

양방향 LSTM을 적용한 단어의미 중의성 해소 감정분석

Emotion Analysis Using a Bidirectional LSTM for Word Sense Disambiguation

기호연 · 신경식[†]

이화여자대학교 일반대학원 빅데이터분석학 협동과정

요 약

어휘적 중의성이란 동음이의어, 다의어와 같이 단어를 2개 이상의 의미로 해석할 수 있는 경우를 의미하며, 감정을 나타내는 어휘에서도 어휘적 중의성을 띄는 경우가 다수 존재한다. 이러한 어휘들은 인간의 심리를 투영한다는 점에서 구체적이고, 풍부한 맥락을 전달하는 특징이 있다. 본 연구에서는 양방향 LSTM을 적용하여 중의성을 해소한 감정 분류 모델을 제안한다. 주변 문맥의 정보를 충분히 반영한다면, 어휘적 중의성 문제를 해결하고, 문장이 나타내려는 감정을 하나로 압축할 수 있다는 가정을 기반으로 한다. 양방향 LSTM은 문맥 정보를 필요로 하는 자연어 처리 연구 분야에서 자주 활용되는 알고리즘으로 본 연구에서도 문맥을 학습하기 위해 활용하고자 한다. GloVe 임베딩을 본 연구 모델의 임베딩 층으로 사용했으며, LSTM, RNN 알고리즘을 적용한 모델과 비교하여 본 연구 모델의 성능을 확인하였다. 이러한 프레임워크는 SNS 사용자들의 감정을 소비 욕구로 연결시킬 수 있는 마케팅 등 다양한 분야에 기여할 수 있을 것이다.

■ 중심어 : 양방향 LSTM, GloVe, 감정분석, 어휘적 중의성, 자연어처리

Abstract

Lexical ambiguity means that a word can be interpreted as two or more meanings, such as homonym and polysemy, and there are many cases of word sense ambiguity in words expressing emotions. In terms of projecting human psychology, these words convey specific and rich contexts, resulting in lexical ambiguity. In this study, we propose an emotional classification model that disambiguate word sense using bidirectional LSTM. It is based on the assumption that if the information of the surrounding context is fully reflected, the problem of lexical ambiguity can be solved and the emotions that the sentence wants to express can be expressed as one.

Bidirectional LSTM is an algorithm that is frequently used in the field of natural language processing research requiring contextual information and is also intended to be used in this study to learn context. GloVe embedding is used as the embedding layer of this research model, and the performance of this model was verified compared to the model applied with LSTM and RNN algorithms. Such a framework could contribute to various fields, including marketing, which could connect the emotions of SNS users to their desire for consumption.

■ Keyword : Bidirectional LSTM, GloVe, Emotion Analysis, Lexical Ambiguity, Word Sense Disambiguation, NLP

I. 서론

많은 사람들은 SNS 상에서 자신의 경험을 공유하고 일련의 사건에 대해 자신의 견해와 태도를 나타낸다. 이러한 텍스트 속에 일부 감정이 담겨 있으며, 평소에 느끼는 감정을 SNS 상에서 표현함으로써 심리를 표출하기도 한다.

감정은 기쁨, 슬픔 등과 같은 인간이 느끼는 심리적 상태로 감정 어휘는 자신의 심리를 투영시킨다는 점에서 구체적이며, 풍부한 맥락을 전달한다. 감정은 추론, 의사 결정 및 상호 작용과 같은 합리적인 작업을 포함해 전반적으로 인간의 행동에 영향을 미치는 중요한 요소이다. SNS 상의 사용자 감정을 이해한다면, 개인 혹은 대중의 선호도를 파악해 마케팅, 여론 조사 등 다양한 분야에서 활용할 수 있다. 특히, 마케팅 분야에서 소비자의 감정을 소비 욕구로 연결시키는 적극적 감정 마케팅 활동을 하고 있으며, 소비자들이 일상에서 경험하는 구체적이고 다양한 감정들을 제대로 이해하고 어떻게 대응하는지에 따라 성공적인 방향으로 나아갈 수 있다 [1]. 즉, 소비자의 감정을 이해하는 것이 먼저 선행되어야 하며, 이에 따라 감정을 자동 분석하려는 연구도 점차 늘고 있다. 그러나 이러한 감정 어휘는 어휘적 중의성(Lexical Ambiguity)을 띄는 경우가 다수 존재한다.

어휘적 중의성은 중의성의 한 부류로, 동음이의어(homonym)와 다의어(polysemy)를 포함한다. 어휘의 형태가 동일해도 문맥에 따라 뜻이 달라질 수 있으며, 각 문장은 다른 분위기로 해석되기도 한다. 자연어 처리 연구 분야에서 어휘적 중의성은 고질적인 문제로 이를 해소(Word Sense Disambiguation; WSD)하려는 연구는 꾸준히 이루어지고 있다. 과거에는 Support Vector Machine(SVM)과 Naïve Bayes와 같은 기법을 사용하여 감정을 분석하는 연구들이 주를 이루었다. 하지만, 최근에는 Recurrent Neural

Network(RNN), Long Short-term Memory (LSTM), Convolutional Neural Network(CNN) 등의 딥러닝을 분류기로 사용한 연구가 계속해서 발표되고 있다. 특히, 양방향 LSTM(Bidirectional LSTM)은 문장을 각각 앞, 뒤로 학습하여 문장 내 모든 정보를 균형 있게 담을 수 있어 문맥 정보를 필요로 하는 텍스트 분류, 음성 인식 등의 NLP 분야에서 두각을 나타낸다. 또한, 기울기 소실 문제를 극복하는데 탁월해 긴 시퀀스를 입력 받아도 충분히 모든 정보를 반영한다. 어휘적 중의성을 띄는 경우에 문장 내 각각의 맥락 정보들이 충분히 고려된다면, 어휘의 의미를 더욱 정확하게 파악하여 어휘적 중의성을 해소하고, 문장의 의미가 어느 하나로 고정될 수 있다[2].

