

딥러닝과 감성사전을 결합한 하이브리드 감성분석 시스템 개발

김백기¹, 장경배^{2*}

¹고려사이버대학교 융합정보대학원 석사과정, ²고려사이버대학교 기계제어공학과 교수

Development of Sentiment Detection combined with Deep Learning and Sentiment Dictionary

Bae-Ki Kim¹, Kyung-Bae Jang^{2*}

¹M.S. Course, The Graduate School of Interdisciplinary Information Studies, The Cyber University of Korea

²Professor, Dept. of Mechanical and Control Engineering, The Cyber University of Korea

요약 스마트 디바이스와 소셜미디어의 보급 확산, 온라인에서의 제품 구매 증 로 인하여 많은 기업들이 소비자의 소비패턴, 생각 등을 이해하려고 한다. 이에 따라 온라인에서 제품이나 서비스에 대한 소비자들의 의견이 포함된 리뷰를 수집하여 소비자들의 감성을 이해하는 필요성이 증대되고 있으며 국내·외 기업이나 연구기관에서 관련 연구가 진행되고 있다. 그러나 아직 영어로 표현된 데이터를 대상으로 한 연구가 대부분이며, 영어 텍스트에 대한 어휘사전이나 머신러닝 접근법으로 감성분석(Sentiment Analysis)에 대한 많은 연구와 성과가 발표되고 있다. 그에 반해 국어는 한국어가 갖고 있는 복잡성과 딥러닝을 위한 레이블링 데이터가 부족하기 때문에 러신머닝 접근법에 의한 감성분석 정확률이 상대적으로 낮다. 이러한 문제점을 개선하고자 본 연구에서는 한글 온라인 리뷰를 대상으로 딥러닝과 감성사전 기법의 장점을 활용하여 감성분석의 정확도를 향상시키는 하이브리드 접근법의 시스템을 활용했다. 이를 통해 정확도, 정밀도, 재현율 등의 지표들이 얼마나 개선되는지 확인했다. 본 연구 결과는 향후 기업이 다량의 온라인 리뷰를 자동으로 분석 및 활용하는데 도움이 될 수 있을 것으로 기대한다.

주제어 : 스마트 디바이스, 소셜미디어, 소비자 감성, 감성 분석, 딥러닝, 감성 사전, 양방향 장단기 기억

Abstract Due to the spread of smart devices and social media, and the increase in product purchases online, many companies are trying to understand consumers' consumption patterns and thoughts. Accordingly, the need to understand consumers' emotions by collecting reviews including consumers' opinions on products or services online is increasing, and related research is being conducted by domestic and foreign companies and research institutes. However, most of the studies are still focused on data expressed in English, and many studies and results on sentiment analysis as a lexicon or machine learning approach for English text have been published. On the other hand, the Korean language has relatively low accuracy due to the complexity of Korean and the lack of labeling data for deep learning. To improve these problems, this study utilized a hybrid approach system that improves the accuracy of sentiment analysis by utilizing the advantages of deep learning and sentiment dictionary techniques for Korean online reviews. Through this, it was confirmed how much the indicators such as accuracy, precision, and recall improved. The results of this study are expected to help companies automatically analyze and utilize a large amount of online reviews in the future.

Key Words : Smart devices, Social media, Consumer sentiment, Sentiment analysis, Deep learning, Sentiment dictionary, Bi-LSTM

본 논문은 2020년 석사학위 논문으로 수행되었음.

*교신저자 : 장경배(kbjang60@cuk.edu)

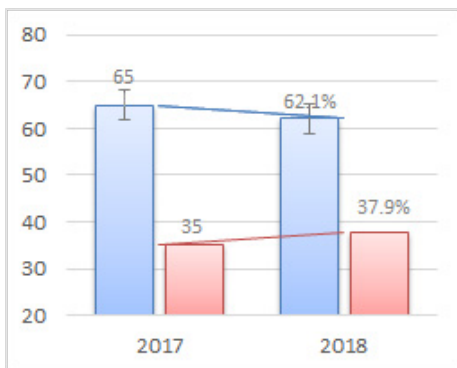
접수일 2023년 4월 22일 수정일 2023년 6월 4일 심사완료일 2023년 6월 7일

1. 서론

리뷰, 포럼, 토론, 블로그 및 뉴스 등 인터넷 상의 다양한 미디어들이 성장함에 따라 점점 더 많은 사람들이 온라인에서 자신의 견해와 의견을 공유합니다[1]. 이에 분석 가능한 텍스트 데이터의 양이 증가함에 따라 텍스트 분석을 통해 새로운 지식을 창출하고자 하는 수요가 증가하고, 이러한 수요 증가로 인해 텍스트 분석에 대한 관심과 연구 활발해지고 있다[2],[15],[16].

Fig. 1과 같이 오프라인은 62.1%로 감소된 반면에 온라인 매출은 증가하고 있고, 개인과 기업 모두 인터넷에 연결되어 언제 어디서나 판매와 구매가 이루어지고 제품이나 서비스에 대해 온라인 커뮤니티나 블로그에 의견을 제출하거나 이미 구매하여 사용하고 있는 사용자의 의견을 듣고 구매 여부를 결정하는 경험과 의견을 공유한다.

이렇듯 온라인에서 다양하고 많은 정보들이 생산되고 있고 빠르게 확산되어 영향을 줄 수 있는 환경이 구축되면서 기업은 실시간으로 VoC(Voice of Customer)를 수집하여 소비자의 의견을 청취하고 이해하여 이를 기업 활동에 반영하는 것이 중요하다. 더욱이 모바일 시장 백서에 따르면 한국의 스마트폰 이용률은 약 92%로 미국의 78%보다 높은 세계 최고이고 사용자의 활동성과 관련 있는 카테고리별 MAU(Monthly Active User)를 보면 미국은 커뮤니케이션과 생산성 순인 데 비해 한국은 커뮤니케이션과 라이프스타일 순으로 온라인에서의 활동성에 대한 관심과 의존도를 보임으로써 온라인 리뷰의 중요성을 알 수 있다.



[Fig. 1] Development of Online-Offline Sales number

최근에는 블로그, SNS, 제품홍보 등에서 작성되는 특정 제품 또는 기업에 대한 글들이 기업의 브랜드 차치를 결정하거나 다른 소비자들의 구매의사를 결정하는데 큰

역할을 하고 있다. 이에 각 기업에서는 이러한 여론 동향이나 작성된 내용의 긍정 또는 부정적 감정을 파악하고자 한다[3],[4].

