

CNN을 적용한 한국어 상품평 감성분석: 형태소 임베딩을 중심으로*

박현정

이화여자대학교 경영연구소
(sparrowpark@ewha.ac.kr)

송민채

이화여자대학교 빅데이터분석학
(nicenara84@naver.com)

신경식

이화여자대학교 경영대학
(ksshin@ewha.ac.kr)

고객과 대중의 니즈를 파악하기 위한 감성분석의 중요성이 커지면서 최근 영어 텍스트를 대상으로 다양한 딥러닝 모델들이 소개되고 있다. 본 연구는 영어와 한국어의 언어적인 차이에 주목하여 딥러닝 모델을 한국어 상품평 텍스트의 감성분석에 적용할 때 부딪히게 되는 기본적인 이슈들에 대하여 실증적으로 살펴본다. 즉, 딥러닝 모델의 입력으로 사용되는 단어 벡터(word vector)를 형태소 수준에서 도출하고, 여러 형태소 벡터(morpheme vector) 도출 대안에 따라 감성분석의 정확도가 어떻게 달라지는지를 비정태적(non-static) CNN(Convolutional Neural Network) 모델을 사용하여 검증한다. 형태소 벡터 도출 대안은 CBOW(Continuous Bag-Of-Words)를 기본적으로 적용하고, 입력 데이터의 종류, 문장 분리와 맞춤법 및 띄어쓰기 교정, 품사 선택, 품사 태그 부착, 고려 형태소의 최소 빈도수 등과 같은 기준에 따라 달라진다.

형태소 벡터 도출 시, 문법 준수도가 낮더라도 감성분석 대상과 같은 도메인의 텍스트를 사용하고, 문장 분리 외에 맞춤법 및 띄어쓰기 전처리를 하며, 분석불능 범주를 포함한 모든 품사를 고려할 때 감성분석의 분류 정확도가 향상되는 결과를 얻었다. 동음이의어 비율이 높은 한국어 특성 때문에 고려한 품사 태그 부착 방안과 포함할 형태소에 대한 최소 빈도수 기준은 뚜렷한 영향이 없는 것으로 나타났다.

주제어 : 감성분석, 형태소 벡터, 단어 벡터, 딥러닝, CNN, CBOW

논문접수일 : 2018년 3월 28일 논문수정일 : 2018년 5월 29일 게재확정일 : 2018년 5월 30일
원고유형 : 일반논문 교신저자 : 신경식

1. 서론

감성분석(Sentiment Analysis) 또는 오피니온 마이닝(Opinion Mining)은 제품, 서비스, 조직, 개인, 쟁점(issues), 이벤트(events), 주제(topics) 및 속성(attributes)과 같은 객체(entity)에 대한 사람들의 의견(opinions), 감성(sentiments), 감정(emotions), 평가(appraisals) 및 태도(attitudes)를

전산적인(computational) 방법으로 분석하는 것이다(Liu, 2015; Pang and Lee, 2008; Wiebe et al., 2004, Zhang et al., 2018). 여기에서 인간의 행동(behavior) 및 의사결정(decision making)에 영향을 미치는 주요 요인이 바로 의견, 감성, 감정, 평가 및 태도라는 사실은 감성분석 연구의 중요성을 다시 한번 일깨워준다. 감성분석은 다양한 사회 및 정치 관련 이슈에 대한 대중의 의견을 분

* 이 논문 또는 저서는 2017년도 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A5B5A02024287).

석하여 정책적인 방향을 결정하거나, 온라인 쇼핑 사이트(site)나 여러 SNS에 축적되는 소비자의 의견을 분석하여 비즈니스 인사이트(insight)를 도출하는 과정에 활용될 수 있다(An et al., 2015; Cui et al., 2016; Kim et al., 2016). 뿐만 아니라, 소셜(social) 마케팅이나 정치 캠페인의 효과 측정, 주가 예측, 기업 및 개인의 신용평가, 영화 흥행 예측 등 자연어 데이터가 축적되는 초연결 가상화사회의 거의 모든 영역에 적용될 수 있다(An et al., 2015; Cui et al., Hong et al., 2016; Jeong et al., 2015; 2016; Kim et al., 2016). 그런데 이러한 감성분석은 언어 표현의 복잡성으로 인해 자연어처리(Natural Language Processing, NLP)의 주요 난제 중 하나이기도 하다(Shirani-Mehr, 2015).

최근 감성분석에 딥러닝이 효과적으로 적용될 수 있다는 사실이 입증되면서(Kim, 2014), CNN(Convolutional Neural Network) 모델을 비롯하여, Bi-RNN(Schuster and Paliwal, 1997), LSTM(Hochreiter and Schmidhuber, 1997), GRU(Cho et al., 2014) 등의 RNN(Recurrent Neural Network) 모델 및 이들을 조합하여 연결한 구조 등 다양한 딥러닝 모델들이 주로 영어 텍스트를 대상으로 적용되고 있다(Chen et al., 2017; Dhanush and Thakur, 2016; Guan et al., 2016; Jebbara and Cimiano, 2017; Shin et al., 2017; Shirani-mehr, 2015; Wang and Liu, 2015; Wang et al., 2016; Yang et al., 2016). 딥러닝 모델은 정교한 언어학적 특징을 데이터셋으로부터 스스로 학습하기 때문에 언어 종속적인 감성 단어 목록이나 파서(parser) 등이 필요 없어 상당한 분량의 수작업을 덜어준다(Mikolov et al., 2013; Socher, 2011). 뿐만 아니라 분류 성능도 다른 방법에 비해 우수하다는 결과들이 보고되고 있으

며(Kim, 2014; Socher, 2011; Wang and Liu, 2015; Zhang et al., 2018), 속성기반(aspect-based) 감성 분석 분야에서도 정교한 모델들이 경쟁적으로 소개되고 있다(Dhanush and Thakur, 2016; Guan et al., 2016; Jebbara and Cimiano, 2017; Ma et al., 2017; Ruder et al., 2016a; Ruder et al., 2016b; Tay et al., 2017; Tang et al., 2016; Wang and Liu, 2015; Wang et al., 2016; Zhang et al., 2018).

그런데 한국어 텍스트에 딥러닝을 적용한 학계의 감성분석 연구는 찾아 보기 힘들다. 이것이 굴절어(inflexional language)인 영어에 비해 조사와 어미가 발달하고 동음이의어 비중이 높은 한국어의 복잡한 특성 때문인지(Kim, 2004; Nam and Jo., 2017), 언어에 덜 종속적인 딥러닝 모델의 특성 때문인지(Mikolov et al., 2013; Ruder et al., 2016b; Socher, 2011), 또는 적절한 훈련 및 테스트 데이터셋의 부재 때문인지 정확하게 원인 규명을 하기는 어렵다. 그러나 감성분석의 중요성과 딥러닝의 무한한 가능성, 영어 및 타언어 텍스트에 대한 딥러닝 감성분석 연구의 활성화 등을 고려할 때, 한국어에 대한 딥러닝 감성분석 연구도 활발하게 축적되고 교류되어야 할 시점이라고 판단된다.

영어 텍스트에 대한 딥러닝 감성분석에서는 흔히 훈련 및 테스트 데이터에 포함된 자연어문장을 단어 벡터(word vector)들의 순열로 표현하여 딥러닝 모델에 입력한다(Kim, 2014; Tang et al., 2016; Wang et al., 2016). 단어 벡터는 주로 문장을 공백문자(space)를 기준으로 분리한 어절에 해당하는 단어(word)에 대한 벡터(vector) 표현을 말한다. 단어 벡터를 도출하는 방식은 다양하지만, 한 가지만 예로 들면, 구글(Google)이 천억 개의 단어로 이루어진 구글 뉴스 데이터를 기반으로 도출한 300차원의 Word2Vec 단어 벡터

가 있다(Mikolov et al., 2013a; Mikolov et al., 2013b). 이 단어 벡터는 뉴스 도메인이 아닌 레스토랑, 영화, 랩탑(laptop), 카메라 등 다른 여러 도메인의 리뷰(review) 텍스트 감성분석 연구에 범용적으로 활용되고 있다(Kim, 2014; Wang and Liu, 2015).

그런데 한국어는 조사와 어미가 발달한 전형적인 교착어(agglutinative language or affixing language)로서, 영어와 달리 단어(word)가 아닌 형태소(morpheme)가 감성분석이나 문장의 구조 분석에서 중요한 역할을 한다(Nam and Jo, 2017). 형태소는 ‘뜻을 가진 가장 작은 말의 단위’로 정의되며, 단어는 하나 이상의 형태소로 이루어진다(Wikibaeggwa, 2018). 예를 들어, ‘예쁘고’가 단어라면 이에 대한 형태소는 ‘예쁘(=형용사)’와 ‘고(=연결 어미)’가 된다. 이러한 형태소의 중요성을 반영하여 한국어를 대상으로 하는 감성분석에서는 단어 보다는 형태소를 기본 단위로 하는 것이 합리적으로 보인다. 따라서, 본 연구에서는 딥러닝 감성분석 모델의 입력으로 영어 텍스트에서 주로 사용되는 ‘단어 벡터(word vector)’가 아닌, ‘형태소 벡터(morpheme vector)’를 사용한다. 형태소 벡터는 형태소에 대한 벡터 표현이라 할 수 있으며, 형태소 단위로 표현된 문장에 기본적인 단어 벡터 도출 메커니즘을 적용하여 도출할 수 있다.

그런데 여기에서 몇 가지 의문점이 생긴다. 먼저, 다양한 품사의 형태소 중 어느 범위의 품사까지 고려하는 것이 딥러닝 감성분석의 정확도 향상을 위해 효과적이라는 것이다. 예를 들어, 한국어 텍스트의 사전기반 감성분석에서는 “~ 예쁘고 ~”, “~ 예쁜데 ~”, “~ 예쁘면 ~” 등에

대한 형태소분석을 통해 ‘형용사’ 형태소인 ‘예쁘-’를 추출하여, 이에 대한 감성점수를 미리 구축해놓은 감성어 사전에서 찾아 사용한다. 이것은 어미에 해당하는 ‘-고’, ‘-는데’, ‘-면’ 등의 형태소는 버리고, 어간에 해당하는 ‘형용사’ 형태소만 고려하여도 대부분의 감성을 파악할 수 있다는 가정이 전제되어 있다고 할 수 있다. 딥러닝 감성분석에서도 이와 같은 가정을 그대로 적용할 수 있을 것이다. 반면에 어미 ‘-고’, ‘-는데’, ‘-면’ 등은 앞뒤 구문을 대등, 상반, 또는 조건적으로 연결하여 문장의 극성을 유지하거나 전환시키므로 딥러닝 감성분석의 자질(feature) 값으로 활용하는 방식을 선택할 수도 있다. 또, 이러한 어미는 포함한다고 해도 감성의 극성과는 별로 관계가 없어 보이는 ‘체언 접두사’, ‘명사파생 접미사’, ‘형용사파생 접미사’ 등의 형태소까지 고려했을 때 감성분석 결과에 오히려 부정적인 영향을 미치는 건 아닌지 등이 분명하지 않다.

