Lee, "Semi-supervised learning for sentiment analysis in mass social media," *Journal of Korean Institute of Intelligent Systems*, vol. 24, no. 5, pp. 482-488, 2014. https://doi.org/10.5391/JKIIS.2014.24.5.482

[2] In-Su Kang, "A Comparative Study on Using Senti-WordNet for English Twitter Sentiment Anaylsis," *Journal of Korean Institute of Intelligent Systems*, vol. 23, no. 4, pp. 317-324, 2013. https://doi.org/10.5391/JKIIS.2013.23.4.317

[3] Sung-min Park, Chul-won Na, Min-seong Choi, Da-hee Lee, Byung-won On, "KNU Korean Sentiment Lexicon : Bi-LSTM-based Method for Building a Korean Sentiment Lexicon," *Journal of Intelligence and Information Systems*, vol. 24, no. 4, pp. 219-240, 2018.

https://doi.org/10.13088/jiis.2018.24.4.219
[4] E. Guzman and W. Maalej, "How Do Users Like This Feature? A Fine Grained Sentiment Analysis of App Reviews," *2014 IEEE 22nd International Requirements Engineering Conference (RE), Karlskrona*, pp. 153-162, 2014. doi : 10.1109/RE.2014.6912257
[5] Muhammad Abdul-Mageed, Lyle Ungar, "EmoNet : Fine-Grained Emotion Detection with Gated Recurrent Neural Network," *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, vol. 1, pp. 718-728, July, 2017.
[6] Yukun Ma, Haiyun Peng, Erik Cambria, "Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM," 32nd AAAI Conference, 2018.
[7] Sang-hoon Lee, Jing Cui, Jong-woo Kim, "Sentiment analysis on movie review through building modified sentiment dictionary by movie genre," *Journal of Intelligence and Information Systems*, vol. 22, no. 2, pp. 97-113, 2016.
[8] Y. Bengio, "Deep Learning of Representations : Looking Forward," arXiv preprint, 2013.
[9] Tomas Mikolov, "Efficient Estimation of Word Representations in Vector Space," arXiv preprint, 2013.
[10] Jin-kwang Som, "A Study on Content Recommendation System through Sentiment Analysis using RNN LSTM and ACO," *The Korean Institute of Information Scientists and Engineers*, 2017.
[11] Min Kim, Jeunghyun Byun, Chunghee Lee, Yeonsoo Lee, "Multi-channel CNN for Korean Sentiment Analysis," *Annual Conference on Human and Language Technology*, pp. 79-83, 2018.
[12] L. Park. "Naver sentiment movie corpus v1 [Online]." Available : https://github.com/e9t/nsmc, 2011, [Accessed: June 10, 2020]

저 자 소 개

김호승 (Ho-Seung Kim)
2013년 : 육군사관학교 물리학과 학사
2020년~현재 : 성균관대학교 대학원 인공지능학과 석사과정

관심분야 : deep learning, natural language processing
ORCID Number : 0000-0002-1416-5004
E-mail : tree901024@g.skku.edu

이지형 (Jee-Hyong Lee)
1993년 : 한국과학기술원 전산학과 학사
1995년 : 한국과학기술원 전산학과 석사
1999년 : 한국과학기술원 전산학과 박사
2002년~현재 : 성균관대학교 컴퓨터공학과 교수

관심분야 : machine learning, deep learning, intelligence system
ORCID Number : 0000-0001-7242-7677
E-mail : john@skku.edu