셀 수 없이 쏟아지는 이러한 텍스트 자원들은 자동화된 정보 분석기법을 통해 소비자들의 동향을 파악하기 위해 다양한 연구가 진행되고 있다. 여기에 적합한 분석 기법으로는 오피니언 마이닝 분야 중 감성분석기법을 활용해야 하는데, 오피니언 마이닝이란 텍스트에 포함된 내용이 주관적, 객관적인지 판별하고 주관이 드러난 내용을 분석하는 기법이다[1],[5].

이러한 텍스트 자원들은 비정형 형태를 가지고 있으며, 비정형 형태의 텍스트 자원들을 통해 사용자들이 가지고 있는 감정을 이해하려면 텍스트를 구성하는 기본 단위에 대한 감정 정보를 가지고 있어야 하는데, 감정 정보를 얻을 수 있는 적절하고도 공개된 자원이 마련되어 있지 않아 감정 인식과 관련된 연구와 서비스가 활성화 되지 못하고 있다[6].

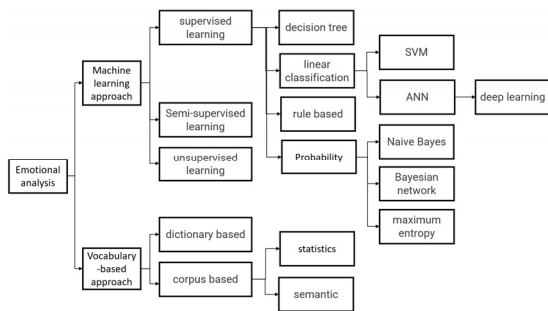
따라서, 이러한 비정형 데이터들은 구조화되지 않은 대규모 텍스트에서 유용한 정보를 추출하여 분류(Classification) 또는 요약(Summarization)하는 텍스트 마이닝(Text mining)과 상품 평을 작성한 소비자의 감성(Sentiment)을 추출하는 오피니언 마이닝(Opinion mining) 기술 중 하나인 감성분석을 활용한다.

감성분석은 어떤 주제에 대한 주관적인 감정이나 태도와 같은 의견을 텍스트로부터 뽑아내는 기술로 일반적으로 긍정(Positive), 부정(Negative), 중립(Neutral)으로 극성(Polarity)을 분류하는 것을 기본으로 슬픔, 기쁨 같은 감정의 상태도 분석할 수 있어, 기업이 신제품 개발이나 제품에 대해 소비자들의 의견을 분석함으로써 불만과 만족하는 이유를 알 수 있어 신속하고 효과적인 대응을 할 수 있고 경쟁사의 제품을 분석하여 산업 동향을 파악하고 브랜드나 기업에 대한 부정적 이슈 또는 평판 관리를 위하여 온라인 리뷰에 대한 감성 분석이 필요하다[2],[6].

2. 관련 연구

아래 Fig. 2와 같이 감성분석은 문맥에서 제품이나 서비스에 대한 저자의 논조, 태도, 감정 등을 정량화하는 유용한 기술로 일반적으로 사전이나 말뭉치(Corpus)를 사용하는 어휘기반 접근법과 머신러닝 접근법으로서의 사결정나무 (Decision tree), SVM (Support Vector Machine), 신경망, 규칙 기반, 확률(나이브 베이즈

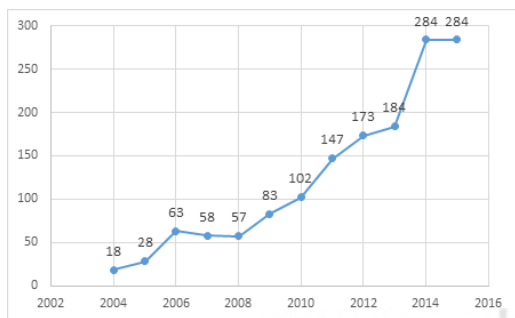
(Naive Bayes)), 베이지안 네트워크, 최대 엔트로피) 등의 기법을 사용하는 지도 학습, 비지도 학습, 반지도 학습의 방식, 여러 기법을 결합한 하이브리드(Hybrid) 접근법 그리고 이모티콘, 퍼지 세트(Fuzzy Set), 온톨로지(Ontology) 등의 기법을 사용한다. 최근 감성분석에서 인공지능망(Artificial neural networks)의 딥러닝(Deep Learning) 학습방식을 이용하는 연구가 늘고 있다[7-10].



[Fig. 2] Sentiment Classification techniques

학술연구에서의 텍스트 마이닝 활용 현황 및 주요 분석기법 논문에 따르면 디비피아를 활용하여 텍스트 마이닝 연구 현황 1570개의 논문 명에 대해 주요 연구주제 및 분야와 분석기법들을 확인하였는데 Fig. 3과 같이 1998년부터 꾸준히 증가하는 추세이며 2011년의 증가율은 44% 2014년 증가율은 55%로 2011년 이후 텍스트 마이닝 관련 연구가 빠르게 증가하고 있음을 알 수 있고 전체 177개의 다양한 학회에서 텍스트 마이닝을 활용한 학술연구가 꾸준히 증가하였고, 상위 7개 학회 논문이 전체논문의 50% 이상을 차지하는 것은 아직 일부 학술분야에서만 텍스트 마이닝 기법을 연구에 활용하고 있음을 의미한다[11].

또한 키워드 및 토픽분석을 통해 오피니언 마이닝, 감



[Fig. 3] Text mining paper by year

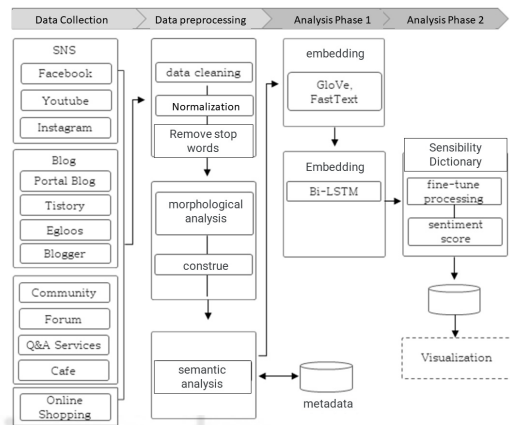
성분석, 토픽모델링 등이 텍스트 마이닝 연구에 주로 활용되고 분석 자료로는 SNS, 고객리뷰, 특허정보가 있으며 활용분야로는 기술예측, 마케팅 등의 분야에서 활용되고 있음을 확인하였다.