따라서, 본 연구에서는 어휘의 중의성을 해소하는 감정 분석을 수행하기 위해 양방향 LSTM에 기반한 새로운 감정 분류 모형을 제안한다. 실제로 문맥 정보를 충분히 활용하여 어휘적 중의성을 해소해 감정 분류 성능을 높일 수 있는지 가설을 세운 후 실험을 진행하고자 한다. 크라우드 소싱 플랫폼 2곳에서 공개한 감정 레이블이 부착된 텍스트 데이터를 사용해 학습 데이터를 구축했으며, 이후, 양방향 LSTM으로 과거, 미래의 모든 정보를 반영하되, 전역적인 단어의 의미 정보를 활용하는 GloVe 임베딩을 적용하여 감정 분석을 시행하였다. 또한, 활성화 함수를 양방향 LSTM의 특성을 최대화할 수 있는 ReLU 함수를 적용하는 등의 하이퍼튜닝 과정을 거쳐 모델의 성능을 높이는데 기여하였다. 최종적으로, 다른 분류기인 LSTM과 RNN을 적용했을 때의 결과와 비교하여 성능에 얼마나 차이가 있는지 확인하였으며, 정확도와 f1-score 지표를 통해 연구자가 내세운 가설이 타당하고, 제안 모형이 가장 적절하다는 사실을 증명하였다.

본 논문의 구성은 다음과 같다. 2절에서는 감정 분석, 어휘적 중의성 해소, 그리고 양방향

LSTM의 활용과 관련된 선행 연구를 살펴본다. 3절에서는 본 연구에서 사용한 기법을 설명하고, 4절에서는 구체적으로 본 논문에서 제안하는 모형과 실험을 설계한 과정을 서술한다. 5절에서는 진행한 실험의 결과를 도출하고 성능을 비교 평가하고, 마지막 6절에서는 결론을 제시한다.

II. 선행 연구

2.1 감정 분석

심리학에서 감정은 생각과 행동에 영향을 미치는 복잡한 상태로 정의된다. 이는 사회적으로 형성되기도 하며, 실제 신체의 생리적, 행동적 변화를 수반하는 의식적 정신 반응으로 여긴다[3]. 즉, 인간의 행동에 영향력을 행사하기에 감정을 이해하는 일은 중요하며, 관련 연구도 끊임없이 등장하고 있다. 대표적으로 감정을 범주화 한 연구는 4가지로 볼 수 있다[4]. 인간의 가장 기본적인 감정을 총 6가지로 ‘anger’, ‘disgust’, ‘fear’, ‘joy’, ‘sadness’, ‘surprise’로 구분 짓기도 하며[5], ‘anger’, ‘anticipation’, ‘disgust’, ‘fear’, ‘joy’, ‘sadness’, ‘surprise’, ‘trust’ 등 총 8가지의 감정을 분류해 서로 반대되는 네쌍의 감정으로 이루어져 바퀴를 만들고 행위에 방아쇠를 당긴다고 설명한다[6]. 이 뿐만 아니라 신경 전달 물질이 행동에 영향을 미쳐 발생하는 8가지 감정 ‘anger’, ‘disgust’, ‘distress’, ‘fear’, ‘joy’, ‘interest’, ‘shame’, ‘surprise’를 3차원적으로 표현한 연구도 있다[7]. 마지막으로 Shaver의 연구[8]에서는 기본적인 감정은 135개의 하위 감정 용어를 클러스터링 했을 때 생성된 감정 집합 {love, joy, anger, sadness, fear, surprise}이라고 정의한다. 여기서 하위 감정 용어는 기본 수준의 감정 집합 내 6가지 감정 중 하나이거나, 여러 감정의 혼합된 형태로 간주한다. 감정은 사

회적 상호 작용에 중요한 역할을 한다. 상호 작용의 상당 부분은 자신과 다른 사람의 감정 반응을 해석하고, 다른 사람의 감정에 영향을 미치려고 시도하는 것을 포함한다. 이러한 프로토타입 접근법은 사회적 상호 작용을 포함한 다양한 실제 상황에서 감정 관련 정보가 처리되는 방식을 결정하는 데 유용하다. 그렇기 때문에 본 연구에서는 감정 클래스로 Shaver의 연구[8]에서 제시한 6가지 기본적인 감정을 채택하고자 한다.

2.2 어휘적 중의성 해소

자연어 처리 연구는 최근 딥러닝(Deep Learning) 처리 기술과 함께 발달하고 있지만, 각국마다 자연어의 특성이 다르고, 처리하기 까다롭기 때문에 완벽하게 처리하기란 쉽지 않다. 이 중에서도 단어의 중의성은 고질적인 문제로 이를 꾸준히 해결하려는 시도가 이어져 오고 있다. 단어의 중의성은 어휘적 중의성, 구적 중의성, 구조적 중의성으로 유형을 나눌 수 있는데, 주로 어휘적 중의성을 다룬다[2]. 여기에는 다의어와 동음이의어가 포함되는데, 모두 한 단어가 두 가지 이상의 뜻을 가진다는 공통점이 있다. 어휘적 중의성은 다의성 혹은 동음이의성 특징을 지닌 어떤 단어가 포함된 문장의 의미를 특정할 문맥이 충분하지 않기 때문에 발생한다. 어휘적 중의성을 지닌 단어들은 충분한 의미 정보를 필요로 하는데 지식 기반 학습, 지도 학습, 비지도 학습으로 구분 지어 중의성을 해소할 수 있다.

지식 기반 학습(Knowledge-based Learning) 방법은 WordNet, 시소러스(thesaurus) 등의 어휘 의미망을 사용하여 그 안에 담겨 있는 사전적 의미와 예제를 기반으로 중의성을 해소하는 방법이다. 이러한 기법에는 대표적으로 WordNet의 뜻풀이에서 어휘가 가장 많이 중복되는

뜻풀이를 고르는 방식인 Lesk Algorithm이 있다[9].

