

## 딥러닝 기반 감성분석 연구동향

서상현\* · 김준태\*\*

### 1. 서 론

최근 인공지능 분야에서 많은 주목을 받는 딥러닝(Deep Learning)은 깊은 구조로 이루어진 인공신경망의 학습기법을 일컫는다[1-3]. 딥러닝은 인공신경망 학습에서 발생하던 기존의 제약사항들을 부분적으로 해소함으로써 이미지, 음성, 자연어처리 등의 분야에서 뛰어난 성과를 거두며 지속적인 관심을 받고 있다.

이미 딥러닝은 인공지능의 다양한 영역에 적용되고 있으며 그 분야를 막론하고 기존의 성과기록들을 갱신하고 있다. 한동안 세간의 주목을 받은 알파고(AlphaGo)는 기존의 게임 알고리즘에 딥러닝 기법을 적용하여 성능을 극대화하였으며[4], 페이스북은 딥러닝을 사용자의 얼굴인식에 적용하여 정확도를 크게 높였다[5]. 이미지인식 분야에서는 이미지넷(ImageNet)과 같은 경연에서 딥러닝을 적용한 모델들이 기존의 성과들을 매년 갱신하고 있으며[6], 음성인식 분야에서도 딥러닝이 활발히 활용되고 있다[7-8]. 자연어처리 분야에서는 질의응답이나 기계번역 등에서 딥러닝이

활용되고 있으며[9-11], 텍스트 마이닝(text mining)의 한 영역인 감성분석(sentiment analysis)에도 다양한 딥러닝 기법이 적용되어 의미 있는 성과를 거두고 있다[12-21].

감성분석은 텍스트 데이터에 내포되어 있는 텍스트 작성자의 의견이나 감성, 평가, 태도 등을 분석하는 일련의 과정이다[22]. 감성분석은 상품에 대한 만족도 조사에서부터 정치적 이슈에 대한 여론조사까지 다양한 산업분야에서 활용되고 있으며[23], 최근에는 사회연결망서비스(social network service)에서 생산되는 다량의 텍스트 데이터를 바탕으로 의견분석(opinion mining)이 활발히 수행되고 있다[22-24].

감성분석은 감성사전 등의 어휘자원을 기반으로 수행되는 어휘기반(lexicon-based) 접근방식[25-26]과 기계학습(machine learning)의 다양한 기법들을 활용하는 기계학습 기반의 방법론들이 주로 사용되어 왔다[27-28]. 최근에는 자연어처리 분야에 여러 가지 딥러닝 기법이 적용되어 기존의 성과들보다 뛰어난 기록들이 보고되면서, 앞서 언급한 바와 같이 감성분석 분야에도 이러한 기법들을 적용하는 다양한 시도가 이루어지고 있다. 이에 본 논문에서는 감성분석과 딥러닝의 개념에 대해 개괄하고, 딥러닝 기반의 감성분석에 대한 최근의 연구동향을 살펴보고자 한다.

\* 교신저자(Corresponding Author): 김준태, 주소:서울특별시 중구 필동로 1길 30, 동국대학교 공과대학 컴퓨터공학과  
전화: 02-2260-3712

\* 동국대학교 컴퓨터공학과  
(E-mail: shseo@dongguk.edu)

\*\* 동국대학교 컴퓨터공학과  
(E-mail: jkim@dongguk.edu)

## 2. 관련 연구

### 2.1 텍스트 마이닝과 감성분석

텍스트 마이닝은 텍스트 데이터에 숨겨진 정보를 찾아 유용하게 활용하는 일련의 과정을 일컫는다[29]. 넓은 의미에서 텍스트 마이닝의 주요 과업(task)은 정보검색(information retrieval), 정보추출(information extraction), 텍스트 범주화(text categorization), 텍스트 군집화(text clustering), 텍스트 요약(text summarization) 등으로 구분할 수 있다[30].

텍스트 마이닝의 한 분야인 텍스트 범주화는 단어, 구, 문장, 문단 등 구조화 되어있지 않은 다양한 형태의 텍스트 데이터에 대해 그 특성에 따라 범주를 구분하는 과정이다[29]. 그 중에서도 감성분석은 “긍정”과 “부정”같이 상반되는 극성(polarity)의 분류부터 “기쁨”, “슬픔”, “지루함”, “흥분”, “놀람” 등 다수의 범주에 대한 분류까지 텍스트에 내포된 의견이나 감성 등의 의미를 범주화하는 작업을 의미한다.