‘Recent Trends in Deep Learning Based Natural Language Processing’ 논문을 통해 딥러닝 아키텍처 및 알고리즘은 이미 컴퓨터 비전 및 패턴 인식과 같은 분야에서 놀라운 발전을 이루었고 수년 이전의 머신러닝 접근법은 매우 높은 차원의 희소 특징(Sparse features)에 대해 훈련된 얇은(Shallow) 모델 예를 들면 SVM 또는 로지스틱 회귀 등을 기반으로 하는데 최근 몇 년간은 밀집 벡터(Dense vector) 표현을 기반으로 한 신경망이 다양한 NLP 테스크에서 우수한 결과를 만들어 냈는데, 워드 임베딩과 딥러닝 기법의 성공에 의하여 시작되었으며, 딥러닝은 멀티레벨 자동 피쳐 표현 학습을 가능하게 한다고 하였다[12].

3. 연구 방법

3.1 감성분석 시스템 아키텍처

제안하는 하이브리드 감성분석 시스템은 화장품 도메인에서 제품이나 서비스에 대한 한글 온라인 리뷰에 대해 딥러닝 접근법의 Bi-LSTM 기법을 이용하여 극성의 방향을 위한 이진분류를 수행한 후, 정확도 평가 및 향상을 위해 감성사전 접근법을 통해 제품의 속성(Attributes)과 관점(Aspect)에서 5단계의 세분화된 감성지수에 따른 감성분석의 정확도가 기존의 연구와 동등하거나 향상되는 것을 목표로 한다. Fig. 4와 같이 이 시스템은 단계별로

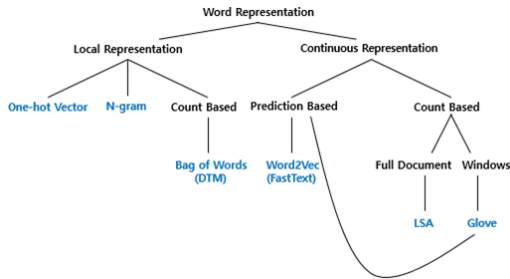


[Fig. 4] System Architecture

진행되며 감성지수를 정교화하고 정확도를 높여나가는 분석 파이프라인 구조이다.

3.2 단어 임베딩

딥러닝을 이용하여 기계번역뿐만 아니라 자동 요약, 담화 분석(Discourse Analysis), 자연어 생성 및 이해, 감성 분석, 스피치 인식(Speech recognition), 이미지 분류, 물체의 위치나 움직임 인식, 자율주행, 체적의 뇌 이미지 분할 등 응용분야가 많다. Fig. 5를 보면 쉽게 이해할 수 있다. 이러한 다양한 응용을 위해선 우선 컴퓨터 즉 인공신경망이 이해할 수 있도록 텍스트를 숫자의 나열인 벡터로 변환하기 위한 단어 임베딩은 텍스트 데이터를 숫자로 이루어진 벡터 공간에 매핑(Mapping)시키는 과정이다.



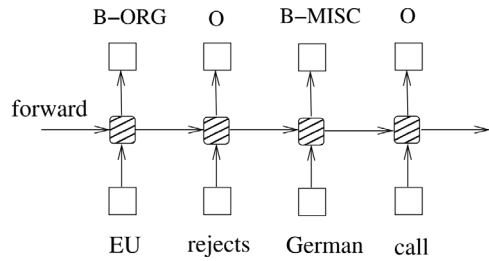
[Fig. 5] Variety of word assortment

이러한 임베딩은 범주형 자료를 연속형 벡터 형태로 변환시키는 것으로 차원을 축소할 수 있고, 연속형으로 표현할 수도 있으며 의미를 도출하기 용이하다. 임베딩 모델 중 Word2Vec과 확장한 FastText는 구글이 공개한 모델로 Skip-gram과 CBOW (Continuous Bag Of Words)가 제공되고, FastText는 페이스북에서 개발 공개한 모델로 각 단어를 문자 단위 n-gram으로 표현하고, GloVe는 스탠포드대학에서 개발한 임베딩 기법으로 Word2Vec의 단점을 극복하고자 시도됐다.

3.3 양방향 장단기 메모리

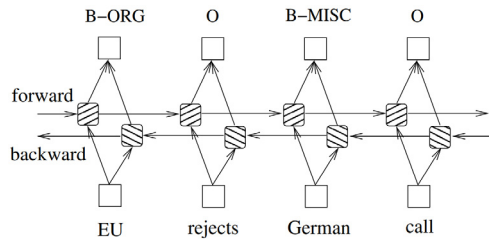
시퀀스 라벨링(Sequence labeling) 태스크를 해결하는 딥러닝 모델이 발표되면서 다양한 구조의 순환신경망이 감성분석에 적용되고 있지만 순환신경망은 일반적으로 문장이 길어질 경우 성능이 저하된다. 이러한 장기의 존성의 문제를 극복하고자 하는 다양한 구조의 순환신경망의 모델들이 제안되었으며, 그 중 대표적인 모델로 Fig. 6과 같이 LSTM(Long Short-Term Memory)이

있다. LSTM은 RNN 모델이 가지고 있는 단점인 시퀀스(Sequence)의 길이가 길어질수록 역전파시 기울기가 완만해져 경사도가 사라지는 문제인 Vanishing Gradient Problem을 해결한 모델로서, 순환신경망의 오랜 학습시간, 불필요한 연산, 오차의 사라짐 현상 등의 문제를 해결하고자 제안되었는데, 입력 값과 이전 시점에서 전달 받은 값을 입력 값으로 하여 활성화수를 활성화시켜 노드의 상태를 결정한다. 그러나 LSTM은 은닉 층의 노드에 분리된 메모리 공간(Cell)을 설정하고 그 메모리 공간의 상태 값들이 입력되거나 출력되는 길목에 특수한 게이트(Gate)를 설치하여 장기의존성 문제에 접근하였다.



[Fig. 6] LSTM network

Fig. 7과 같이 기존의 Bi-LSTM(Bidirectional Long Short Term Memory)은 입력을 양방향으로 받는 것 외에 전체적인 셀 스테이트(Cell State)의 구성은 LSTM과 같고, 최적화 알고리즘으로 adam을 선택하며 파라미터를 최적화하고 Epoch는 10, batch size는 50으로 한다.



[Fig. 7] Bi-LSTM network

3.4 수집과 전처리

3.4.1 데이터 수집

인터넷에 흩어져 있는 웹 페이지의 다양한 데이터를 수집하는 기술은 서비스의 품질을 결정할 수 있어 중요하다. 분석을 위한 수집 대상은 유튜브, 인스타그램, 페이스북, 블로그, 커뮤니티, 포럼, 카페, 온라인 쇼핑몰, 해당 기업의 웹사이트에서 스크래핑(Scraping)하는데,

통상 크롤링(Crawling), OpenAPI, RSS 등의 방법으로 자동화하여 수집하나 본 연구는 지정된 사이트에서 필요한 데이터만 수집하는 스크래핑을 통해 수집하고, 이 온라인 리뷰는 구조화되지 않은 비정형 데이터이므로 분석과 활용 그리고 처리 비용을 고려하여 전처리한 csv 형태의 파일로 저장하며 처리 과정에서 메타데이터와 최소의 필요 정보만 데이터베이스에 별도 저장하여 이후 실시간 분석과 시각화(Visualization)를 용이하게 한다.