지도 학습(Supervised Learning) 방법은 지식 기반 학습과 달리 세부 의미가 부착된 말뭉치를 바탕으로 한다. 주변 단어에 품사를 부착(Part-of-Speech Tagging)하고, 단어 간 연어, 구문 관계만을 활용해 SVM으로 지도 학습 기반의 새로운 중의성 해소 기법을 제안하기도 하였다[10]. 또한, [11]에서는 Naïve Bayes로 해당 단어의 주변 맥락에 있는 단어들과 연어들의 집합을 이용해서 세부 의미를 판별하였다. [12]에서는 말뭉치에 품사를 태깅하여 Word2Vec으로 학습 자질을 훈련시킨 다음, LSTM을 적용하여 중의성을 해소하고자 하였다. 이때, SVM, k-NN, Naïve Bayes를 적용했을 때와 비교 실험을 진행했으나, LSTM이 가장 성능이 우수하다고 판단되었다. 지도 학습으로 중의성을 해소한다면 세부 의미가 부착된 양질의 말뭉치를 학습 데이터로 사용하여 다른 학습 방법보다 성능이 우수할 수 있지만, 이를 구축하는데 시간적, 비용적으로 소모가 큰 단점도 존재한다.

비지도 학습(Unsupervised Learning) 방법은 지도 학습과 동일하게 말뭉치를 기반으로 하지만, 학습 방법이 다르다. [13]에서는 단어가 각각 한 담화당, 한 연어당 의미를 가진다는 제약 조건을 기반으로 Decision List를 적용한 비지도 학습을 소개한다. Context Clustering 기법은 문맥 안에 등장하는 단어들이 비슷하다면 그들의 세부적인 의미도 비슷할 것이라는 아이디어에서 출발한다[14]. 지도 학습과 달리 세부 의미가 부착된 말뭉치는 필수가 아니며, 말뭉치로부터 자동적으로 훈련하기 때문에 최근 가장 많이 활용하는 방법으로 떠오르고 있다. 하지만, 여전히 성능이 떨어지는 문제점이 존재한다. 이 외에도 지도 학습과 지식 기반 학습을 적절히 조합해 말뭉치와 사전 전부 활용해서 중의성 해소에 기여하기도 한다.

2.3 문맥을 반영한 양방향 LSTM

정방향, 역방향으로 각각 시퀀스를 학습하는 양방향 LSTM(Bidirectional Long Short-term Memory)은 문맥적 정보를 요하는 sequence learning tasks에서 뛰어난 성능을 보여준다[15]. 개체명 인식에도 양방향 LSTM이 활용되는데, 데이터베이스에서 질의에 포함된 단어를 검색해 개체명을 인식하는 기존의 방식에서 양방향 LSTM을 적용해 질의가 가지는 문맥적 의미를 반영한다[16]. 글을 작성할 때 유사한 어휘로 대체하고, 문법적 오류를 정정할 수 있는 방법을 모색하기도 하는데, 양방향 LSTM에 WordNet의 사전적 정의를 사용해 태깅을 함으로써 적절한 어휘 선택에 도움을 주기도 한다[17].

양방향 LSTM은 중의성 해소에도 여러 차례 활용되었으며, [18]에서는 단일 심층 양방향 LSTM 모델을 구축하여 모든 중의적인 단어를 고려하였다. [19]에서는 양방향 LSTM으로 한국어에서의 중의적 표현을 해소했으며, 이 때, LSTM, GRU(Gated Recurrent Units)과 성능을 비교하였다. 이 외에, 의공학 분야에서도 의공학 특징 상 중의적인 의료 용어가 많아 이를 해결하려는 시도가 여럿 나타난다[20]. [21]에선 어텐션 기반한 신경망과 양방향 LSTM을 결합하였으며, 양방향 LSTM에 Label Propagation(LP)에 기반한 준지도 학습을 적용해 의공학 데이터의 의존도를 줄이고 더욱 의미 있는 정보를 추출해 임베딩을 할 수 있도록 하였다.

III. 연구 기법

3.1 GloVe

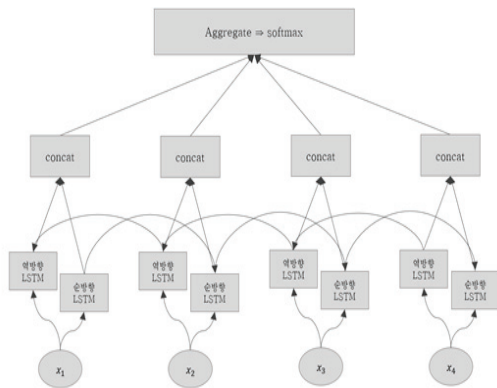
기계 학습 시 텍스트 데이터를 입력값으로 넣기 위해서는 데이터로부터 학습 자질을 추출한 다음, 적절한 벡터 값으로 수치화 하는 과정이 필요하다. 이러한 임베딩 작업은 분류 성능을

향상시키는데 중요한 역할을 하며, GloVe 임베딩을 예로 들 수 있다[22].

GloVe는 2014년 스탠포드 대학(Stanford University)에서 개발한 카운트 기반의 비지도 학습 방법론으로, 전체 말뭉치 내 단어 간 동시 등장 확률을 사용해 차원을 축소하는 등 기존의 BoW, TF-IDF, LSA 및 Word2Vec을 보완한다. 본 연구에서는 제한적인 데이터 수의 한계를 극복하고 도메인에 상관없이 사람들이 SNS에 쓴 텍스트로부터 감정을 추출하기 위해 GloVe를 사용하고자 한다. 본 논문의 연구 모형인 양방향 LSTM과 결합해 자질을 학습시키고, 풍부한 의미 정보를 제공 받을 것이다.

3.2 양방향 LSTM

양방향 LSTM(Bidirectional LSTM)은 순방향으로만 학습하는 LSTM에 역방향으로 학습하는 레이어를 추가한 LSTM 응용 기법이다[23]. <그림 1>은 양방향 LSTM의 구조를 나타낸다.