그리고 영어에 비해 한국어는 동음이의어 비율이 매우 높으므로¹⁾ 단어 벡터 모델을 한국어의 형태소 벡터 도출에 적용했을 때 문제점은 없는지, 이를 보완할 방안은 없는지 등도 주목할만한 이슈로 떠오른다. 기존 단어 벡터 모델은 단어의 형태를 기준으로 단어에 대한 벡터 표현을 도출하기 때문에 형태는 같은데 서로 다른 의미를 갖는 동음이의어에 대한 벡터 표현이 희석되는 한계점을 지니고 있기 때문이다. 이외에도 문법적인 오류 및 변형이 많은 한국어 상품평을 형태소 벡터 도출을 위한 데이터로 사용할 때 맞춤법이나 띄어쓰기 등의 전처리 여부에 따라 분류 정확도가 어떻게 변화할지도 미심쩍

1) 현대 표준 한국어에서 품사 구분을 하지 않은 경우 동음이의어 비율은 약 30% 정도이다(Kang, 2005).

은 부분이다.

이와 같이 본 연구는 다양한 딥러닝 모델을 한국어 텍스트에 적용할 때 가장 먼저 부딪히게 되는 기본적인 이슈들에 대한 실증적인 해답을 찾고자 한다. 본 연구에서 이러한 이슈들을 정리하여 도출한 세 가지 주요 연구 질문은 다음과 같다. 첫째, 한국어에서는 문법이 잘 지켜지지만 분석 대상과 다른 도메인의 텍스트로 훈련한 형태소 벡터와, 같은 도메인이지만 그리 문법적이지 않은 텍스트로 훈련한 형태소 벡터 중 어느 것을 딥러닝 모델의 초기 입력으로 사용하는 것이 더 효과적일까? 둘째, 품사의 범위, 동음이의어, 텍스트 전처리, 최소 빈도수 등의 측면에서 한국어에 적합한 형태소 벡터 도출 방법은 무엇이며, 어떤 방법이 딥러닝 감성분석의 정확도 향상을 위해 유리할까? 셋째, 딥러닝을 한국어 감성분석에 적용했을 때 어느 정도 만족할만한 정확도를 얻을 수 있을까?

이러한 연구 질문들에 대한 접근 방법으로, 먼저 연구 질문을 반영한 형태소 벡터 도출 대안들을 선정하고, 이에 대해 비정태적(non-static) CNN(Convolutional Neural Network) 모델을 적용하여 분류 정확도를 비교분석한다. 훈련 및 테스트 데이터로는 네이버 쇼핑의 화장품 상품평 17,260건을 사용한다. 그리고 형태소 벡터 도출을 위해서는 분석 대상과 다른 도메인이면서 구글 뉴스 데이터와 대응된다고 할 수 있는 네이버 뉴스 데이터 약 52만건 및 같은 도메인의 네이버 쇼핑 화장품 상품평 약 2백만건을 활용한다.

이후 본 논문의 구성은 다음과 같다. 먼저 2장에서 이론적 배경 및 관련 연구에 대해 정리하고, 3장에서는 데이터셋 및 실험 방법에 대해 설명한다. 그리고 4장에서는 실험 결과를 정리하고 5장에서 결론 및 향후 연구 방향을 제시한다.

2. 이론적 배경 및 관련 연구

2.1 단어 벡터(word vector)

단어 벡터(word vector)는 앞에서 언급한 바와 같이 단어(word)를 실수(real number) 원소로 이루어진 수십에서 수백 차원의 벡터(vector)로 변환한 것이다. 이렇게 변환된 단어 벡터들은 단순한 수치 이상의 의미적 자질들을 포함하고 있다. 실제로 의미적(semantic) 문법적(syntactic) 성질이 비슷한 단어들은 벡터 공간상에서 유클리디안(Euclidean) 거리나 코사인유사도(cosine similarity) 거리가 가까운 벡터들로 표현된다. 신경망(neural networks)을 이용하여 단어를 벡터로 표현하는 대표적인 방식으로는 2013년 구글(Google)에서 공개한 Word2Vec(Mikolov et al., 2013a; Mikolov et al., 2013b) 외에, 2014년 Stanford 대학 NLP 연구진에 의해 제안된 Glove(Pennington et al., 2014), 2016년 Facebook 연구진에 의해 제안된 fastText(Joulin et al., 2016) 등이 있다.

이 중 Word2Vec 모델은 내부적으로 CBOW(Continuous Bag-Of-Words)나 Skip-Gram이라는 신경망 구조를 이용해 단어들의 벡터 표현을 학습하게 된다. CBOW는 컨텍스트 단어들로부터 타겟 단어를 예측한다. 예를 들어, “This lipstick is beautiful in _____ and has a good sustainability.”라는 문장에서 _____ 앞 뒤에 나오는 컨텍스트 단어들로부터 _____에 해당되는 ‘color’와 같은 타겟 단어를 예측하는 방식이다. Skip-Gram 구조는 타겟 단어로부터 컨텍스트 단어들을 역으로 예측한다. CBOW는 Skip-Gram보다 훈련에 소요되는 시간이 짧고 고빈도 단어에 대한 정확도가 높은 반면에, Skip-Gram은 훈

런 데이터의 양이 적을 때나 저빈도 단어에 대한 정확도가 요구될 때 적합하다고 한다(Mikolov et al., 2013a; Mikolov et al., 2013b).

2.2 언어의 분류와 한국어 텍스트 분석

세계의 다양한 언어들은 구조나 형태의 관점에서 대개 고립어(isolating language), 교착어(agglutinative language or affixing language), 굴절어(inflexional language) 등 세 종류로 분류된다(Kim, 2004). 고립어에서는 단어가 형태상의 변화 없이 문장을 구성하고 다른 단어와의 문법적인 관계는 어순에 의해 표현된다. 굴절어에서는 단어 자체의 형태변화를 통해 해당 단어의 문법적인 기능을 나타낼 수 있다. 교착어는 고립어와 굴절어의 중간적인 특성을 가진 것으로, 어근(root)에 접사(affix)가 결합되는 형태로 문장 내에서의 단어의 기능을 표현한다.

대표적인 고립어로는 중국어, 교착어로는 한국어, 굴절어로는 영어를 들 수 있다. 예를 들어, 한국어의 “나는 책을 읽는다.”라는 문장에서, ‘나’에 ‘는’이, ‘책’에 ‘을’이 부착되어 ‘나’와 ‘책’의 문법적인 기능이 표현된다. 굴절어인 영어는 “I read a book.”으로 ‘I’는 ‘나’라는 뜻 외에 ‘는’(주격)의 뜻을 가지고 있다. 이 문장을 고립어인 중국어로 표현하면 ‘我看書’가 된다. 중국어의 ‘我’에는 ‘나는’이나 ‘I’처럼 문법적인 기능의 표시가 포함되어 있지 않다.

영어의 ‘drive-drove-driven’과 같은 예를 보면, 어휘적 의미와 문법성이 한 단어 안에 밀접하게

결합되어 있음을 더욱 실감할 수 있다. 물론 이러한 구분이 항상 모든 경우에 해당되는 것은 아니다. 영어의 명사(noun)는 과거에는 굴절이 많았으나, 근래에는 소유격 -’s 이외의 격변화를 상실하고 격관계의 표시를 어순에만 의존하는 고립어에 상당 부분 가까워졌다. 중국어의 경우에도 문법적인 기능을 표현하는 말이 아예 없는 것은 아니다.

그렇지만 결국 한국어가 영어나 중국어에 비해 조사와 어미를 포함한 형태소의 역할이 크다는 점에서(Nam and Jo., 2017), 한국어 텍스트에 대한 감성분석은 형태소를 단위로 하는 것이 효과적일 수 있다. 예를 들어, “이 립스틱은 컬러도 예쁘고 지속력도 좋다.”라는 문장이 있을 때, 어절 단위로 분리한 단어를 CBOW에 대한 입력으로 사용하면 ‘립스틱은’이나 ‘컬러도’와 같은 단어 벡터를 얻게 될 것이다²⁾. 그런데 이렇게 하면 ‘립스틱을’, ‘립스틱이’나 ‘컬러가’, ‘컬러를’, ‘컬러는’ 등도 모두 다른 단어로 인식하게 되므로, 고려해야 할 단어 수가 비효율적으로 증가하게 된다. 그런데 이 문장을 형태소분석하면, 선택한 형태소분석기나 분절 정도에 따라 약간 달라지겠지만, 다음과 같은 결과를 얻을 수 있다.

“이(=관형사) 립스틱(=일반 명사) 은(=보조사) 컬러(=일반 명사) 도(=보조사) 예쁘(=형용사) 고(=연결 어미) 지속력(=일반 명사) 도(=보조사) 좋(=형용사) 다(=종결 어미) .(=마침표)”

이로부터 “이 립스틱 은 컬러 도 예쁘 고 지속력 도 좋 다.”와 같은 형태소 단위의 문장 표현을 얻을 수 있고, 이것을 CBOW에 입력하면 형

2) 단어는 ‘스스로 일정한 뜻을 담고 있고, 자립성이 있는 최소 단위’로 정의되며, 하나 이상의 형태소로 이루어져 있다(Wikibaeggwa, 2018). 그런데, 용언의 경우에는 어간과 어미가 한 어절로서 한 단어를 이루지만, 체언과 관련된 독립성이 없는 조사 ‘이/가’, ‘은/는’, ‘을/를’이나 의존 명사 ‘줄’, ‘것’, ‘바’ 등을 단어로 인정하는 경우도 있다.

태소 단위의 형태소 벡터를 얻을 수 있게 된다. 본 연구에서는 이러한 형태소 벡터를 CNN 딥러닝 모델의 입력으로 사용한다.