파이썬에서 수집을 위해 일반적으로 request 라이브러리를 많이 사용하나, 본 연구에서는 자바스크립트의 동적인 변화까지 자동화할 수 있는 selenium 라이브러리를 사용하여 데이터는 화장품 도메인으로 포털사이트의 해당 전문 카페, 전자상거래 사이트 등으로부터 일부 수집하고 오픈 데이터 셋과 혼합하여 사용한다.

데이터 셋을 훈련 데이터와 테스트 데이터로 분할할 때 가장 이상적인 비율은 데이터의 성격과 목적에 따라 일반적으로 80대 20 또는 70대 30으로 나누어 수행하는데, 본 프로젝트에서는 비교를 위해 8:2의 비율을 적용한다.

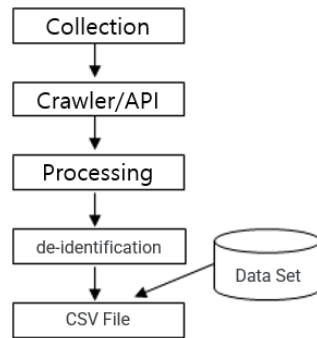
3.4.2 전처리

전처리의 주요 작업은 결측값(Missing Value)을 채우거나, 잡음값(Noisy Data) 완화, 이상점(Outlier)을 발견하여 이를 제거하는 정제(Data Cleaning) 과정, 데이터 통합(Data Integration), 데이터 정리(Data Reduction), 데이터 변환(Data Transformation) 등이며 자연언어의 분석 과정은 구조화되지 않은 텍스트 데이터인 온라인 리뷰에 대해 전처리 과정이 필요한데 이 과정은 가장 시간이 많이 소요되고 연구 결과에 영향을 줄 수 있다. 통상 분석을 위한 전처리 과정에서 이모티콘, 관용어, 등을 불용어로 분류하여 삭제하나 본 연구는 관용어, 일부 이모티콘 등을 분석 대상에 포함시켜 저자의 감성이 왜곡 또는 편향되지 않도록 하며, 첫 단계는 이모지(Emoji)와 특수문자, 주민등록번호, 여러 개의 '\n' 등을 삭제한 후 문자열을 쉽게 처리하기 위해 정규화표현식을 활용한다. 또한 결측 값이나, 특이 값의 NaN 등은 빈 공간으로 대체한다.

- (1) Data cleaning : 감성분석에 무의미한 데이터를 삭제 또는 filtering
 - HTML 태그 (BeautifulSoup 이용)
 - 쉼표, 느낌표, 물음표, 이모지 등 특수문자 및 불용어 제거 (정규표현식을 이용)
 - 여러 개의 CR, LF 제거

- 결측값, NaN 등 제거 또는 빈 공간으로 대체
- (2) 개인정보의 비식별화 처리 : 비식별화 (de-identification) 단계별 조치 대상 수집과 동시에 아래와 같은 개인 식별정보는 모두 삭제 처리
 - 성명, 주소, 주민등록번호, 여권번호, 외국인등록번호 등
 - 계좌번호, 신용카드번호, 각종 자격증 및 면허 번호, 이미지, 신체 식별정보 등

본 연구에서 제안하는 감성분석 시스템의 처리 프로세스는 다음과 같다.

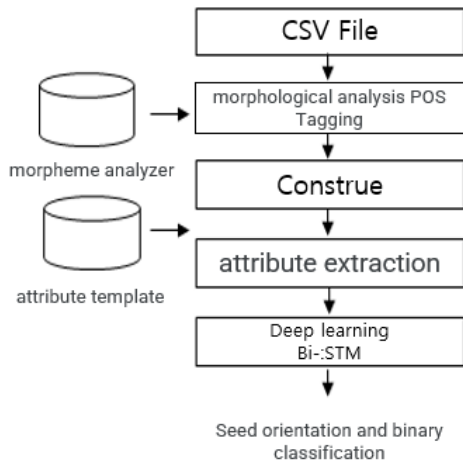


[Fig. 8] Preprocessing Process

3.5 형태소 분석

의미가 있는 최소의 단위인 형태소로 분리하고 품사정보를 얻기 위해 형태소분석기를 이용하는데 불 용어를 제거한 이후 형태소 분석에서 최종 분석까지 본 연구방식과 같은 파이프라인 분석 처리 구조에서는 부정확한 품사정보로 인해 이전 단계의 오류가 다음 단계로 전파되는 문제가 존재하므로 형태소 분리 태스크가 매우 중요하며 이를 보완해야 한다. 이때 형태소 분석 이전에 띄어쓰기가 중요하다.

범언어적 자연어처리를 위한 NLTK 패키지는 한글처리에 한계가 있으므로 파이썬에서 사용할 수 있고 한국어 형태소 분석기를 포함하는 KoNLPy(Korean NLP in Python) 패키지를 사용한다. 이 패키지에서 제공하는 형태소 분석과 문장의 품사 정보를 판별하는(POS tagging) 분석기를 이용하여 의미를 가진 최소의 문법 단위로서 자립 가능한 단위인 형태소의 품사 정보를 확보한다.



[Fig. 9] Sentiment Analysis flow chart

3.6 구문분석과 속성 처리

대량의 비정형 데이터를 자동으로 효과적인 극성 분류를 위해 리뷰 글의 주제와 그 속성과 관점을 추출한 후 긍정·부정의 판단을 한다. 구문분석을 통해 명사구, 동사구, 형용사구를 추출하여 활용하고, POS 태깅을 통해 얻은 품사 정보 중 명사를 이용하여 제품 또는 서비스, 속성에 대한 정보를 추출하는데 특정 도메인에서는 긍정적일 수 있지만 다른 도메인에서는 부정적일 수 있어 이와 같이 속성에 따른 극성이 도메인에 따라 달라지는 것이 감성분석의 어려운 과제이나 본 연구는 특정 도메인을 대상으로 접근방식에 의한 정확도 향상이므로 생각한다.

속성을 추출하기 위해 Table 1과 Table 2 같이 연관 배열(Associative array)이라는 자료구조의 파이썬 딕셔너리를 이용하여 연구자가 수작업으로 작성한 속성 템플릿을 이용한다.