<그림 1> 양방향 LSTM 구조

문장의 감정을 나타내는 주요 어휘의 의미를 파악하고자 할 때, 그 어휘 이전에 출현한 단어보다 이후에 출현한 단어들을 살폈을 때 더욱 풍부한 의미 정보를 얻을 수도 있다. RNN, LSTM 두 기법은 순방향으로만 학습하기 때문

에 뒤의 시퀀스는 고려하지 못한 채, 앞의 시퀀스만으로 의미를 유추해야 한다.

그러나 양방향 LSTM은 양쪽 방향에서 시퀀스를 읽고 처리하여 문서 내 전체 패턴을 활용할 수 있다. LSTM 알고리즘은 은닉층에 input gate, forget gate, output gate 3개의 cell state가 있어 정보의 양을 조절할 수 있는데, 양방향 LSTM 또한 이 구조를 유지하고 있어 장기 의존성 문제도 해결한다. 다시 말해서, 양방향 LSTM은 가변적인 시퀀스의 길이에 구애 받지 않아 먼 과거의 정보도 유연하게 수용할 수 있으며, 양방향으로 시퀀스를 처리하여 문서 내의 모든 문맥적 정보를 반영할 수 있다. 양방향 LSTM의 각 은닉층은 순방향, 역방향으로 학습하는데, 서로의 레이어에 영향을 주지 않고 독립적으로 학습한다. 입력층으로부터 데이터를 입력 받으면, 첫 번째 LSTM 레이어는 좌에서 우로, 두 번째 LSTM 레이어는 우에서 좌로 데이터를 각각 처리한다. 이러한 방법은 균형 잡힌 문맥 정보를 획득함으로써 단어와 문서 간의 유사도를 구하는데 기여해 분류 성능을 높일 수 있다.

본 연구는 사전 학습한 GloVe 워드 임베딩과 양방향 LSTM의 구조로부터 2가지 의미를 얻는다. 기존에는 random initialized된 임베딩 층을 사용했으나, GloVe 벡터로 해당 토큰들을 밀집된 표현(distributed representation)으로 나타내 의미적, 문법적 정보를 최대한 반영할 수 있다. 또한, 문서 내 모든 문맥 정보를 반영함으로써 감정적 어휘의 중의적인 표현들을 해소해 감정 분류의 성능을 높일 수 있다.

IV. 실험 설계

4.1 실험 데이터

본 연구에서는 감정을 분류하기 위한 학습 및 평가 데이터로 크라우드 소싱 플랫폼(Crowd

Sourcing Platform) 2곳에서 공개한 감정 레이블이 부착된 감정 텍스트 데이터를 활용하였다. 감정 레이블은 anger, fear, joy, love, sadness, surprise로 6가지이며, 각각의 개수는 2,490개다 <표 1>.

<표 1> 실험 데이터 유형

실험 데이터		
사전 학습용 데이터	120만 개의 트위터(twitter) 단어	
학습 및 평가 데이터	클라우드소싱 플랫폼 14,940 개의 문장	
	감정 레이블	Raw data
	Anger	2,490
	Fear	2,490
	Joy	2,490
	Love	2,490
	Sadness	2,490
Surprise	2,490	

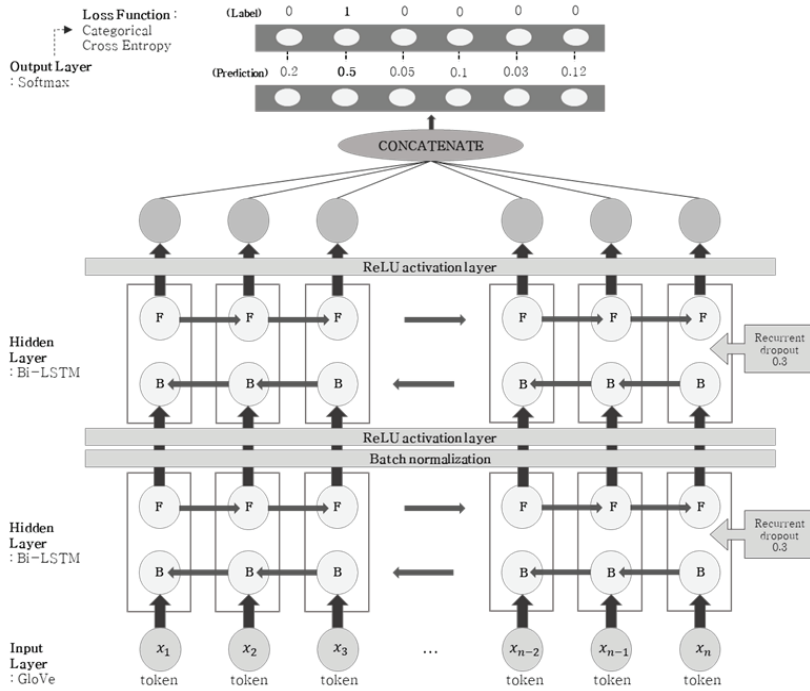
또한, 훈련 및 평가 데이터 외에 트위터에서 크롤링하여 사전 학습시킨 120만 개 GloVe 벡터를 활용해 일반적인 SNS의 의미적, 구문적 정보를 활용하고자 하였다. 감정 어휘는 동일한 형태일지라도 사용자에게 따라 의미가 달리 해석되기도 하며, 더 나아가 함께 쓰이는 단어와 문맥에 따라 각 문장이 나타내고자 하는 주요 감정, 즉 문장의 분위기도 달라질 수 있다. 연구자가 보유한 <표 2>는 이에 해당하는 예시를 보여준다. ‘Emotional’ 단어가 포함된 텍스트를 각 감정 레이블 당 1개씩 나타냈으며, 각 텍스트의 주요 감정을 나타내는 어휘가 동일함에도 불구하고 감정별로 고루 분포되어 있음을 확인할 수 있다. 따라서, 감정 어휘의 중의성을 문장 내 같이 등장하는 단어들이 구성하고 있는 문맥을 이해함으로써 해소한다면, 각 문장이 암시하려는 감정을 하나의 레이블로 분류할 수 있게 된다.