2.3 CNN(Convolutional Neural Network)

다층퍼셉트론(multi-layer perceptron)이라고도 불리는 전형적인 피드포워드 뉴럴넷(feedforward neural network)은 인접한 층(layer)의 유닛(unit)이 서로 모두 연결된(fully-connected) 형태이다. CNN은 이러한 전결합층(fully-connected layer) 외에 인접층의 특정한 유닛만이 연결되는 합성곱층(convolution layer)과 풀링층(pooling layer)을 갖는다. 합성곱층에서는 가중치 행렬 형태로 표현된 이미지의 가중치 영역에 해당 이미지 보다 사이즈가 작은 가중치 행렬 형태의 필터(filter)를 순차적으로 이동시키면서 대응되는 원소끼리(element-wise) 곱한 후 합을 구하는 합성곱 연산이 일어난다. 그리고 풀링층에서는 합성곱으로 얻어진 값들에 대해 특정 사이즈 영역별로 순차적으로 최대값이나 평균값 등의 대표값을 추출하는 연산이 일어난다. 합성곱층과 풀링층은 서로 쌍을 이루며 여러번 반복되거나, 합성곱층만 여러번 반복된 후 풀링층이 오는 구조도 가능하다. 이러한 연산을 통해 입력된 이미지로부터 유용한 자질(feature)들이 계층적으로 추출되며, 추출된 자질들은 하나 이상의 전결합층(fully-connected layer)을 통해 입력 데이터를 타겟 클래스로 분류하는 작업에 활용된다.

문장을 단어 벡터의 결합으로 표현한 가중치 행렬에 CNN을 적용하면, 필터(filter)의 윈도우 사이즈에 따라 지역적인 n-그램(n-gram) 자질(feature)들을 추출할 수 있다. 그리고 이러한 자질들을 기반으로 다시 상위 레벨의 자질들을 추

출하는 것이 가능하다. $x_i \in R^k$ 가 임의의 문장의 i -번째 단어에 대한 k -차원의 단어 벡터라고 할 때, 길이 n 의 문장은 다음과 같이 표현된다.

$$x_{1:n} = x_1 \oplus x_2 \oplus \cdots \oplus x_n,$$

(\oplus : 연결(concatenation) 연산자)

그렇다면, $x_{i:i+h}$ 는 단어 $x_i, x_{i+1}, \dots, x_{i+h}$ 를 연결한 형태를 의미한다. 합성곱 연산은 새로운 자질(feature)을 추출하기 위해 h 단어 윈도우에 필터 $w \in R^{h \times k}$ 를 적용한다. 자질 c_i 는 단어 $x_{i:i+h-1}$ 로부터 다음 연산에 의해 생성된다.

$c_i = f(w \cdot x_{i:i+h-1} + b)$, (f : 하이퍼볼릭 탄젠트(hyperbolic tangent) 또는 렐루(relu)와 같은 비선형함수(non-linear function), $b \in R$: 바이어스(bias))

필터는 해당 문장의 모든 가능한 윈도우인 $\{x_{1:h}, x_{2:h+1}, \dots, x_{n-h+1:n}\}$ 에 대해 적용되어 피쳐맵(feature map)인 $c = [c_1, c_2, \dots, c_{n-h+1}]$ 이 생성된다($c \in R^{n-h+1}$). 풀링 연산은 이러한 피쳐맵에 대해 전역(global) 최대값을 구하거나($\hat{c} = \max\{c\}$) 특정 사이즈별로 지역(local) 대표값을 구한다. 윈도우 사이즈가 다른 여러 개의 필터를 적용하면 여러 특성의 자질들을 추출할 수 있고, 이렇게 추출된 자질들은 전결합층으로 연결되어 클래스 분류에 사용된다.

3. 데이터셋 및 실험 방법

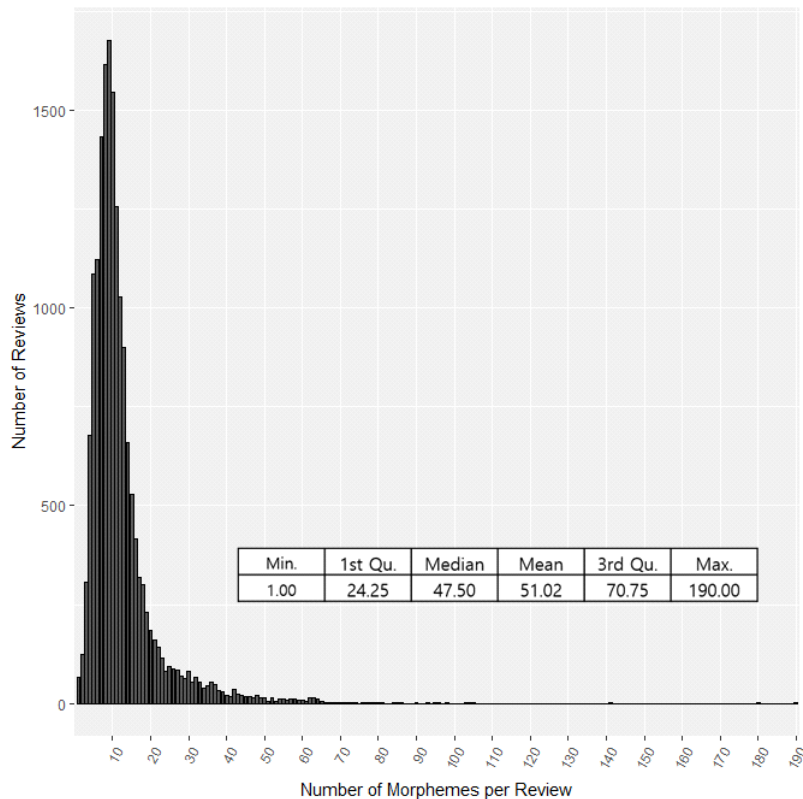
3.1 훈련 및 테스트 데이터셋

먼저, 네이버 쇼핑몰 화장품 138개 카테고리

에 해당하는 상품평 약 280만건을 크롤링하였다. 이 중 내용이 중복되거나 과도하게 띄어쓰기가 무시된 것, 한글/알파벳/숫자/마침표/느낌표/물음표 외의 부호를 제거한 후 글자수가 2개 미만인 것 등을 정제하여 2,053,883건의 상품평을 추출하였다. 이러한 상품평 중 별점 40 이하의 부정(negative) 상품평은 대략 1% 미만이었다. 그런데 신경망 성능 향상을 위해서는 부정과 긍정 라벨 데이터 수를 균등하게 맞추는 것이 좋다고 알려져 있기 때문에 부정 상품평 셋부터 선정하고 이와 같은 수의 긍정 상품평 셋을 도출하였다.

좀 더 구체적으로 설명하면, 부정 상품평 셋

선정을 위해 2,053,883건의 상품평 중 상품평 별점이 40점 이하인 것들을 일차적으로 추출하였다. 그리고 하나의 상품평 안에 극성이 상반되는 문장들이 섞여있을 경우 별점과의 일관성이 떨어지므로 문장 분리를 통해 하나의 문장으로 이루어진 상품평을 선별하였다. 문장 분리는 기본적으로 마침표(.), 느낌표(!), 물음표(?) 다음에 공백이 있으면 분리하였고, 상품평에서는 문법이 무시되는 경우가 많기 때문에 문장 부호가 없어도 되는 꼬꼬마(Kkoma) 형태소분석기의 문장 분리 기능을 추가적으로 적용하였다. 결과적으로 문장 분리가 되지 않은 11,915건 중에서 별점이



〈Figure 1〉 Morpheme Distribution of Data Set (Num_POS=45)

40점 이하임에도 불구하고 극명하게 긍정인 것들은 수작업을 통해 걸러내어 최종적으로 8,630건의 부정 상품평 셋을 선정하였다. 마찬가지로 긍정 상품평 셋은 별점이 80점 이상인 것들 중에서 문장 분리가 되지 않는 것들을 선별하고, 이 중 극성이 부정인 것들을 수작업으로 걸러내어 8,630건을 최종적으로 도출하였다. 전체 훈련 및 테스트 상품평 데이터셋 17,260건의 상품평 당 형태소 수의 분포는 <Figure 1>과 같다.

3.2 형태소 벡터 도출

본 연구에서는 감성분석 딥러닝 모델에 한국어 형태소 벡터를 어떻게 도출하여 입력하는 것이 효과적인가에 대해 검증하기 위해 다양한 형태소 벡터 도출 대안들을 선정하였다. 이들은 입력 텍스트, 문장 분리 및 맞춤법 교정 등의 텍스트 전처리, 포함되는 품사의 범위, 품사 태그 부착 여부 등의 기준에 따라 달라진다. 모든 형태소 벡터는 기본적으로 Mikolov et al.(2013b)에서 소개된 Word2Vec 모델 중 CBOW를 적용하여 도출하였다. 이를 위해 파이썬 Gensim 라이브러리를 사용하였으며, 컨텍스트 윈도우(window) 5, 벡터 차원 수(size) 300으로 설정하였다. CNN 모델을 통한 예비 분석 과정에서 벡터 100, 200 차원도 적용해 보았지만 300 차원의 경우가 대체적으로 다른 설정 이상의 성능을 보였다.

3.2.1 기본 형태소 벡터 셋(set) 구축

본 연구에서 구축한 형태소 벡터 셋(set)은 크게 다음 세 가지 기준에 의해 달라진다. 첫째, 데이터 소스가 문법 준수도가 높은 네이버 기사 또는 문법 준수도가 낮은 네이버 쇼핑 화장품 상품평 두 가지로 구분된다. 네이버 기사는 2017년

정치, 사회, 경제, 세계, 생활문화, 과학(IT 포함) 등 각 분야 약 10만개씩 전체 약 63만건을 수집하여 중복을 제거 한 후 얻은 517,167건을 사용하였다. 그리고 네이버 쇼핑 화장품 상품평은 138개 화장품 카테고리에 해당하는 약 280만건을 수집하여 중복을 제거하고 정제된 상품평 2,053,883건을 사용하였다.

둘째, 문장 분리만 하느냐, 문장 분리 후 맞춤법 및 띄어쓰기 교정까지 적용하고 사용하느냐 등 데이터 전처리의 정도에 따라 달라진다. Word2Vec 단어 벡터 모델은 기본적으로 한 문장 안에서 함께 출현하는 단어들에 대한 정보를 기반으로 하므로 모든 텍스트 데이터에 대해 문장 분리를 적용하였다. 뉴스 텍스트에서는 문법이 잘 지켜지기 때문에 문장 분리를 위해 마침표(.), 느낌표(!), 물음표(?) 다음에 공백 문자가 오는 조건을 사용하였다. 뉴스와 달리 상품평에서는 문법 준수도가 낮기 때문에 마침표, 느낌표, 물음표 및 공백 문자 조건 외에 이러한 문장 부호가 없어도 문장 분리가 되는 꼬꼬마(Kkma) 형태소 분석기의 문장 분리 기능을 적용하였다. 상품평의 경우에는 문법적인 오류 및 변형들을 교정하는 것이 실제로 도움이 되는지를 확인하기 위하여 문장 분리만 적용한 데이터 외에 파이썬 한글 맞춤법 및 띄어쓰기 교정 라이브러리인 파이-한스펠(Py-Hanspell)을 적용한 데이터도 도출하였다.