<Table 1> Cosmetics properties by category

Moisturizer	Makeup-Lip
Moisturizing power	persistence
absorption power	Product effects/features
trouble/stimulation	Product scent
sense of use	Product Formulation
Product scent	Color
Product effects/features	adhesion
Ingredients/Contains	Color development
Multiple	Multiple
etc	etc

<Table 2> Likert scale's 5 Sentiment scale

Sortation	1	2	3	4	5
stage	strong denial	Denial	neutrality	Positive	strong Positive

3.7 감성단어사전

한국어 감성사전은 세종 구문 분석 코퍼스 중에서 조선일보 생활, 사회면과 한국일보, 한겨레신문에서 332개 기사, 7744 문장을 선정하여 주석 처리한 서울대 학교언어학과 컴퓨터 언어학 연구실에서 개발한 KOSAC(Korean Sentiment Analysis Corpus)과 표준국어대사전의 뜻풀이의 감성을 Bi-LSTM을 활용하여 89.45%의 정확도로 분류하였으며 긍정으로 분류된 뜻풀이에서는 긍정에 대한 감성 어휘를, 부정으로 분류된 뜻풀이에서는 부정에 대한 감성 어휘를 1-gram, 2-gram, 어구 그리고 문형 등 다양한 형태로 추출하고 다양한 외부 소스(SentiWordNet, SenticNet, 감정동사, 감성사전0603)를 활용하여 감성 어휘를 확장하였으며 온라인 텍스트 데이터에서 사용되

<Table 3> Example of characteristic and point of view

Domain	Products /Services	Property	관점	내용
Small appliances	cellphone	Battery	- Weight - lifespan	- Positive : 오래 간다 - Denial : 무겁다
		Screen	- Size - sharpness	- Positive : 시원하다 - Denial : 무겁다
Cosmetics	Lipstick	Color	- kind - texture	- Positive : 맑다. 선명하다. 밝다. - Denial : 탁하다, 어둡다, 뵈다.
		Moisturizing power	- durability - water content	- Positive : 오래, 촉촉 - Denial : 마름, 거침, 트러블
		persistence	- kind	- Positive : 하루 종일 편하게, 처음처럼 - Denial : 메마름
		applicability	- touch	- Positive : 바르는 순간, 부드럽다, 촉촉 - Denial : 부담, 거칠다

는 신조어, 이모티콘에 대한 감성 어휘도 포함하고 있는 KNU 한국어 감성사전이다. 이 두 개의 사전은 모두 화장품 도메인에 최적화되어 있지 않으므로 연구 목적 상 특정 도메인에 영향을 받지 않는 14,843개의 감성 어휘로 구성되어 있고 기본적인 감성 분석의 수행에 활용될 수 있는 KNU감성사전을 파이프라인 분석 단계의 하나로써 이용한다.

딥러닝 기법으로 얻은 시드 감성점수에 대해 정확률을 개선하기 위해 감성사전을 이용 극성의 방향이 옳은지 여부와 극성의 정도를 가감하여 세분화된 수치로 표현한다. 본 연구에서 감성 지수는 Table 3과 같이 부정은 1~2(2단계), 긍정은 4~5(2단계), 그리고 중립으로 분류하며 감성사전에서 해당되는 단어가 없는 경우 향후 빈도수 계산 및 피드백 대상으로 고려하기 위해 보관한다.

4. 연구 결과

4.1 데이터 셋과 평가결과

4.1.1 데이터 셋 및 실험환경

본 연구에서는 GPU 등의 개발환경을 이용하기 위해 구글의 코랩을 활용하고, 두 개의 데이터 셋, 하나는 직접 수집을 위해 네이버 블로그 및 화장품 전문 카페에서 화장품을 키워드로 주고 검색결과를 스크래핑 하여 수집하고, 다른 하나는 한국경영학회 연구용 빅 데이터 레파지토리에서 다운로드한 한국어 속성단위 감성분석(Aspect-based Sentiment Analysis)을 위한 화장품 상품평 데이터를 이용하여 평가한다.

데이터 셋은 연구를 위해 수집한 10,000개의 화장품 상품평 데이터이고, 속성은 상품 평에 실제 등장한 aspect term과 aspect term의 상위개념인 aspect category 20개로 구성되어 있고 polarity는 각 속성에 대한 감성을 7,000개의 positive와 3,000개의 negative로 분류되어 있는 데이터 셋이다.

4.1.2 평가 방법

실험 평가를 위해 모델의 성능을 평가하는 지표 중 하나로 많이 사용하는 정확도(Accuracy)와 재현율(Recall), 정밀도(Precision) 그리고 가중조화평균인 F1 score로 평가한다.

(1) 정밀도(Precision)

모델이 True로 분류한 것 중에서 실제 True 인 것의 비율이다.

$$\text{Precision} = \frac{TP}{TP + FP}$$

(2) 재현율(Recall)

실제 True인 것 중에서 모델이 True로 예측한 비율이다.

$$\text{Recall} = \frac{TP}{TP + FN}$$

(3) 정확도(Accuracy)

True와 False를 맞게 예측한 비율이다.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

(4) F1 Score

Precision과 Recall의 조화평균으로 아래의 식과 같다.

$$F\text{-score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

4.2 실험 및 평가

4.2.1 장단기 메모리로 감성 분류

가. 데이터 준비

구글의 코랩 가상환경에 한국어 파이썬 자연어처리 패키지들 “!pip install konlpy” 명령어로 설치하고, 구글 드라이브에서 데이터를 review와 속성 값 그리고 polarity의 정보를 Pandas로 연동하고 10,000건의 데이터 셋을 위 소스코드와 같이 train_set과 test_set을 8:2의 비율로 나누어 데이터의 분포를 고르게 하도록 하였다. 또한 다양한 알고리즘에 따라 적절한 파라미터가 설정되어야 하고 그에 따라 결과가 크게 달라지므로 Scikit-learn의 그리드 서치(GridSearch) 함수 등을 이용, 아래의 모델과 같이 최적의 파라미터를 찾았다. 또한 화장품 데이터 셋의 positive, negative로 표기되어 있는 극성을 딥러닝에서 처리하기 위해 2진화 처리한다.