<표 2> Emotional 포함된 감정 레이블별 문장 예시

감정	Emotional이 포함된 문장
Anger	I am always so emotionally driven that despite being grateful for having work I go through very slow and or dead seasons throughout the year I couldn't help feeling annoyed and resentful because I was being pulled away from one of my passions.
Fear	I chose to attempt a similar feeling in my photographs evoking the vulnerable painful self-doubting emotional tole physical and emotional abuse has on a woman due to bullying sometimes family disapproval or societal abuse.
Joy	@stereophonics - I feel really emotional Its been great seeing your pics and keeping up with you.
Love	I feel like I am expending a lot of effort in emotionally supporting them.
Sadness	last day of school and last concert of my life.. wow.. can you say emotional? im really gonna miss mhs.. had graduation practice today..
Surprise	My sister graduated today. it was amazing and super emotional. Wow ...

4.2 데이터 가공

자연어를 처리하기에 앞서, 이를 정제하는 과정이 선행되어야 한다. 본 논문에서는 해시태그 표현, 특수문자 및 문장 부호와 같은 불용어를 처리하였다. 그 다음, 데이터를 살폈을 때, 문장을 구성하는 단어의 수(n)가 2개 이하일 때는 의미 있는 학습이 불가능하다고 판단하여 제거하였다. 또한, 대문자는 소문자로 모두 바꿔 동일한 어휘로 표현될 수 있도록 했다. 최종적으로, 트위터로부터 추출한 사전 학습시켜 생성한 120만 개의 GloVe 벡터를 활용하였다. 이로써, 어절 단위로 분절된 텍스트를 GloVe 임베딩 기법에 적용시켜 GloVe 벡터의 의미적, 문법적 정보를 담았다.



<그림 2> 본 연구에서 제안하는 모형 구조

4.3 감정 분류 모델

본 연구에서 제안하는 어휘적 중의성을 해소한 감정 분류 모델의 전체적인 구조는 <그림 2>와 같다.

본 연구에서 진행하는 실험 과정은 다음과 같다. 앞서 데이터를 훈련, 검증, 테스트 데이터로 7:1:2 비율에 맞춰 분리하였으며, 이 중 훈련, 검증 데이터를 학습과 10차 교차 검증에 사용하고, 테스트 데이터로 분류 예측을 실시하였다.

데이터를 분리한 후, 모델의 입력층에서 GloVe 임베딩 벡터를 look-up 하는 형태로 해당하는 임베딩 벡터를 추출하여, GloVe의 풍부한 의미 정보를 받았다.

이후 순방향, 역방향으로 각각 독립적으로 작동하는 은닉층을 지나며 문맥적 정보를 학습하였다. 문장을 좌측, 우측 양쪽에서 읽어 타깃 단어를 예측하는데, 타깃 단어 이전 주변 단어들의 정보와 이후 주변 단어들의 정보를 종합적으로 살펴 밀집 표현으로 나타냄으로써, 문장 내

전체 문맥 정보를 빠짐없이 담아냈다.

연구자가 제안하는 모델에서는 양방향 LSTM을 2층으로 쌓은 stacked 형태이며, 내부를 하이퍼 튜닝함으로써 최적의 모델을 찾아가 하였다. 여기엔 배치 정규화(batch normalization), 드롭아웃(dropout), ReLU 활성화 함수 적용이 해당된다. 입력값의 분산을 최소화하여 학습이 잘 되도록 배치 정규화를 실시했으며, 과적합을 방지하기 위해 랜덤하게 은닉층의 뉴런을 삭제하도록 드롭아웃을 설정했다.

ReLU 활성화 함수를 적용하기에 앞서, 각 문장의 단어들을 학습시킨 다음, 이를 결합하여 하나의 벡터로 만드는 과정을 수행하였다. 이 벡터는 문장의 임베딩 벡터라 할 수 있으며, 이를 softmax 활성화 함수에 적용시켜 0에서 1 사이의 확률로 출력하였다. 최종 확률 값과 실제 문장의 레이블을 비교하고, 정답에 근접한 확률이 출력될 때까지 학습을 진행하였다. 이 때 손실 함수인 교차 크로스 엔트로피(categorical

cross entropy)를 사용하는데, 이는 역전파를 계산하여 오차를 줄이는 방향으로 가중치를 업데이트하는 방식이다. 역전파가 문장의 양쪽 끝까지 제대로 전달될 수 있도록 sigmoid 활성화 함수가 아닌 ReLU 활성화 함수를 사용하였다. 0에서 1 사이의 값을 반영하는 sigmoid 함수와는 달리, ReLU 함수는 0보다 작은 값이면 0을 반환하고, 1보다 크면 그 값 그대로 반환하는데, 이러한 특징 덕분에 기울기 소실 문제를 대폭 줄일 수 있다[24].

연구자는 <표 3>에서처럼 GloVe 임베딩 층을 통계 변수로 사용하되, LSTM과 RNN을 벤치마크 모델의 알고리즘으로 사용하여 연구자가 제안하는 연구 모델과 비교하였다. 또한, ReLU 활성화 함수를 적용하는 등의 하이퍼튜닝 과정을 거쳐 모델의 성능을 한층 향상시키고자 하며, 구체적인 실험 설계 관련한 매개변수 설정 값은 <표 4>에 나와 있다.

<표 3> 제안 모델과 비교 모델

	임베딩	ReLU 적용	분류 모델	모델명
제안 모델	GloVe	O	양방향 LSTM	Glo-Re-BLSTM
비교 모델		O	LSTM	Glo-Re-LSTM
		O	RNN	Glo-Re-RNN
		X	양방향 LSTM	Glo-BLSTM
		X	LSTM	Glo-LSTM
		X	RNN	Glo-RNN

<표 4> 매개변수 설정 값

매개변수	설정 값
Input size	79
Max features	30,000
Batch size	64 / 128
Epochs	100
Activation function	ReLU
Optimization	Adam
Output layer	Softmax

V. 연구 결과

5.1 정확도 지표

본 연구에서는 감정 어휘의 중의성을 해소하여 감정을 분류하기 위해 GloVe 임베딩 층과 양방향 LSTM을 조합한 GloB-LSTM 연구 모델을 제안 하였으며, LSTM, RNN을 사용했을 때보다 맥락 정보를 더 반영함으로써 분류 성능을 높일 수 있는지 살펴보았다. GloVe 임베딩 층을 적용한 각 알고리즘 간의 비교 실험을 진행했고, 배치 사이즈에 따라 각각의 정확도를 구하였다 <표 5>.