셋째, 문장 분리나 맞춤법 및 띄어쓰기 교정을 마친 데이터에 대해 형태소분석을 하고 형태소만 단어 벡터 모델에 입력하느냐, 형태소에 해당 품사 태그를 부착한 형태로 입력하느냐에 따라 달라진다. 형태소 끝에 품사 태그를 부착하는 방법은 2.2절에서 예로 든 “이 립스틱 은 컬러 도 예쁘 고 지속력 도 좋 다.”라는 문장 대신에, “이

=MM 립스틱=NNG 은=JX 켈러=NNG 도=JX 예
 =VA 고=EC 지속력=NNG 도=JX 좋=VA 다
 =EF .=SF”와 같은 형태의 문장을 CBOW 단어
 벡터 모델에 입력하는 것이다. 여기에서 품사 태
 그는 [Appendix]의 코모란(Komorán) 형태소분석
 기 품사 분류를 적용한 것이다. ‘=’는 대부분의
 텍스트에서 잘 나타나지 않는 기호문자를 삽입
 함으로써 한글과 영어가 혼합된 형태의 어휘와
 중복될 가능성을 최소화한 것이다. 이렇게 하면,
 예를 들어, 주격 조사와 동사에 해당하는 ‘가’의
 경우 ‘가=JKS’와 ‘가=VV’로 형태를 달리함으로
 써, 서로 다른 의미의 형태소로 구별하게 된다.
 형용사나 동사의 경우 어간의 끝에 ‘~다’를 붙여
 ‘가다’와 같이 형태를 바꿈으로써 구별하는 경우
 도 있는데 품사 태그를 붙이는 방법이 좀 더 안
 전해보인다. 품사 태그를 부착하지 않고 무조건
 ‘~다’를 붙이게 되면, 예를 들어, ‘~을 보다’의
 ‘보다’와 비교를 나타내는 부사격 조사 ‘보다’가
 같은 형태를 갖게 되기 때문에 구별이 어려워진

다. 이렇게 품사 태그를 부착하는 방법은 단어
 형태를 기준으로 하는 기존 단어 벡터 모델의 동
 음이의어 문제를 완화하기 위한 대안으로 본 연
 구에서 고안한 것이다. 전체 한국어 어휘에서 동
 음이의어가 차지하는 비율은 꽤 높으며 한국어
 어휘의 중의성은 영어에 비하여 훨씬 높다고 할
 수 있다(Kang, 2005).

결과적으로 앞에서 제시한 기준에 의해 도출
 한 여섯 가지 기본 형태소 벡터 셋(set)은 <Table
 1>과 같다.

3.2.2 공통 옵션: 품사 태그 셋, 최소 빈도수, 랜덤 초기화 범위

<Table 1>의 기본 형태소 벡터 셋(set)은 다시
 품사 태그의 고려 범위, 포함 형태소의 최소 빈
 도수 기준, 랜덤 초기화 범위에 따라 달라진다.
 첫째, 품사 태그의 고려 범위는 한국어의 의미
 및 감성 전달을 위해 기초적인 정보를 제공하는
 품사 셋으로부터 더욱 미묘한 의미 및 감성을 표

<Table 1> Basic Morpheme Vector Sets

Name	Data Source	Data Preprocessing	POS Tag Attachment
News	Naver News	▪ Sentence Splitting: space after (. ! ?)	no
News_POS	Naver News	▪ Sentence Splitting: space after (. ! ?)	yes
Cosm	Naver Cosmetics Reviews	▪ Sentence Splitting: space after (. ! ?) & Kkma	no
Cosm_POS	Naver Cosmetics Reviews	▪ Sentence Splitting: space after (. ! ?) & Kkma	yes
Cosm_Checked	Naver Cosmetics Reviews	▪ Sentence Splitting: space after (. ! ?) & Kkma ▪ Spelling & Spacing Check: Py_Hanspell	no
Cosm_Checked_POS	Naver Cosmetics Reviews	▪ Sentence Splitting: space after (. ! ?) & Kkma ▪ Spelling & Spacing Check: Py_Hanspell	yes

현하거나 감성과의 관련도가 미미해보이는 품사 셋으로 단계적으로 확장하여 [Appendix]와 같은 네 가지 품사 셋을 도출하였다. 본 연구에서는 형태소분석을 위해 형태소 분절 정도가 적당하다고 판단되는 코모란(Komoran) 형태소분석기를 사용하였다. 코모란(Komoran)의 미등록어에 해당하는 품사를 포함하여 전체 45가지 품사 각각에 대해 포함 여부에 의한 분류 정확도의 변화를 측정할 수도 있겠지만, 비용대비 효익이 크지 않을 것으로 생각되어 이와 같이 그룹 단위로 접근하였다.

둘째, 단어 벡터 모델에서 고려할 단어의 최소 빈도수, 즉, 본 연구에서는 고려할 형태소의 최소 빈도수 기준을 1, 3, 5 또는 1, 3, 5, 7, 10으로 다르게 설정하였다. 최소 빈도수 기준을 높일수록 고려 형태소 수는 감소하므로 도출되는 전체 형태소 벡터 수도 감소한다.

셋째, 형태소 벡터 도출 후 훈련 및 테스트 데이터에 포함된 형태소와 매핑시, 형태소 벡터 셋(set)에 존재하지 않는 형태소들은 균등분포(uniform distribution)에 의해 랜덤으로 초기화하는데, 이 때 균등분포의 범위를 다르게 설정할 수 있다. 균등분포의 범위는 도출된 기본 형태소 벡터들의 각 차원의 분산과 대략 비슷해지도록 $U(-0.5, 0.5)$ 로 설정하거나(Kim, 2014), 예비 분석을 통해 이와 비슷한 성능을 보이는 $U(-0.25, 0.25)$ 로 설정하였다. 최소 빈도수 기준이 높아질수록 랜덤 벡터로 초기화되는 형태소의 수는 증가한다.

3.3 CNN 모델 구조 및 훈련

감성분석을 위한 딥러닝 모델로는 Kim(2014)의 연구에서 소개된 모델 중 여러 데이터셋에 대

해 전체적으로 성능이 우수한 비정태적(non-static) CNN 모델을 참고하여, 예비분석을 통해 세부 구조를 조율하였다. 먼저, 임베딩계층(embedding layer)에서는 훈련 및 테스트 데이터에 포함된 형태소들을 이에 해당하는 형태소 벡터로 변환한다. 이 때 매핑되는 형태소 벡터가 존재하지 않으면 앞에서 언급한 바와 같이 랜덤으로 초기화된다. 초기 입력으로 설정된 단어 벡터들은 비정태적(non-static) CNN 훈련 과정을 통해 좀 더 나은 값으로 업데이트된다. 다음으로, 합성곱계층(convolutional layer)에서는 (2, 3, 4, 5) 사이즈의 필터를 각각 50개씩 팰리드(valid) 패딩(padding) 및 스트라이드(stride) 1 방식으로 합성곱하고 활성화함수 렐루(relu)를 적용한다. 그리고 풀링계층(pooling layer)에서는 전역(global) 맥스풀링(maxpooling) 대신에 사이즈(size) 2의 맥스풀링을 적용한다. 이 후 풀링계층의 여러 출력 값들을 평탄화(flatten) 및 연결(concatenate)하여 50개 유닛(unit)의 히든(hidden) 전결합층(fully-connected layer)으로 보내고 활성화함수 렐루(relu)를 통과시킨 후에 다시 하나의 유닛을 갖는 전결합층으로 연결하여 활성화함수 시그모이드(sigmoid)로 마무리한다.

오버피팅(overfitting)을 완화하기 위해 임베딩 계층 및 연결(concatenate) 후에는 각각 (0.5, 0.8)의 드롭아웃(dropout) 확률을 적용하였다. 손실(loss) 함수는 바이너리-크로스엔트로피(binary-crossentropy) 함수를, 옵티마이저(optimizer)로는 아담(adam)을 사용하였다. 모델 하이퍼패러미터(hyperparameter) 및 훈련 관련 패러미터(parameter)를 정리하면 <Table 2>와 같고, 전체적인 모델 구현은 파이썬(Python) 및 케라스(Keras) 라이브러리를 활용하였다.

(Table 2) Non-static CNN Model Details

Category	Item	Value
Model Hyperparameters	Embedding Dimension	300
	Filter Sizes	(2, 3, 4, 5)
	Number of Filters for Each Filter Size	50
	Dropout Probabilities	(0.5, 0.8)
	Hidden Dimension	50
Training Parameters	Batch Size	64
	Number of Epochs	20
Word2Vec Parameters	MinimumWord Count(min_count)	(1, 3, 5)
	Context(window)	5
	Vector Dimension(size)	300

3.4 분류 정확도 비교 방법

분류 정확도(accuracy)는 긍정과 부정 각각 8,630건으로 이루어진 데이터셋에 대해 5-fold CV(Cross Validation) 기법을 적용하여 각 라운드의 20 에폭(epoch) 중 최대값들의 평균값으로 계산하였다. 좀 더 구체적으로 설명하면, 먼저 긍정과 부정 데이터셋을 각각 랜덤하게 5그룹으로 나누고 4그룹은 훈련 데이터, 1그룹은 테스트³⁾ 데이터로 사용하여 20 에폭 중 테스트 데이터에 대한 최대 정확도를 구하였다. 그리고 이 과정을 테스트 데이터 그룹을 바꿔가면서 5번 반복한 후 5개의 최대 정확도 값에 대한 평균값을 구하였다.