아래의 코드와 같이 형태소분석기의 품사정보를 이용하여 형용사와 부사를 추출, 상위 35,000개의 단어만 토 큰화하고 리뷰의 최대길이를 12로 설정한다.

```
path = '/content/drive/My Drive/proj'
df = pd.read_csv(path + '/sentiment_dataset.csv', sep=',', encoding='utf-8')
df.head()
```

	review	aspect	aspect	polarity
		tera	category	
0	검은 피부엔 맞지 않네요... 뽀얀 피부엔 괜찮아요... 색상 조금 어두워요	색상	색상	negative
1	폼 신상 색상 너무 저렴하게 나와서 모두 구매했는데 다 이쁨	색상	색상	positive
2	싸게 샀는데 ...색은 그렇다 치고 잘 번져요 엄청 건조하고 뽀얀 그냥 색 느낌만 ...	뽀얀	가격	negative
3	배송 3일이나 걸렸지만 저렴하게 구매해서 기분이 좋네요	배송	배송	negative
4	페리페라 입소문 듣고 사용해본 적은 없는데 가격 저렴하게 샀어요 처음 사용해보는 ...	가격	가격	positive

[Fig. 10] Dataset for analysis

```
from sklearn.model_selection import train_test_split
df_train, df_test = train_test_split(df, test_size = 0.2)
print(len(df_train), len(df_test))
8000 2000
```

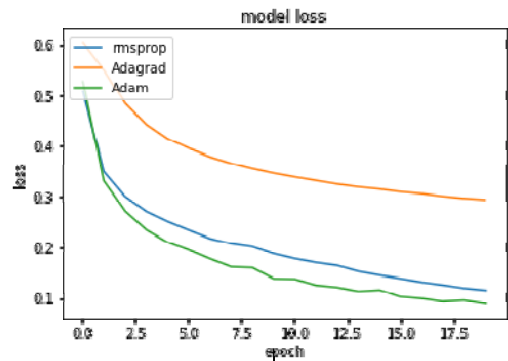
```
X_test = []
for c, sentence in enumerate(df_test['review']): #for sentence in
    review['document']:
        temp_X = []
        temp_X = okt.pos(sentence, stem = True)
        temp_X = [word for word in temp_X if not word in
stopwords]
        temp_e = []
        for i in temp_X:
            if (i[1] == 'Adverb') or (i[1] == 'Adjective'):
                temp_e.append(i[0])
        X_test.append(temp_e)
from tensorflow.keras.preprocessing.text import Tokenizer
max_words = 35000
tokenizer = Tokenizer(num_words=max_words) # 상위
35,000개의 단어만 보존
tokenizer.fit_on_texts(X_train)
X_train = tokenizer.texts_to_sequences(X_train)
X_test = tokenizer.texts_to_sequences(X_test)
```

4.2.2 모델 아키텍처

```
max_features = 15000
batch_size = 64
epoch = 100
model_lstm = Sequential()
model_lstm.add(Embedding(max_features, 100))
model_lstm.add(LSTM(128))
model_lstm.add(Dropout(0.5))
model_lstm.add(Dense(1, activation='sigmoid'))
model_lstm.compile(optimizer='Adam',
loss='binary_crossentropy', metrics=['acc',f1_m,precision_m,
recall_m]) # binary_crossentropy sparse_categorical_crossentropy
history_lstm_Adam = model_lstm.fit(X_train, y_train, epochs=
epoch, batch_size = batch_size, validation_split=0.2, verbose = 0)
```

모델을 개발할 때 학습시키면서 함수의 반환 값으로 히스토리 객체를 얻을 수 있고, 이 객체는 매 에포크 마다의 훈련 손실값(Loss), 훈련 정확도(Acc), 검증 손실값(Val_loss), 검증 정확도(Val_acc)를 히스토리 기능을 통해 확인한다.

옵티마이저는 케라스 모델을 컴파일하기 위해 필요한 파라미터 중 하나로서 해당 옵티마이저의 기본 설정을 사용하여 위의 소스와 같이 모델의 정의한 후 Fig. 11과 같이 일반적으로 RNN에서 많이 사용하는 RMSProp, 파라미터의 값이 업데이트되는 빈도에 의해 학습률이 결정되는 모델 파라미터 학습률을 사용하는 Adagrad, 그리고 RMSProp에 모멘텀이 적용된 Adam, 확률적 경사 하강법(Stochastic Gradient Descent)인 SGD까지 4 종류를 이용하여 성능평가를 다른 옵티마이저는 성능의 큰 차이가 없으므로 Adam으로 성능 평가를 진행하였다. batch_size는 한 번에 전체 학습이 어려우므로 미니배치를 사용하며 몇 개의 훈련 데이터를 사용할 지를 나타내고, iter_num은 가중치 갱신을 몇 번하는지, epoch는 하나의 미니배치에 대하여 갱신을 몇 번 해야 전체 훈련 데이터의 학습이 가능한지를 나타낸다.



[Fig. 11] Training Set's learning loss by epoch

```
1. 테스트1
sparse_categorical_crossentropy
Dropout = 0.5
Dense = 5
Adam loss: 0.8611149497032166 accuracy: 0.747 f1 score: 0.8556414 precision: 0.6885307 recall: 1.1413629

2. 테스트2
sparse_categorical_crossentropy
Dropout = 0.6
Dense = 5
Adam loss: 0.8342471299171448 accuracy: 0.757 f1 score: 0.8514397 precision: 0.6868393 recall: 1.1306947

3. 테스트3
binary_crossentropy
Dropout = 0.5
Dense = 1
Adam loss: 1.107579858191967 accuracy: 0.755 f1 score: 0.8529713 precision: 0.798817 recall: 0.87487495
```

[Fig. 12] Test parameter settings

히스토리 기능을 통해 그래프로 결과를 테스트와 트레인 데이터를 비교하며 오버피팅과 언더피팅이 되는 것을 피하여 최적의 파라미터를 찾기 위해 Fig. 11과 같이 설정하여 Table 4와 같이 결과를 도출했다.

〈Table 4〉 LSTM test result

Model	Precision	Recall	Accuracy	F1-score
Adam	0.798	0.874	0.755	0.832

또한 옵티마이저를 비교한 결과 학습이 진행될수록 오버피팅으로 인해 오히려 loss 값이 증가하는 것을 확인하여 학습 횟수에서 최적의 파라미터, 최적의 학습 순간을 찾는 것이 중요하며 결과적으로 Adam과 RMS 옵티마이저는 0.4 loss로 가장 높은 성능을 나타내고 있음을 확인할 수 있었다.

4.2.3 양방향 장단기 메모리로 감성 분류

가장 최적의 모델과 아키텍처를 찾기 위해 모델 중심적인 접근으로 LSTM과 성능의 차이를 비교하고 성능을 개선하고자 Bi-LSTM을 선택한다.