<표 5> 배치 사이즈별 제안, 비교 모델 간의 정확도

정확도	배치 크기 64			배치 크기 128		
	학습	검증	테스트	학습	검증	테스트
Glo-Re-BLSTM	0.6210	0.6313	0.5940	0.6593	0.6346	0.6007
Glo-Re-LSTM	0.6183	0.6171	0.5847	0.5934	0.6171	0.5888
Glo-Re-RNN	0.3368	0.3921	0.3621	0.3255	0.3737	0.3591
Glo-BLSTM	0.6426	0.6446	0.5917	0.6282	0.6321	0.5961
Glo-LSTM	0.6667	0.6263	0.5776	0.6423	0.6346	0.5907
Glo-RNN	0.3387	0.3997	0.3758	0.3472	0.3905	0.3752

연구자가 제안한 모델(Glo-Re-LSTM)의 정확도가 비교 모델들의 정확도 대비 가장 높게 나왔으며, ReLU 함수의 적용에 따라 먼저 결과를 비교하였다. ReLU 함수를 활성화 함수로 적용하지 않은 상황에서 배치 사이즈가 64, 128일 때 Glo-BLSTM의 테스트 정확도는 0.5917과 0.5961로 Glo-LSTM과 Glo-RNN의 테스트 정확도보다 비교적 높게 나왔다. 배치 사이즈별로 비교했을 때 128로 배치 사이즈를 조절하자 전반적으로 정확도가 향상되었음을 확인하였다.

ReLU 함수를 활성화 함수로 설정했을 때의 결과를 보면, 배치 사이즈가 64일 때, Glo-Re-

BLSTM의 테스트 정확도는 0.5940으로 Glo-Re-LSTM의 정확도 0.5847보다 0.0093 증가하였고, Glo-Re-RNN의 정확도 0.3621보다 0.2319 증가하였다. 배치 사이즈가 128인 경우는 Glo-Re-BLSTM의 테스트 정확도가 0.6007 이며 Glo-Re-LSTM의 정확도 0.5888보다 0.0097 증가하였고, Glo-Re-RNN의 정확도 0.3591보다 0.2416 증가하였다. 또한, Glo-BLSTM보다 배치 사이즈가 64일 때 0.0023만큼, 배치 사이즈가 128일 때 0.0046만큼 정확도가 향상되었다.

이 결과로 5가지 사실을 확인할 수 있다. 첫째, ReLU 활성화 함수 적용 유무에 상관없이 양방향 LSTM을 알고리즘으로 사용했을 때 가장 분류 성능이 높으며, 둘째, 동일한 배치 사이즈 내에서 ReLU 활성화 함수를 적용하자 성능이 향상되었다는 점이다. 셋째, 배치 사이즈와 ReLU 적용 유무에 따라 정확도 간의 차이가 발생하는데, 배치 사이즈가 128일 때 ReLU 활성화 함수를 적용하지 않은 모델(Glo-BLSTM)의 성능이 배치 사이즈가 64일 때 ReLU 활성화 함수를 적용한 모델(Glo-Re-BLSTM)보다 우수할 수 있다는 것이다. 넷째, RNN을 적용한 비교 모델보다 가파르게 성능이 향상된 것을 보아, 양방향 LSTM이 RNN의 장기 의존성 문제를 완화하여 분류 성능에 기여했음을 알 수 있다. 다섯째, LSTM보다 성능이 향상된 결과로부터 역방향으로 학습하는 과정이 추가되어 감정 어휘가 문맥 정보를 충분히 반영함으로써 어휘의 의미를 더욱 정확하게 파악하고, 더 나아가 문장의 의미가 하나로 고정되었음을 알 수 있다.

5.2 F1-score 지표

또 다른 분류 성능 지표인 F1-score로 제안 모델과 비교 모델 간의 비교 분석을 실시했으며, 그 결과는 <표 6>에 나와 있다.

ReLU 활성화 함수를 적용하지 않은 케이스부터 살폈을 때, 배치 사이즈가 64인 경우의

<표 6> 제안, 비교 모델 간의 분류 결과 표

분류 지표	배치 크기 64			배치 크기 128		
	재현율	정밀도	F1-score	재현율	정밀도	F1-score
Glo-Re-BLSRM	0.6115	0.6163	0.6139	0.6436	0.6144	0.6286
Glo-Re-LSTM	0.5778	0.5847	0.5812	0.5951	0.5860	0.5905
Glo-Re-RNN	0.3751	0.4121	0.3928	0.3894	0.4091	0.3990
Glo-BLSTM	0.5900	0.5857	0.5878	0.6036	0.5849	0.5941
Glo-LSTM	0.5618	0.5824	0.5719	0.5886	0.5817	0.5851
Glo-RNN	0.3929	0.3758	0.3842	0.3605	0.3752	0.3677

Glo-BLSTM 정확도 0.5878과 배치 사이즈가 128인 경우의 정확도 0.5941은 Glo-LSTM과 Glo-RNN의 정확도보다 높다. 추가적으로, 배치 사이즈에 따라 비교하자 정확도와 유사하게 128로 모델의 배치 사이즈를 설정했을 때, 64로 설정했을 때보다 성능이 향상되었다.