단어 벡터 모델의 입력으로 사용되는 상품평 데이터와 훈련 및 테스트 용도로 사용되는 상품평 데이터에 대해 같은 정도의 텍스트 전처리를 적용하였다. 즉, 두 가지 데이터 모두 문장 분리

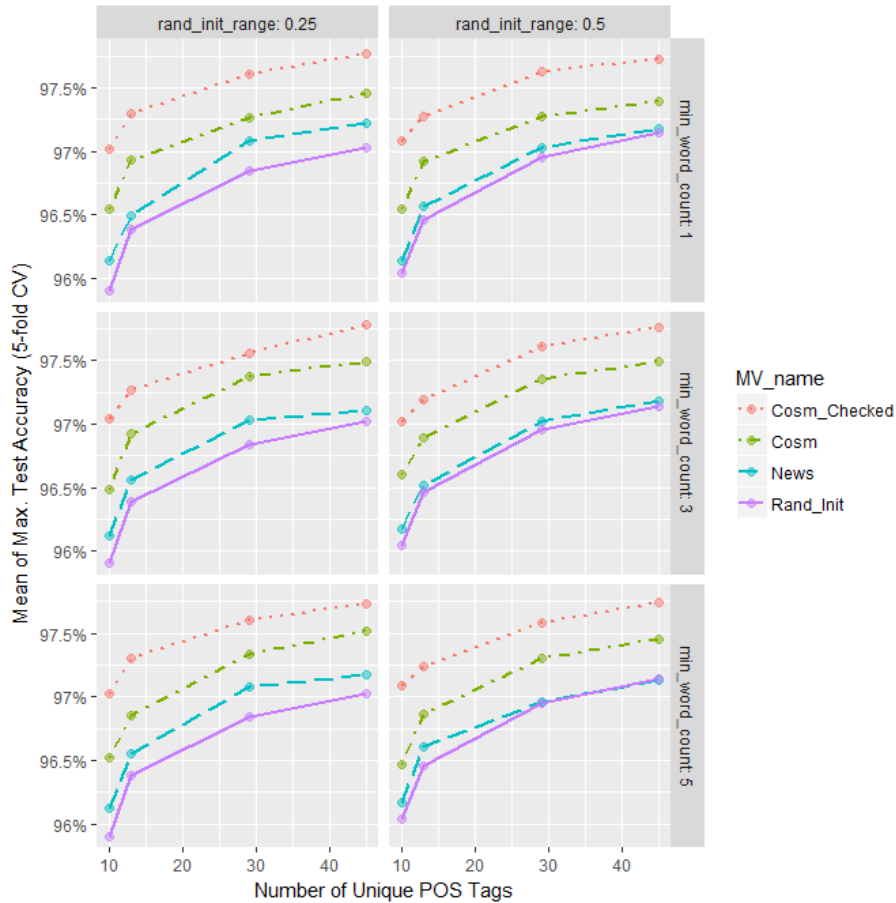
만 하든지, 모두 문장 분리 후 띄어쓰기 및 맞춤법 교정까지 적용하든지, 또는 모두 품사 태그를 부착하였다. 품사의 범위도 두 가지 데이터셋에 대해 똑같이 설정하였다. 즉, [Appendix]의 네 가지 품사 셋별로 기본적인 형태소 벡터를 도출하였고, 훈련 및 테스트 데이터셋에 대해서도 네 가지 품사 셋별로 형태소 표현을 도출하여 품사 범위를 통일하였다. 뉴스 데이터로 형태소 벡터를 도출하는 경우에는 훈련 및 테스트 데이터로 문장 분리한 상품평을 사용하였다.

4. 실험 결과

4.1 형태소 벡터 입력 데이터의 종류

<Figure 2>와 <Figure 3>에서 관찰되는 입력 데이터 소스 관련 결과를 요약해보면 다음과 같

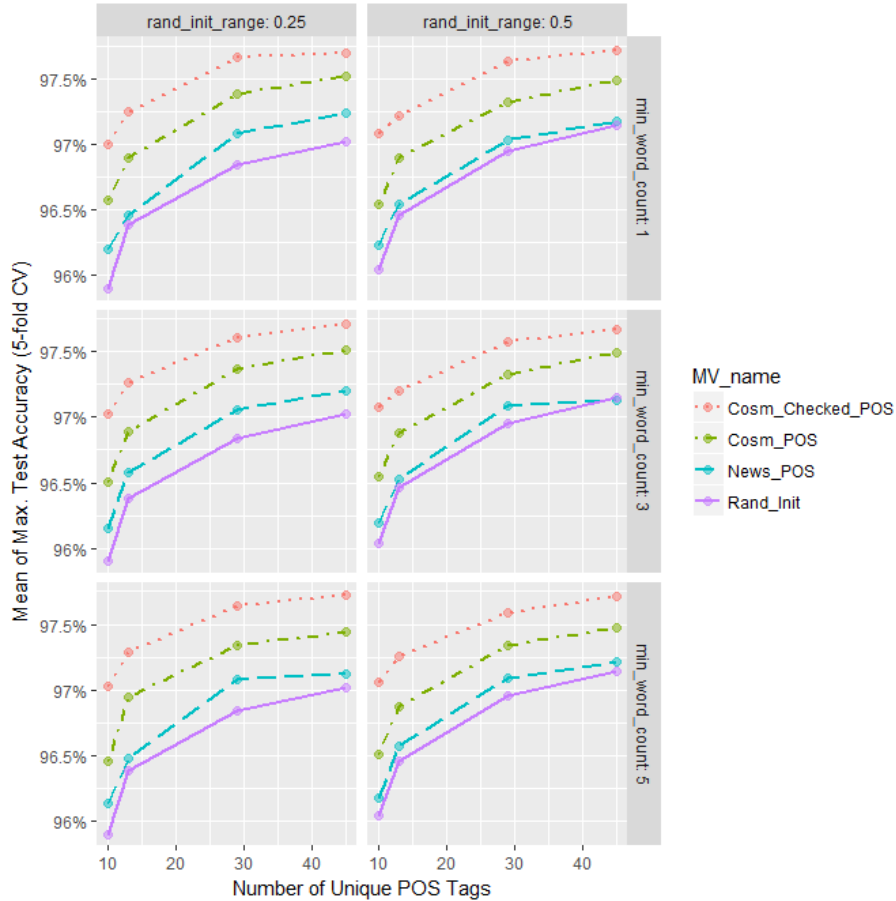
3) 또는 검증(validation) 데이터로도 생각할 수 있다.



〈Figure 2〉 Test Accuracy Distributions on Focal Parameter Combinations (Word2Vec without POS)

다. 첫째, 문법 준수도가 높은 기사로 도출한 형태소 벡터 보다, 다양한 문법적 오류가 포함되어 있어도 훈련 데이터와 같은 도메인의 상품평 데이터로 도출한 형태소 벡터를 입력 데이터로 사용할 때 더욱 높은 정확도를 달성하는 것으로 보인다. 따라서, 분석 대상 도메인 내에서 대규모 데이터를 획득할 수 있다면, 해당 데이터를 활용하는 것이 범용 형태소 벡터를 사용하는 것 보다 효과적인 대안으로 보인다.

둘째, 분석 대상과 다른 도메인의 입력 데이터를 사용하더라도 미리 도출한 형태소 벡터를 사용하는 것이 훈련 및 테스트 데이터에 포함된 형태소에 대한 벡터를 모두 랜덤으로 초기화하는 것보다 나은 대안으로 보인다. 전체 랜덤 벡터에 비해 네이버 뉴스 텍스트로 도출한 형태소 벡터가 약간 더 나은 결과를 제공하고 있다. 이것은 도메인이 다르더라도 비지도적(unsupervised) 방법으로 미리 훈련된 단어 벡터를 여러 영어 텍스



〈Figure 3〉 Test Accuracy Distributions on Focal Parameter Combinations (Word2Vec with POS)

트 연구에서 활용하고 있는 추세와 상통하는 결과라고 할 수 있다(Collobert et al., 2011; Kim et al., 2014).

셋째, 전체 랜덤 벡터로 초기화하는 대안도 상대적으로 크게 나쁘지 않다. 이것은 본 연구에서 랜덤 벡터 초기화 범위를 적절하게 설정하였고, 비정태적 CNN 모델이 초기 입력 형태소 벡터를 훈련 과정을 통해 효과적으로 업데이트하기 때문인 것으로 추측된다.

4.2 형태소 벡터 도출 방법

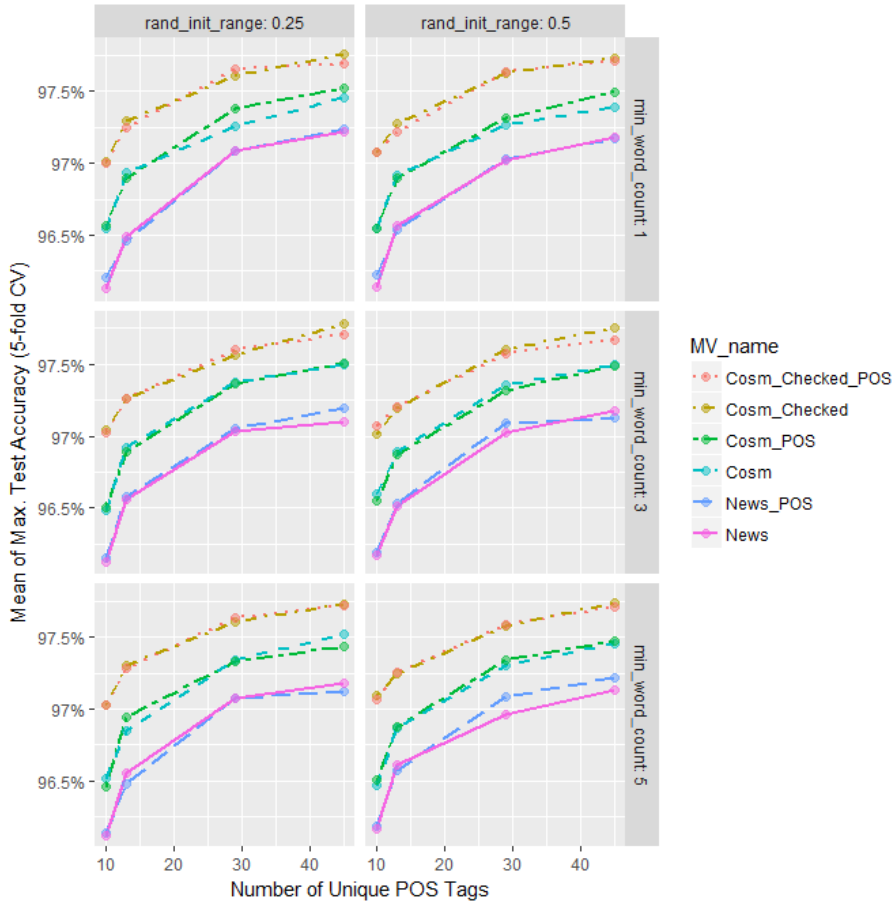
4.2.1 입력 데이터의 전처리 및 품사 태그 부착

형태소 벡터 도출을 위한 입력 데이터에 대한 문장 분리와 맞춤법 및 띄어쓰기 교정, 품사 태그 부착 등과 관련된 결과를 요약하면 다음과 같다. 첫째, 〈Figure 2〉와 〈Figure 3〉를 보면, 상품평을 문장 분리만 하여 사용하는 것 보다 맞춤법 및 띄어쓰기 교정까지 적용한 후 형태소 벡터를

도출하는 것이 분류 정확도 향상을 위해 유익하다. 이것은 맞춤법 및 띄어쓰기를 교정하면 다양한 변형을 하나의 표준형으로 통일해주기 때문에 좀 더 품질이 좋은 형태소 벡터를 얻을 것으로 예상한 것과 일치하는 결과라고 할 수 있다. 그런데 이 결과가 맞춤법 및 띄어쓰기 교정의 품질이 완벽하다는 것을 의미하지는 않는다. 본 연구에서 사용한 파이-한스펠(Py-Hanspell)은 화장품 도메인에 특화된 라이브러리가 아니기 때문에 교정 내용에 약간의 오류가 포함되어 있다.