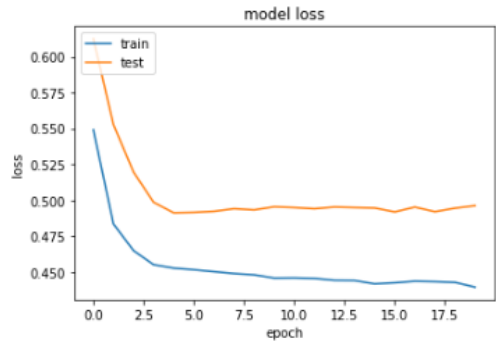
```

from tensorflow.keras.layers import Dense, Embedding, Bidirectional, LSTM, Concatenate, BatchNormalization
from tensorflow.keras import Input, Model
from tensorflow.keras import optimizers
sequence_input = Input(shape=(max_len,), dtype='int32')
embedded_sequences = Embedding(vocab_size, 128, input_length=max_len)(sequence_input)
lstm, forward_h, forward_c, backward_h, backward_c = Bidirectional(LSTM(128, dropout=0.5, return_sequences=True, return_state=True, recurrent_activation='relu', recurrent_initializer='glorot_uniform')(embedded_sequences), learning_rate = 0.0001)
state_h = Concatenate()(forward_h, backward_h) # 은닉 상태
state_c = Concatenate()(forward_c, backward_c) # 셀 상태
attention = BahdanauAttention(128) # 가중치 크기 정의
context_vector, attention_weights = attention(lstm, state_h)
hidden = BatchNormalization()(context_vector)
output = Dense(1, activation='sigmoid')(hidden)
biresult = []
op_adm = optimizers.Adam(lr=learning_rate, clipnorm=1.)
op_rms = optimizers.RMSprop(lr=learning_rate, clipnorm=1.)
    
```

모델 정의와 결과를 확인하기 위해 위, 아래의 소스코드와 같이 실험한다.

```

loss_SGD, accuracy_SGD, f1_score_SGD, precision_SGD, recall_SGD = model_SGD.evaluate(X_test, y_test, verbose=0)
print(history_SGD.history.keys())
plt.plot(history_SGD.history['loss'])
plt.plot(history_SGD.history['val_loss'])
plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
    
```



〔Fig. 13〕 Loss by epochs

Test(1)
Adam Dropout = 0.5, Dense = 1 Loss : 0.4955210409164429 Accuracy : 0.767 Precision : 0.8222986 Recall : 0.85974735
Test(2)
RMS Dropout = 0.5, Dense = 1 Loss : 0.4961269335746765 Accuracy : 0.7685 f1 Score : 0.83813566 Precision : 0.82440823 Recall : 0.8581106

〔Fig. 14〕 Bi-LSTM Paramater setting

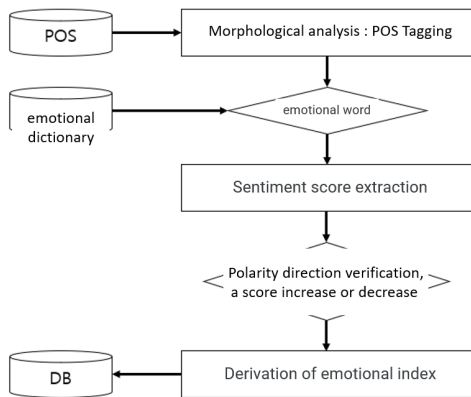
〈Table 5〉 LSTM and Bi-LSTM evaluation comparison

Model	Precision	Recall	Accuracy	F1-score
LSTM	0.686	1.130	0.757	0.851
Bi-LSTM	0.824	0.858	0.768	0.838

위와 같이 LSTM과 Bi-LSTM의 precision과 F1 score를 확인할 수 있다.

4.2.4 감성사전을 이용한 감성점수의 미세조정

원본 데이터를 전처리 후 감성사전을 이용해 형태소 분석기의 okt.nouns 함수를 이용한 명사 추출과 속성 처리, okt.pos 함수를 이용해 동사, 형용사, 부사를 추출하여 감성의 계량화 및 조정, 그리고 okt.phrases 함수를 이용한 정교화를 위해 리뷰를 대상으로 분석하고 감성지수를 추출하는 것이 이 단계의 목표로서 Fig. 14와 같이 진행된다.



[Fig. 15] Elaboration processing sequence using emotional dictionary

실험이 진행되면서 일반 감성사전의 감성지수를 본 연구자가 샘플링하여 검사한 결과 정확도가 너무 낮아 형태소 분석을 통해 동사, 부사, 형용사만을 추출하여 트레이닝 데이터로서 학습시킨 후 Bi-LSTM 기법을 이용해 F1 score를 비교하면서 최종 Table4의 결과를 도출하였다. 그러나 정확도가 낮아진 이유는 화장품 데이터 셋의 극성은 긍정, 부정 이진분류가 되어 있는 반면에 KNU 감성사전은 Table6과 같이 5단계로 테스트 데이터를 실행하는 과정에서 해당 단어가 사전에서 매칭되는 케이스조차 임의로 스케일링을 하여 진행하였다.

<Table 6> Result of Okt morpheme analyzer

원본	검은 피부엔 맞지 않네요... 보오얀 피부엔 괜찮아요... 색상 조금 어두워요
형태소	'검은', '피부', '엔', '맞지', '않네요', '...', '보오얀', '피부', '엔', '괜찮아요', '...', '색상', '조금', '어두워요'
명사	'피부', '보오얀', '피부', '색상', '조금'
품사	('검은', 'Adjective'), ('피부', 'Noun'), ('엔', 'Josa'), ('맞지', 'Verb'), ('않네요', 'Verb'), ('...', 'Punctuation'), ('보오얀', 'Noun'), ('피부', 'Noun'), ('엔', 'Josa'), ('괜찮아요', 'Adjective'), ('...', 'Punctuation'), ('색상', 'Noun'), ('조금', 'Noun'), ('어두워요', 'Adjective')
어절	'검은 피부', '보오얀', '보오얀 피부', '색상', '색상 조금', '피부', '조금'

<Table 7> 5 Scale score for dataset

Polarity of Data Set	Polarity score of emotional dictionary	evaluation
Positive	2	strong positive
	1	Positive
-	0	neutrality
negative	-1	Denial
	-2	strong Denial

따라서 순수 감성사전을 이용한 감성점수의 정확도 개선은 화장품 도메인의 감성 단어에 대해 정확도가 낮아 LSTM과 Bi-LSTM을 트레이닝 데이터로서 학습시키는데 수준으로 실험했다.

```

RMS / epoch 20 batch 60
loss: 0.46640743947824309 / accuracy: 0.7944664 / f1 score: 0.8639856 / precision: 0.81444055 / recall: 0.9246205

RMS / epoch 40 batch 120
loss: 0.4760542898646591 / accuracy: 0.7810785 / f1 score: 0.8610153 / precision: 0.8195644 / recall: 0.9117173

RMS / epoch 60 batch 60
loss: 0.4756133239891984 / accuracy: 0.76939454 / f1 score: 0.85751593 / precision: 0.8331971 / recall: 0.8889869

ADAM / epoch 20 batch 60
RMS loss: 0.4877341540209056 / accuracy: 0.7953371 / f1 score: 0.8602336 / precision: 0.82785237 / recall: 0.9014289
    
```