각 감정별 성능도 비교했는데, 그 결과는 <표 7>부터 <표 12>에 나와 있다. 배치 사이즈가 128일 때 Anger 감정부터 Surprise 감정까지의 Glo-Re-LSTM F1-score은 0.7033, 0.5441, 0.5762, 0.6948, 0.5401, 0.7090으로 Glo-Re-LSTM과 Glo-Re-RNN 대비 가장 높은 값을 나타낸다. Glo-Re-LSTM의 배치 사이즈가 64일 때의 F1-score

<표 7> Anger 감정 분류 결과 표

Anger	배치 크기 64			배치 크기 128		
	재현율	정밀도	F1-score	재현율	정밀도	F1-score
Glo-Re-BLSRM	0.7203	0.6827	0.7010	0.7416	0.6687	0.7033
Glo-Re-LSTM	0.7206	0.6526	0.6849	0.6953	0.6844	0.6898
Glo-Re-RNN	0.4621	0.5884	0.5177	0.4196	0.4297	0.4246
Glo-BLSTM	0.7055	0.6928	0.6991	0.7378	0.6547	0.6938
Glo-LSTM	0.6998	0.6647	0.6818	0.6929	0.6888	0.6908
Glo-RNN	0.5070	0.5803	0.5412	0.5144	0.5743	0.5427

〈표 8〉 Fear 감정 분류 결과 표

Fear	배치 크기 64			배치 크기 128		
	재현율	정밀도	F1-score	재현율	정밀도	F1-score
Glo-Re-BLSRM	0.5108	0.5212	0.5159	0.5681	0.5221	0.5441
Glo-Re-LSTM	0.4594	0.5562	0.5032	0.5662	0.4378	0.4938
Glo-Re-RNN	0.2870	0.3988	0.3338	0.2846	0.3703	0.3218
Glo-BLSTM	0.5266	0.4578	0.4898	0.5010	0.4819	0.4913
Glo-LSTM	0.4872	0.4578	0.4720	0.4740	0.4759	0.4749
Glo-RNN	0.2232	0.3012	0.2564	0.2668	0.2068	0.2330

〈표 11〉 Sadness 감정 분류 결과 표

Sadness	배치 크기 64			배치 크기 128		
	재현율	정밀도	F1-score	재현율	정밀도	F1-score
Glo-Re-BLSRM	0.5222	0.5478	0.5347	0.5820	0.5039	0.5401
Glo-Re-LSTM	0.5049	0.4157	0.4560	0.4665	0.4759	0.4712
Glo-Re-RNN	0.3775	0.2145	0.2736	0.3495	0.2610	0.2988
Glo-BLSTM	0.4867	0.4398	0.4621	0.5175	0.4810	0.4986
Glo-LSTM	0.4465	0.4600	0.4531	0.4781	0.4739	0.4760
Glo-RNN	0.3168	0.2570	0.2838	0.3246	0.2490	0.2818

〈표 9〉 Joy 감정 분류 결과 표

Joy	배치 크기 64			배치 크기 128		
	재현율	정밀도	F1-score	재현율	정밀도	F1-score
Glo-Re-BLSRM	0.5252	0.5643	0.5440	0.5684	0.5843	0.5762
Glo-Re-LSTM	0.5364	0.5181	0.5271	0.4992	0.6205	0.5533
Glo-Re-RNN	0.2649	0.3755	0.3106	0.2882	0.3976	0.3342
Glo-BLSTM	0.4975	0.5864	0.5383	0.5797	0.5402	0.5593
Glo-LSTM	0.5238	0.5522	0.5376	0.5437	0.5743	0.5586
Glo-RNN	0.3150	0.4036	0.3538	0.3130	0.2169	0.2562

〈표 12〉 Surprise 감정 분류 결과 표

Surprise	배치 크기 64			배치 크기 128		
	재현율	정밀도	F1-score	재현율	정밀도	F1-score
Glo-Re-BLSRM	0.6807	0.7129	0.6964	0.7217	0.6967	0.7090
Glo-Re-LSTM	0.5476	0.6928	0.6117	0.6375	0.6426	0.6400
Glo-Re-RNN	0.4977	0.3169	0.3872	0.4870	0.4518	0.4687
Glo-BLSTM	0.6341	0.6647	0.6490	0.6236	0.6687	0.6454
Glo-LSTM	0.5591	0.6747	0.6115	0.6503	0.6124	0.6308
Glo-RNN	0.3754	0.2510	0.3008	0.3099	0.3635	0.3346

〈표 10〉 Love 감정 분류 결과 표

Love	배치 크기 64			배치 크기 128		
	재현율	정밀도	F1-score	재현율	정밀도	F1-score
Glo-Re-BLSRM	0.7100	0.6687	0.6887	0.6795	0.7108	0.6948
Glo-Re-LSTM	0.6979	0.6727	0.6851	0.7056	0.6546	0.6791
Glo-Re-RNN	0.3615	0.5787	0.4450	0.5075	0.5442	0.5252
Glo-BLSTM	0.6893	0.6727	0.6809	0.6888	0.6888	0.6888
Glo-LSTM	0.6545	0.6847	0.6693	0.6925	0.6647	0.6783
Glo-RNN	0.6199	0.4618	0.5293	0.4340	0.6406	0.5174

은 0.7010, 0.5159, 0.5440, 0.6887, 0.5347, 0.6964로 배치 사이즈가 128일 때보다 다소 떨어지지 만, 다른 비교 모델 대비 우수한 성능을 보인다. ReLU 활성화 함수 적용 유무에 따라 결과를 비교 한 경우에도 배치 사이즈에 관계없이 ReLU 함수를 적용했을 때 성능이 높은 것을 알 수 있다. 이러한 결과는 양방향 LSTM이 LSTM과 RNN의 단점을 완화하고 어휘의 의미를 적용하 는데 있어서 양쪽 문맥을 파악함으로써 분류 성 능을 높일 수 있음을 제시한다. 또한 ReLU 활성 화 함수의 적용은 양방향 LSTM의 성능 제고에 기여함을 보여준다.