예를 들어, ‘가격대비’를 ‘자격대비’로 교정하는 것을 발견하였다. 이러한 오류에도 불구하고 더 나은 결과를 제공하는 이유는 형태소 벡터 도출 텍스트와 분석 대상 텍스트를 같은 방식으로 교정했으므로 감성분석 측면에서는 부정적인 영향이 거의 없기 때문인 것으로 판단된다.

둘째, <Figure 4>를 보면, 각 형태소에 품사 태그를 부착하여 단어 벡터 모델에 입력하는 방법은 예상한 만큼 효과가 뚜렷하지는 않은 것으로 보인다. 품사 범위나 랜덤 초기화 범위 및 최소



<Figure 4> Test Accuracy Comparison Between With-POS and Without-POS Cases

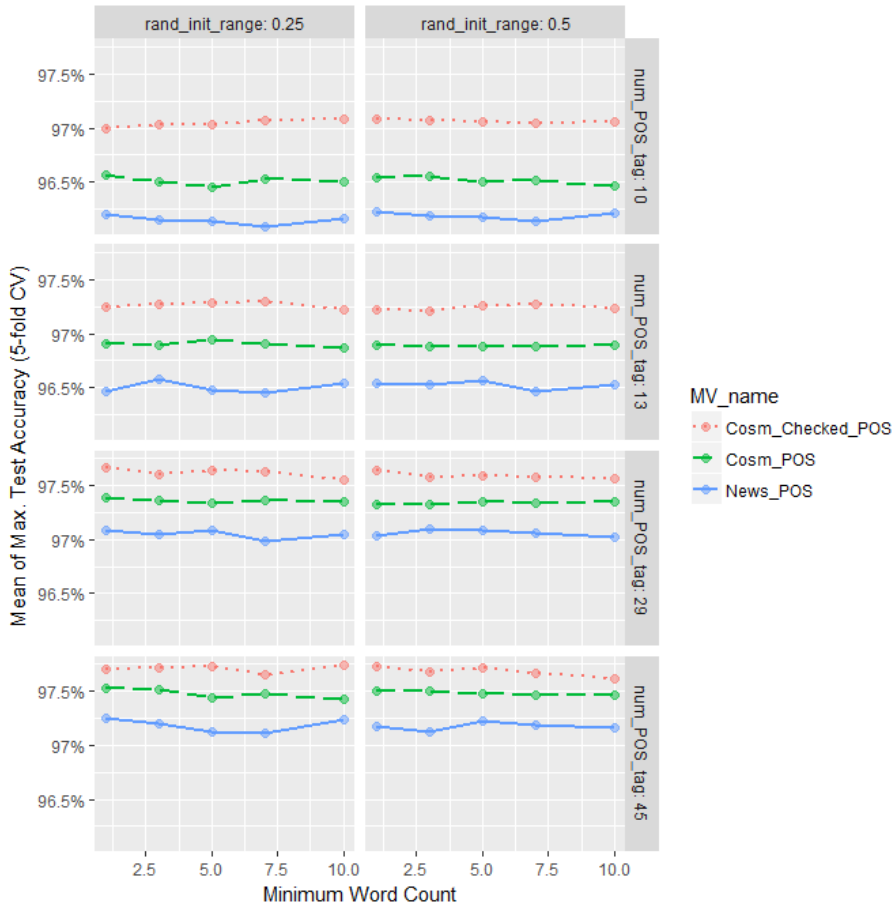
빈도수 값에 따라 품사 태그를 부착하는 경우의 정확도가 약간 높거나 비슷한 경우도 있지만, 약간 낮은 경우도 존재한다. 이것은 기본적으로 한국어의 동음이의어 비중은 높지만 화장품 상품평 텍스트에서 언급되는 형태소들 중 품사가 다른 동음이의어의 비중은 높지 않기 때문인 것으로 추측된다. 그런데 품사 태그를 부착했을 때의 정확도가 더 낮다고도 할 수 없으므로, 동음이의어 비중을 예측하기 어려운 다른 여러 도메인이나 개체(entity) 인식 및 속성 추출(aspect extraction) 등에 형태소 벡터를 적용할 때에는 품사 태그를 부착하여 도출하는 방법이 더욱 안전할 것으로 판단된다.

4.2.2 품사 셋, 최소 빈도수, 랜덤 초기화 범위 설정

형태소 벡터 도출 시 고려할 형태소의 품사 셋, 최소 빈도수 및 랜덤 벡터 초기화 범위와 관련된 결과를 요약하면 다음과 같다. 첫째, <Figure 2>, <Figure 3>, <Figure 4>에서와 같이, 고려하는 품사의 범위가 넓을수록 정확도가 높아지는 것으로 보인다. 특히, 코모란 형태소분석기의 경우 상품평과 뉴스 공통적으로 ‘분석불능 범주(NA)’에 속하는 것들의 비율이 상대적으로 높게 나타나는데 이들도 포함하는 편이 나온 것으로 보인다. 분석불능 범주에는 ‘론칭’, ‘홈커밍’, ‘크로노그래프’ 등의 미등록 명사 외에 ‘아 쉽다...πππ’, ‘ㅋㅋ’, ‘ㅎㅎ’, ‘ππππ’, ‘찹찹했던’, ‘이빠요’ 등의 미등록 감성어들도 많이 포함되어 있다. 그리고 감성 표현과 비교적 관련도가 낮아 보이는 네 번째 품사 그룹인 ‘POS Tag Set 4’에 속하는 품사들도 전체 품사에서 차지하는 비율은 높지 않지만 정확도 측면에서는 유효한 것으로 보인다.

둘째, <Figure 5>에서와 같이, 형태소 벡터 도출 시 고려할 형태소에 대한 최소 빈도수 기준을 1, 3, 5, 7, 10으로 설정할 때 이 차이가 분류 정확도에 미치는 영향은 미미한 것으로 보인다. 이러한 현상은 빈도수 기준을 높게 설정할수록 형태소 벡터의 품질은 나아지더라도, 훈련 및 테스트 데이터에 포함된 형태소와의 매칭 비율이 낮아지기 때문에 효과가 서로 상쇄되기 때문인 것으로 생각된다. <Table 3>에서와 같이, 훈련 및 테스트 데이터셋의 형태소 수는 띄어쓰기 및 맞춤법 교정을 한 경우에는 5,945개, 문장 분리만 한 경우에는 6,255개이다. 이중 미리 훈련된 형태소 벡터와의 매칭 비율은 상품평으로 형태소 벡터를 도출한 경우 최소 빈도수 기준을 10에서 1까지 낮춤에 따라 89%에서 100%까지, 뉴스를 사용한 경우 77%에서 85%까지 상승하였다. 매칭이 되지 않는 경우에는 랜덤으로 설정하더라도, 비율이 그리 높지 않고 훈련 과정을 통해 업데이트되기 때문에 부정적인 영향이 완화되는 면도 있는 것으로 보인다.

셋째, 랜덤 초기화 범위는 고려하는 품사 범위나 최소 빈도수 등의 조합에 따라 $U(-0.25, 0.25)$ 일 때의 정확도가 $U(-0.5, 0.5)$ 일 때 보다 약간 높은 경우와 낮은 경우가 혼합되어 나타난다. 이것은 랜덤 초기화 범위는 도출된 형태소 벡터나 딥러닝 모델의 특성에 따라 최적 범위가 달라질 것으로 예상하여, 초기화 범위를 두 가지 모두 적정값으로 설정했기 때문일 것이다. 형태소 벡터를 모두 랜덤으로 초기화하는 베이스라인 모델의 경우에는 $U(-0.5, 0.5)$ 일 때 약 0.1% 정도 높다.



(Figure 5) Test Accuracy Distributions Focused on Minimum Word Frequency

4.3 딥러닝 감성분석 결과의 정확도 수준

맞춤법 및 띄어쓰기 교정을 한 상품평 데이터를 사용하고 모든 품사를 고려하여 도출한 형태소 벡터를 사용할 때 본 연구의 비정태적 CNN 모델은 최대 97.8%의 정확도를 보였다. 형태소 벡터 도출 시 10가지 품사만 고려하거나 모두 랜덤으로 초기화하는 경우, 또는 고려하는 형태소의 빈도수를 1이나 10으로 설정하는 경우에도 95% 이상의 정확도를 제공하고 있다. 이러한 결

과는 비정태적 CNN 모델 디자인의 적합성뿐만 아니라 문장 분리를 통해 별점의 의미가 분명한 상품평들로 구성된 훈련 및 테스트 데이터셋의 특성에도 기인할 것이다. 본 연구의 훈련 및 테스트 데이터셋 구성 방법은 별점과 상품평 극성과의 일관성을 유지하기 위한 것이며, 분석 대상 상품평 텍스트도 문장 분리를 한 후 감성분석을 하고 각 문장의 감성을 전체적으로 통합하는 방식으로 활용될 수 있다.

〈Table 3〉 Morpheme Vectors & Their Vocabulary Matching Status

Name	Min_Wd_Ct	M2V_Size	Vocab_Size	M2V_Matched	M2V_Matched_Ratio
Cosm_Checked_POS	1	79,907	5,945	5,942	99.95
	3	28,423	5,945	5,644	94.94
	5	21,880	5,945	5,538	93.15
	7	18,714	5,945	5,454	91.74
	10	16,063	5,945	5,371	90.34
Cosm_POS	1	101,289	6,255	6,252	99.95
	3	31,504	6,255	5,869	93.83
	5	23,679	6,255	5,756	92.02
	7	20,022	6,255	5,662	90.52
	10	17,005	6,255	5,573	89.10
News_POS	1	146,965	6,255	5,324	85.12
	3	89,734	6,255	5,153	82.38
	5	72,824	6,255	5,023	80.30
	7	63,495	6,255	4,933	78.86
	10	54,791	6,255	4,818	77.03

*(Num_POS=45, Max_Sentence_Morph=190)

5. 결론 및 향후 연구 방향

본 연구는 한국어 감성분석에 딥러닝 모델을 적용하기 위한 기초 연구로서, 예비 분석을 통해 조율한 비정태적 CNN 모델을 기반으로 다양한 형태소 벡터 도출 대안의 효과성을 검증하였다. 그리고 전형적인 교착어이면서 동음이의어 비중이 높은 한국어인 경우에도 만족할만한 정확도를 얻을 수 있는 가능성을 확인하였다. 최근에 활발하게 소개되고 있는 다양한 RNN 모델에서도 초기화한 단어 벡터를 훈련 과정에서 업데이트하는 비정태적 방식을 사용하는 경우가 많으므로, 본 연구의 비정태적 CNN 모델에 의한 결

과를 적용 가능할 것으로 예상된다.