[Fig. 16] Hybrid approach method results

5. 결론

5.1 연구의 의의와 한계

본 연구는 레이블된 데이터가 중요한 딥러닝과 도메인 적합성에 의한 감성사전의 단점을 해결하는 두 기법의 결합을 통해 한국어 감성분석의 정확도를 높이고자 하는 연구로서 화장품 도메인에 대하여 각 기법의 개별적 활용보다 두 개의 장점을 결합하기 위해 1차 레이블된 데이터를 감성사전을 통해 추가 학습시킴으로써 정교화 및 Accuracy와 F1 score의 개선을 아래 Table 8과 같이 하이브리드 접근법에 의한 감성분석의 정확성 향상이란 결과를 보여줌으로써 하나의 연구 방향을 제시했다고 할 수 있다.

이러한 연구의 의의가 있음에도 불구하고, 첫째 영어에 비해 한국어 감성 데이터 셋의 필요성, 둘째 명사, 형용사, 관용어의 규칙기반 및 정확한 품사 정보를 얻기 위해 띄어쓰기 등의 사전처리에 많은 노력이 필요하며, 해결하는 방법으로 여러 개의 형태소 분석기를 활용하여 각 장점을 이용해야 비교적 원하는 품사 정보를 얻을 수 있다는 점, 셋째 훈련데이터를 통해 리뷰의 극성 오리엔테이션과 극성의 시드 점수 확보를 위한 단계에서 Bi-LSTM 알고리즘의 하이퍼 파라미터 튜닝에 좀 더 정밀한 연구와 실험이 부족했던 점, 넷째 감성사전의 도메인 적합성 문제를 해결하고자 시도했던 구문 및 의미 분석을 위해 연구자 중심의 템플릿을 통해 규칙 기반의 감성점수 조정을 적용하지 못함에 따라 90% 이상의 정확도 그리고 특정 도메인이 아닌 다중 도메인을 적용하는 연구가 진행되지 못했다는 점에서 한계가 있다.

REFERENCES

[1] Chen T., Xu R., He Y. and Wang X., "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," Expert Systems With Applications, Vol.72, pp.221-230, 2017.

[2] N.Y.Kim, D.H.Lee, H.H.Choi and W.X.S.Wong., "Investigations on Techniques and Applications of Text Analytics", The Journal of Korean Institute of Communications and Information Sciences, Vol.42, No. 2, pp.471-492, 2017.

[3] S.R.Hong, Y.O.Jeong and J.H.Lee, "Semi-supervised learning technique for large-scale social media sentiment analysis," Journal of the Korea Intelligent Systems Society, Vol.24, No.5, pp.482-488, 2014.

[4] Liu, B. and Zhang, L., "A survey of opinion mining and sentiment analysis," In:Aggarwal C., Zhai C. (eds) Mining Text Data, pp.415-463, Springer, Boston, MA, 2012.

[5] Y.H.Jeong, "A Study on Sentiment Classification Techniques for Korean Short Sentences Based on Machine Learning," Doctoral dissertation, Korea University, 2017.

[6] S.D.Choi and O.B.Kwon, "Study on Korean SentiWordNet development plan for big data analysis: focusing on feelings of anger," The Journal of Society for e-Business Studies Vol.19, No.4, pp.1-19, 2014.

[7] M.P.Hong, M.Y.Shin, S.H.Park and H.M.Lee, "Automatic analysis of hybrid text tone based on syntax analysis and machine learning," Korean Language and Information Society, Language and Information Vol.14, No.2, pp.159-181, 2010.

[8] Collomb, Costea, Joyeux, Hasan and Brunie, "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation," University of Lyon, France, 2014.

[9] Jurgita K., Algis K. and Tomas K "A Comparison of Approaches for Sentiment Classification on Lithuanian Internet Comments," 4th Biennial International Workshop on Balto-Slavic Natural Language Processing, pp.2-11, 2013.

[10] Medhat W., Hassan A. and Korashy H., "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, Vol.5, No.4, pp.1093-1113, 2014.

[11] S.G.Kim, H.J.Cho and J.Y.Kang, "The Status of Using Text Mining in Academic Research and Analysis Methods," Journal of Information Technology and Architecture, Vol.13, No.2, 2016.

[12] Young T., Hazarikaz D., Poria S. and Cambria E, "Recent Trends in Deep Learning Based Natural Language Processing," arXiv:1708.02709v8 [cs.CL] 25 Nov, 2018.

[13] Liu, B. and Zhang, L., "A survey of opinion mining and sentiment analysis," In:Aggarwal C., Zhai C.

(eds) Mining Text Data, pp. 415-463. Springer, Boston, MA, 2012

[14] Appel. O, Chiclana. F, Carter. J and Fujita, H, "A Hybrid Approach to the Sentiment Analysis Problem at the Sentence Level," Knowledge-Based Systems, Vol.108, pp.110-124, 2016.

[15] J.K.Lee and K.H.Y, "Trend Analysis of Barrier-free Academic Research using Text", Journal of Internet of Things and Convergence, Vol.9, No.2, pp.19-31, 2023.

[16] J.W.Mok, H.J.Jang and H.S.Lee, "HTML Tag Depth Embedding: An Input Embedding Method of the BERT Model for Improving Web Document Reading", Journal of Internet of Things and Convergence, Vol.8, No.5, pp.17-25, 2022.

김 백 기(Bae-Ki Kim)

[정회원]



- 2020년 8월 : 고려사이버대학교 융합정보대학원(정보학석사)
- 2018년 6월 : 디노플러스 연구개발 부문 부사장

<관심분야>

사물인터넷, AI, 감성분석

장 경 배(Kyung-Bae Jang)

[중신회원]



- 1995년 2월 : 광주과학기술원 기전공학과(공학석사)
- 2006년 8월 : 고려대학교 일반대학원 전기공학과(공학박사)
- 1997년 2월 ~ 2000년 3월 : SK 하이닉스반도체 주임연구원

- 2000년 4월 ~ 2008년 8월 : 현대모비스 선임연구원
- 2008년 12월 ~ 2013년 5월 : 보건복지부 국립재활원 공업연구관
- 2014년 3월 ~ 현재 : 고려사이버대학교 기계제어공학과 교수

<관심분야>

사물인터넷, 제어시스템, 로봇