VI. 결 론

본 연구에서 SNS 이용률 증가에 따른 사용자의 감정을 이해하기 위해 감정 분석 시스템을 제시하였다. 여기에 감정 어휘의 중의성을 해소할 수 있도록 문맥적 정보를 이용하려는 자연어 처리 분야에서 주로 쓰이는 양방향 LSTM을 적용하여 모델을 설계하였다. 본 논문은 어휘적 중의성을 문맥을 통해 해소한다면, 문맥의 풍부한 의미적, 맥락적 정보를 반영한 것으로 각 문장이 나타내는 감정이 하나로 통일된다는 가설을 세우고 실험을 진행하였다. 이를 검증하기 위해 연구 모델로 양방향 LSTM을 설정했으며, GloVe 임베딩 층을 양방향 LSTM 모델의 임베딩 층으로 사용하거나 ReLU 활성화 함수 등을 추가하는 등의 하이퍼튜닝을 실시해 모델의 성능에 기여하는지 살펴보았다. 실험 결과, 정확도와 f1-score 지표를 통해 제안한 가설과 모형이 적합한 것으로 밝혀졌다. GloVe를 임베딩 층으로 사용한 양방향 LSTM 모델은 LSTM, RNN 모델에 적용한 경우보다 높은 성능을 보였으며, ReLU 활성화 함수를 적용함으로써 모델의 성능을 한층 향상시킬 수 있었다. 즉, 중의성을 띄는 어휘들이 LSTM과 RNN 모델보다 문맥적 의미를 학습하여 각 문장이 의미하는 감정을 하나로 귀결시켰음을 밝혀냈다.

참 고 문 헌

- [1] 김해룡, 안광호, 마케팅을 결정하는 소비 감정의 힘: 감정을 팔아라, 서울:원앤원북스, 2019.
- [2] 도재학, “국어 문장의 중의성에 대하여: 언표 및 발화 문장에 따른 유형 분류를 중심으로”, 아시아문화연구, 제46권, pp.39-72, 2018.
- [3] M. Munezero, “Are they different? Affect feeling emotion sentiment and opinion detection in text,” IEEE Trans, Affective Comput. Vol.5, No.2, pp.101-111, 2014.
- [4] A. Yadollahi, “Current State of Text Sentiment Analysis from Opinion to Emotion Mining,” ACM Computing Surveys, Vol.50, No.2, pp.1-33, 2017.
- [5] P. Ekman, “Emotion in the human face: Guidelines for research and an integration of findings,” New York, Permagon, 1972.
- [6] R. Plutchik, “Emotion: Theory, Research and Experience,” Academic Press, New York, NY, 1986.
- [7] H. Lovheim, “A new three-dimensional model for emotions and monoamine neurotransmitters,” Med Hypotheses, Vol.78, No.2, pp.341-348, 2012.
- [8] P. Shaver, “Emotion knowledge: Further exploration of a prototype approach,” J. Pers. Soc. Psychol, Vol.52, No.6, 1061, 1987.
- [9] M, Lesk, “Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone,” Proceedings of the 1986 SIGDOC Conference, pp.24-26, 1986.
- [10] Lee, “Supervised Word Sense Disambiguation with Support Vector Machines and Multiple Knowledge Sources,” Proceedings of Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, pp.137-140, 2004.
- [11] A. Le, “High WSD Accuracy Using Naive Bayesian Classifier with Rich Features,” PACLIC 18, Vol.18, pp.105-14, 2004.
- [12] A. Yepes, “Word embeddings and recurrent neural networks based on Long-Short Term Memory nodes in supervised biomedical word sense disambiguation,” Journal of biomedical informatics, Vol.73, pp.137-147, 2017.
- [13] D. Yarowsky, “UNSUPERVISED WORD SENSE DISAMBIGUATION RIVALING SUPER-

- VISeD METHODS,” In ACL 95, pp.189-196, 1995.
- [14] C. Niu, “Context Clustering for Word Sense Disambiguation Based on Modeling Pairwise Context Similarities,” In: SENSEVAL- 3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, pp.187-190, 2004.
- [15] A. Graves, A., and Schmidhuber, J. (2005). “Frame-wise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures,” Neural Networks, Vol.18 No.5-6, pp.602-610, 2005.
- [16] 진승희, “온톨로지 지식 기반 특성치를 활용한 Bidirectional LSTM-CRF 모델의 시퀀스 태깅 성능 향상에 관한 연구”, 지능정보연구, 제24권, 제1호, pp.253-267, 2018.
- [17] V. Makarekovic, “Choosing the right word: Using bidirectional LSTM tagger for writing support systems,” Engineering Applications of Artificial Intelligence, Vol.84, pp.1-10, 2019.
- [18] A. Pesaranhader, “One Single Deep Bidirectional LSTM Network for Word Sense Disambiguation of Text Data,” Springer Advances in Artificial Intelligence: 31st Canadian Conference on Artificial Intelligence Canada, pp.96-107, 2018.
- [19] J. Min, “A Study on Word Sense Disambiguation Using Bidirectional Recurrent Neural Network for Korean Language,” Journal of the Korea Society of Computer and Information, Vol.22, No.4, pp.41-49, 2017.
- [20] C. Zhang, “Biomedical word sense disambiguation with bidirectional long short-term memory and attention-based neural networks,” BMC Bioinformatics, Vol.20, No.502, 2019.
- [21] Z. Li, “Context Embedding Based on Bi-LSTM in Semi-Supervised Biomedical Word Sense Disambiguation,” IEEE Access, Vol.7, pp.72928-72935, 2019.
- [22] P. Jeffrey, “Glove: Global vectors for word representation,” Proceedings of the Empirical Methods in Natural Language Processing, 2014.
- [23] S. Mike, “Bidirectional recurrent neural networks,” Signal Processing, IEEE Transactions on, Vol.45, No.11, pp.2673-2681, 1997.
- [24] V. Nair, “Rectified Linear Units Improve Restricted Boltzmann Machines,” In Proceedings of the 27th International Conference on Machine Learning, pp.807-814, 2010.

저자 소개



기 호 연(Ho-Yeon Ki)

- 2018년 8월 : 이화여자대학교 통계학사
- 2020년 8월 : 이화여자대학교 일반대학원 빅데이터분석학 석사
- 관심분야 : 자연어처리



신 경 식(Kyung-shik Shin)

- 1998년 : 한국과학기술원 경영공학 박사
- 1999년-현재 : 이화여자대학교 경영대학 교수
- 관심분야 : 빅데이터 분석 및 응용