본 연구의 결과를 반영하여 한국어 상품평에 대한 딥러닝 감성분석 시스템을 구축한다면, 형태소 벡터 도출 전략은 대략 다음과 같을 것이다. 먼저, 분석 대상 도메인에 속하는 풍부한 텍스트를 확보한다. 다음으로 해당 도메인에 특화된 맞춤법 및 띄어쓰기 교정 프로그램이 있다면 좋고 없다면 범용의 교정 프로그램으로 텍스트 정제 과정을 거친다. 그리고 해당 텍스트의 문법 준수도에 적합한 문장분리 방법을 적용한다. 문장별 형태소 분석을 한 후 모든 품사의 형태소에 대한 품사 태그를 부착하여 형태소 벡터를 도출한다. 도출된 형태소 벡터를 딥러닝 모델에 입력

하여 문장별 극성을 산출한 후 합산하여 문서에 대한 최종 극성을 계산한다. 분석 대상 도메인에 속하는 대량의 텍스트를 구하기 어렵다면 앞서와 비슷한 방법으로 동음이의어 문제를 고려하여 도출한 범용의 형태소 벡터를 사용하는 대안이나 랜덤으로 초기화하는 방법을 순차적으로 고려한다.

연구 과정에서 느낀 본 연구의 한계점 및 향후 연구 주제는 다음과 같다. 첫째, 한국어의 경우 동음이의어 비중이 상대적으로 높기 때문에 고품질의 형태소 벡터를 도출하기 위해서는 이에 대한 추가적인 연구가 요구된다. 특정 도메인 내에서는 동음이의어 비율이 낮아 감성분석 정확도에 큰 영향을 미치지 않는다해도 형태소 벡터 자체를 사용하여 속성(aspect)이나 개체(entity)를 추출하는 경우에는 좀 더 미세한 조정이 필요할 수 있다. 특히 여러 도메인을 아우르는 범용 한국어 형태소 벡터를 구축하는 경우에는 본 연구에서 제안한 방법과 같이 품사를 구별해도 해결되지 않는 부분이 있기 때문에, 품사가 같은 동음이의어에 대한 연구가 필요한 것 같다. 이를 위해서는 일차적으로 전체 형태소 벡터를 도출하고 동음이의어들을 컨텍스트 형태소들을 기준으로 클러스터링하여 형태를 다르게 조정한 후 최종 형태소 벡터를 도출하는 식의 접근이 이루어져야 할 것으로 보인다.

둘째, 한국어 텍스트 극성과 관련이 있는 다양한 패턴들을 정리하고 이러한 패턴들을 충분히 포함하는 훈련 및 테스트 데이터셋을 효율적으로 구축하기 위한 연구가 필요하다. 기존의 사전이나 규칙 기반 감성분석에 비해 딥러닝을 활용한 방법은 힘겨운 수작업을 상당 부분 덜어주지만, 훈련 및 테스트 데이터셋의 구축 문제는 여전히 과제로 남아 있다. 상품평의 경우 부정 극

성을 가진 상품평이 훨씬 적을뿐더러 별점이 도움은 되지만 별점과의 일관성이 떨어지는 상품평도 다수 존재함을 발견하였다. 특히, 긍정이나 부정으로 분류하는 수준을 넘어 극성의 강도를 더욱 세분화하여 분류하는 연구로 확장해나가기 위해서는 이에 맞는 훈련 및 테스트 데이터셋의 구축이 요구될 것이다.

셋째, 본 연구에서 제안한 방법을 다른 언어 및 도메인의 데이터셋에 대해 적용해보거나 본 연구에서 살펴보지 못한 딥러닝 모델 및 다양한 패러미터 조합에 대해 지속적으로 검증해보는 과정이 필요하다. 본 연구에서는 여러 형태소 도출 대안들을 비교하느라 뉴스와 화장품 상품평을 사용했지만 특성이 다른 언어나 도메인에 대해서도 확인해볼 필요가 있다. 그리고 본 연구에서는 충분하다고 판단되는 정도의 예비 분석을 통해 대안들을 선정하고 모델 구조를 조율하였지만, 딥러닝 모델 관련 변수의 종류가 많고 변화 범위도 넓기 때문에 더 나은 변수 조합을 발견할 가능성도 배제할 수 없다.

끝으로, 본 연구의 결과가 형태소 벡터를 입력으로 사용하는 다양한 딥러닝 모델을 활용한 한국어 감성분석 연구에 실질적인 도움이 되길 바란다.

참고문헌(References)

- An, J.-y., J.-w. Bae, N.-g. Han, and M. Song, "A Study of 'Emotion Trigger' by Text Mining Techniques," *Journal of Intelligence and Information Systems*, Vol.21, No.2(2015), 69~92.
- Chen, P., Z. Sun, L. Bing, and W. Yang,

- “Recurrent Attention Network on Memory for Aspect Sentiment Analysis,” *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, (2017).
- Cho, K., van M. Bart, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation,” arXiv preprint arXiv:1406.1078, (2014).
- Collobert, R., J. Weston, L. Bottou, M. Karlen, K. Kavukcuglu, and P. Kuksa, “Natural Language Processing (Almost) from Scratch,” *Journal of Machine Learning Research*, Vol.12, (2011), 2493~2537.
- Cui, M.-n., Y.-s. Jin, and O.-b. Kwon, “A Method of Analyzing Sentiment Polarity of Multilingual Social Media: A Case of Korean-Chinese Languages,” *Journal of Intelligence and Information Systems*, Vol.22, No.3(2016), 91~111.
- Dhanush, D. and A. K. Thakur, “Aspect-based Sentiment Summarization with Deep Neural Networks,” *International Journal of Engineering Research and Technology*, Vol.5, No.5(2016), 371~375.
- Guan, Z., L. Chen, W. Zhao, Y. Zheng, S. Tan, and D. Cai, “Weakly-supervised Deep Learning for Customer Review Sentiment Classification,” *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16)*, (2016), 3719~3725.
- Hochreiter, S. and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, Vol.9, No.8(1997), 1735~1780.
- Hong, T.-h., T.-w. Lee, and J.-g. Li, “Development of Sentiment Analysis Model for the Hot Topic Detection of Online Stock Forums,” *Journal of Intelligence and Information Systems*, Vol.22, No.1(2016), 187~204.
- Jebbara, S. and P. Cimiano, “Aspect-Based Relational Sentiment Analysis Using a Stacked Neural Network Architecture,” arXiv preprint arXiv:1709.06309, (2017).
- Jebbara, S. and P. Cimiano, “Aspect-Based Sentiment Analysis Using a Two-Step Neural Network Architecture,” arXiv preprint arXiv:1709.06311, (2017).
- Jeong, J. S., D. S. Kim, and J. W. Kim, “Influence Analysis of Internet Buzz to Corporate Performance: Individual Stock Price Prediction Using Sentiment Analysis of Online News,” *Journal of Intelligence and Information Systems*, Vol.21, No.4(2015), 37~51.
- Joulin, A., E. Grave, P. Bojanowski, and T. Mikolov, “Bag of Tricks for Efficient Text Classification,” arXiv preprint arXiv, (2016).
- Kang, B.-m., “Aspects of the Use of Homonyms,” *Language Research*, Vol.41, No.1(2005), 1~29.
- Kim, B. G., “A Study on the Homonym in Korean,” *Our Language Study*, Vol., No.45 (2013), 181~200.
- Kim, J. H., *Understanding Linguistics*, Yeog-Lag, 2004.
- Kim, S. I., D. S. Kim, and J. W. Kim, “Public Sentiment Analysis of Korean Top-10 Companies: Big Data Approach Using Multi-Categorical Sentiment Lexicon,” *Journal of Intelligence and Information Systems*, Vol.22, No.3(2016), 45~69.

- Kim, Y., "Convolutional Neural Networks for Sentence Classification," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing(EMNLP)*, (2014), 1746~1751.
- Liu, B., *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, The Cambridge University Press, 2015.
- Ma, D., S. Li, X. Zhang, and H. Wang, "Interactive Attention Networks for Aspect-Level Sentiment Classification," *Proceedings of the Internal Joint Conference on Artificial Intelligence(IJCAI 2017)*, (2017).
- Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," *Advances in Neural Information Processing Systems*, (2013a), 3111~3119.
- Mikolov, T., K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv preprint arXiv:1301.3781, (2013b).
- Nam, G.-I. and E.-g. Jo., *Korean Text Sentiment Analysis*, Communication-Books, 2017.
- Ouyang, X., P. Zhou, C. H. Li, and L. Liu, "Sentiment Analysis Using Convolutional Neural Network," *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, (2015), 2359~2364.
- Pang, B. and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, Vol.2, No.1-2(2008), 11~35.
- Pennington, J., R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation," *Proceedings of the Conference on Empirical Methods on Natural Language Processing(EMNLP 2014)*, (2014).
- Ruder, S., P. Ghaffari, and J. G. Breslin, "A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis," *Proceedings of the Conference on Empirical Methods on Natural Language Processing(EMNLP 2016)*, (2016a).
- Ruder, S., P. Ghaffari, and J. G. Breslin, "INSIGHT-1 at SemEval-2016 Task 5: Deep Learning for Multilingual Aspect-based Sentiment Analysis," *Proceedings of SemEval*, (2016b), 330~336.
- Schuster, M. and K. K. Paliwal, "Bidirectional Recurrent Neural Networks," *IEEE Transactions on Signal Processing*, Vol.45, No.11(1997), 2673~2681.
- Shin, B., T. Lee, and J. D. Choi, "Lexicon Integrated CNN Models with Attention for Sentiment Analysis," *Proceedings of the EMNLP Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis of WASSA'17*, (2017).
- Shirani-Mehr, H., "Applications of Deep Learning to Sentiment Analysis of Movie Reviews," *Technical Report*, Stanford University, (2015).
- Socher, R., C. Lin, A. Y. Ng, and C. D. Manning, "Parsing Natural Scenes and Natural Language with Recursive Neural Networks," *Proceedings of the 28th International Conference on Machine Learning(ICML-11)*, (2011).

- Stevenson, M., *Word Sense Disambiguation: The Case for Combinations of Knowledge Sources*, Stanford: CSLI Publications, 2003.
- Tang, D., B. Qin, X. Feng, and T. Liu, “Effective LSTMs for target-dependent sentiment classification,” *Proceedings of the International Conference on Computational Linguistics (COLING 2016)*, (2016).
- Tay, Y., L. A. Tuan, and S. C. Hui, “Dyadic Memory Networks for Aspect-based Sentiment Analysis,” *Proceedings of the International Conference on Information and Knowledge Management(CIKM 2017)*, (2017).
- Wang, B. and M. Liu, “Deep Learning for Aspect-Based Sentiment Analysis,” *Stanford University Report*, 2015. Available at <https://cs224d.stanford.edu/reports/WangBo> (Downloaded 19 May, 2018).
- Wang, Y., M. Huang, X. Zhu, and L. Zhao, “Attention-based LSTM for Aspect-level Sentiment Classification,” *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, (2016), 606~615.
- Wiebe, J., T. Wilson, R. Bruce, M. Bell, and M. Martin, “Learning Subjective Language,” *Computational Linguistics*, Vol.30, No.3 (2004), 277~308.
- Wikibaeggwa, Available at <https://ko.wikipedia.org/wiki/%EB%82%B1%EB%A7%90/> (Downloaded 20 March, 2018).
- Yang, Z., D. Yang, C. Dyer, X. He, A. J. Smola, and E. H. Hovy, “Hierarchical Attention Networks for Document Classification,” *HLT-NAACL(North American Chapter of the Association for Computational Linguistics: Human Language Technologies)*, (2016).
- Zhang, L., S. Wang, and B. Liu, “Deep Learning for Sentiment Analysis: A Survey,” arXiv preprint arXiv:1801.07883, (2018).

[Appendix] POS Tags and 4 POS Tag Sets from Komoran Morpheme Analyzer

N	Tag	Description(Eng.)	Description(Kor.)	POS Tag Set			
1	NNG	general noun	일반 명사	POS Tag Set 1: 10 POS	POS Tag Set 2: 13 POS	POS Tag Set 3: 19 POS	POS Tag Set 4: 45 POS
2	NNP	proper noun	고유 명사				
3	VV	verb	동사				
4	VA	adjective	형용사				
5	VX	secondary verb	보조 용언				
6	XR	radix	어근				
7	VCP	positive designator	긍정 지정사				
8	VCN	negative designator	부정 지정사				
9	MAG	general adverb	일반 부사				
10	IC	interjection	감탄사				
11	JKS	subjective postposition	주격 조사				
12	JKO	object postposition	목적격 조사				
13	NA	incomprehensible category	분석불능 범주				
14	EC	connective ending	연결 어미				
15	EP	prefinal ending	선어말 어미				
16	EF	final ending	종결 어미				
17	SF	period, question mark, exclamation mark	마침표, 물음표, 느낌표				
18	SE	ellipsis	줄임표				
19	JKC	supplementary postposition	보격 조사				
20	JKG	genitive postposition	관형격 조사				
21	JKB	adverbial postposition	부사격 조사				
22	JKV	vocative postposition	호격 조사				
23	JKQ	quotation postposition	인용격 조사				
24	JC	connective postposition	접속 조사				
25	JX	auxiliary particle	보조사				
26	NF	noun-assumed category	명사추정 범주				
27	NV	predicate-assumed category	용언추정 범주				
28	MAJ	connective adverb	접속 부사				
29	MM	prenoun	관형사				
30	NNB	dependent noun	의존 명사				
31	ETN	noun transformation ending	명사형 전성 어미				
32	ETM	prenoun transformation ending	관형형 전성 어미				
33	XPN	substantive prefix	체언 접두사				
34	XSN	noun derivational suffix	명사파생 접미사				
35	XSV	verb derivational suffix	동사파생 접미사				
36	XSA	adjective derivational suffix	형용사파생 접미사				
37	SS	quotes, parentheses, strings	따옴표, 괄호표, 줄표				
38	SL	foreign language	외국어				
39	NR	numeral	수사				
40	NP	pronoun	대명사				
41	SP	comma, centered dot, colon, slash	쉼표, 가운데점, 콜론, 빗금				
42	SN	number	숫자				
43	SO	dash(wave, hidden, missing)	불임표(물결, 숨김, 빠짐)				
44	SW	other symbols(mathematical symbols, monetary symbols)	기타기호(논리수학기호, 화폐기호)				
45	SH	chinese character	한자				

Abstract

Sentiment Analysis of Korean Reviews Using CNN: Focusing on Morpheme Embedding

Hyun-jung Park* · Min-chae Song** · Kyung-shik Shin***

With the increasing importance of sentiment analysis to grasp the needs of customers and the public, various types of deep learning models have been actively applied to English texts. In the sentiment analysis of English texts by deep learning, natural language sentences included in training and test datasets are usually converted into sequences of word vectors before being entered into the deep learning models. In this case, word vectors generally refer to vector representations of words obtained through splitting a sentence by space characters. There are several ways to derive word vectors, one of which is Word2Vec used for producing the 300 dimensional Google word vectors from about 100 billion words of Google News data. They have been widely used in the studies of sentiment analysis of reviews from various fields such as restaurants, movies, laptops, cameras, etc.

Unlike English, morpheme plays an essential role in sentiment analysis and sentence structure analysis in Korean, which is a typical agglutinative language with developed postpositions and endings. A morpheme can be defined as the smallest meaningful unit of a language, and a word consists of one or more morphemes. For example, for a word '예쁘고', the morphemes are '예쁘(= adjective)' and '고(=connective ending)'. Reflecting the significance of Korean morphemes, it seems reasonable to adopt the morphemes as a basic unit in Korean sentiment analysis. Therefore, in this study, we use 'morpheme vector' as an input to a deep learning model rather than 'word vector' which is mainly used in English text. The morpheme vector refers to a vector representation for the morpheme and can be derived by applying an existent word vector derivation mechanism to the sentences divided into constituent morphemes.

By the way, here come some questions as follows. What is the desirable range of POS(Part-Of-Speech) tags when deriving morpheme vectors for improving the classification accuracy of a deep learning model? Is it proper to apply a typical word vector model which primarily relies on the form of words to Korean with a high homonym ratio? Will the text preprocessing such as correcting spelling

* Management Research Center, Ewha Womans University

** Big Data Analytics, Ewha Womans University

*** Corresponding Author: Kyung-shik Shin

School of Business, Ewha Womans University

52 Ewhayeodae-gil, Seodaemun-gu, Seoul, 120-750, Korea

Tel: +82-2-3277-2799, Fax: +82-2-3277-2766, E-mail: ksshin@ewha.ac.kr

or spacing errors affect the classification accuracy, especially when drawing morpheme vectors from Korean product reviews with a lot of grammatical mistakes and variations?

We seek to find empirical answers to these fundamental issues, which may be encountered first when applying various deep learning models to Korean texts. As a starting point, we summarized these issues as three central research questions as follows. First, which is better effective, to use morpheme vectors from grammatically correct texts of other domain than the analysis target, or to use morpheme vectors from considerably ungrammatical texts of the same domain, as the initial input of a deep learning model? Second, what is an appropriate morpheme vector derivation method for Korean regarding the range of POS tags, homonym, text preprocessing, minimum frequency? Third, can we get a satisfactory level of classification accuracy when applying deep learning to Korean sentiment analysis?

As an approach to these research questions, we generate various types of morpheme vectors reflecting the research questions and then compare the classification accuracy through a non-static CNN(Convolutional Neural Network) model taking in the morpheme vectors. As for training and test datasets, Naver Shopping's 17,260 cosmetics product reviews are used. To derive morpheme vectors, we use data from the same domain as the target one and data from other domain; Naver shopping's about 2 million cosmetics product reviews and 520,000 Naver News data arguably corresponding to Google's News data.

The six primary sets of morpheme vectors constructed in this study differ in terms of the following three criteria. First, they come from two types of data source; Naver news of high grammatical correctness and Naver shopping's cosmetics product reviews of low grammatical correctness. Second, they are distinguished in the degree of data preprocessing, namely, only splitting sentences or up to additional spelling and spacing corrections after sentence separation. Third, they vary concerning the form of input fed into a word vector model; whether the morphemes themselves are entered into a word vector model or with their POS tags attached. The morpheme vectors further vary depending on the consideration range of POS tags, the minimum frequency of morphemes included, and the random initialization range. All morpheme vectors are derived through CBOW(Continuous Bag-Of-Words) model with the context window 5 and the vector dimension 300.

It seems that utilizing the same domain text even with a lower degree of grammatical correctness, performing spelling and spacing corrections as well as sentence splitting, and incorporating morphemes of any POS tags including incomprehensible category lead to the better classification accuracy. The POS tag attachment, which is devised for the high proportion of homonyms in Korean, and the minimum frequency standard for the morpheme to be included seem not to have any definite influence on the classification accuracy.

Key Words : Sentiment Analysis, Morpheme Vector, Word Vector, Deep Learning, CNN, CBOW

Received : March 28, 2018 Revised : May 29, 2018 Accepted : May 30, 2018

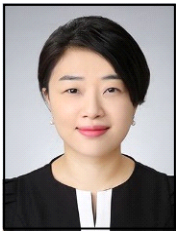
Publication Type : Regular Paper Corresponding Author : Kyung-shik Shin

저 자 소개



박현정

현재 이화여자대학교 경영연구소 연구교수로 재직 중이다. KAIST 경영과학과에서 학사와 석사 학위를, 서울대학교 경영학과에서 경영정보시스템 전공으로 박사 학위를 취득하였다. 주요 연구분야는 빅데이터 분석 및 비즈니스 응용, 텍스트 마이닝, 소셜 네트워크 분석, 비즈니스 인텔리전스, 가상화(Virtualization) 및 가상 협업(Virtual Collaboration) 등이다.



송민채

현재 이화여자대학교 빅데이터 분석학 박사과정에 있다. 이화여자대학교에서 경제학으로 석사 학위를 받고, 산업연구원, 한국은행, 한국개발연구원의 연구원으로 재직했다. 주요 연구분야는 데이터 마이닝과 텍스트 마이닝, 딥러닝 등이다.



신경식

현재 이화여자대학교 경영대학 경영학부 교수로 재직 중이다. 연세대학교 경영학과를 졸업하고, 미국 George Washington University에서 MBA, KAIST에서 경영공학 Ph.D.를 취득하였다. 주요 연구분야는 데이터 마이닝과 비즈니스 인텔리전스, 빅데이터 분석 / 비즈니스 애널리틱스(Business Analytics), 인공지능 응용과 지식공학, 가상화(Virtualization) 및 가상 협업(Virtual Collaboration) 등이다.