감성분석은 수행되는 기법에 따라서 어휘기반, 기계학습, 혼합방식(hybrid) 등으로 분류할 수 있다[31]. 어휘기반 접근방식은 기보유한 어휘자원(lexical resource)을 바탕으로 범주화 작업을 선행하여 감성사전(sentiment dictionary)을 구축한 뒤, 감성사전을 바탕으로 분석하고자 하는 텍스트 데이터에 대한 감성점수를 계산하여 범주화하는 방식을 일컫는다[26]. 어휘기반의 감성분석은 감성사전 등의 어휘자원에 대한 의존도가 높기 때문에 성능향상을 위해서는 감성사전의 구축에 많은 노력이 필요하다. 영어권에서는 SentiWordNet과 같은 어휘자원이 공개되어 다양한 연구에 활용되고 있다[26][32]. 한국어의 경우, 보편적으로 사용되는 한국어 감성사전이 부재한 상황에서 공공

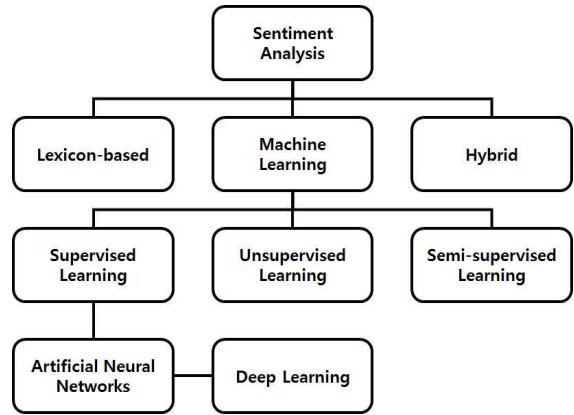


그림 1. 감성분석에 대한 접근방식 분류

한국어 감성사전 제작이 시도되고 있다[33].

기계학습 기반의 감성분석은 감독(supervised)·비감독(unsupervised)·반감독(semi-supervised) 학습의 다양한 기법들을 통해 수행된다. 그 중에서도 감성분석에 활용되는 감독 학습의 대표적인 기법으로는 나이브 베이저안 분류기(naive bayesian classifier), 서포트 벡터 머신(support vector machine), 의사결정나무(decision tree), 인공신경망(artificial neural networks) 등이 있다[34]. 그림 1은 감성분석에 대한 분류와 더불어 기계학습 기반의 학습방식을 구분하고 있다.

하지만 기존의 감성분석 접근방식은 학습 데이터의 특성(feature)에서 많은 영향을 받기 때문에 영역 적응(domain adaption)에 취약한 모습을 보인다. 영역 적응의 문제는 특정 영역의 데이터로 학습된 분류모델이 다른 영역에 적용될 때 저조한 성능을 기록하는 것을 말하며, 감성분석 분야에서도 이러한 문제를 해결하고자 다양한 시도가 이루어졌다[27][35]. 최근에는 감성분석 분야의 영역 적응 문제에 딥러닝 기법이 적용되어 학습 데이터에서 높은 수준의 특성을 추출하는 방법을 통해 뛰어난 성과를 기록한 바 있다[36].

## 2.2 딥러닝

인공신경망은 입력된 값들에 대한 연산을 통해 특정 상태값을 갖는 노드들의 연결을 통해 입력층과 출력층, 그 사이에 존재하는 다수의 은닉층으로 구축된다. 인공신경망은 출력층의 결과값과 학습 데이터의 실제값을 비교하여 그 오차를 줄이는 방향으로 노드 사이의 연결 파라미터를 수정하는 오류 역전파 기법(error backpropagation)을 통해 특정 입력에 대해 적절한 출력을 낼 수 있도록 학습된다[37]. 인공신경망은 입력과 출력의 관계가 복잡한 문제일수록 입력층과 출력층 사이에 다수의 은닉층을 구성함으로써 성능을 향상시킬 수 있다. 하지만 다수의 은닉층을 통해 인공신경망을 구성하는 접근방식은 몇몇 문제점들이 지적되어 왔다.

첫 번째 문제점은 인공신경망의 학습과정에서 기울기 사라짐 현상(vanishing gradient)이 발생하는 것이다. 기울기 사라짐 현상은 출력층에서 연산된 오류값이 입력층 근처의 연결 파라미터까지 충분히 전달되지 않아 인공신경망의 학습이 적절히 이루어지지 않는 현상을 말한다. 이로 인해 깊은 구조의 인공신경망은 오차의 학습이 적절하게 이루어지지 못하게 된다.

두 번째 문제점은 오류 역전파 기법을 통해 학습하는 인공신경망의 노드의 수가 늘어나 학습되는 연결 파라미터의 수가 늘어날수록 연산량이 증가하는 것이다. 즉, 인공신경망이 깊은 구조를 갖게 되면서 기울기 사라짐 현상과 더불어 학습속도에 있어서도 성능저하가 발생하게 된다.

마지막 문제점은 인공신경망 모델이 학습 데이터에만 적절하게 학습되는 과적합(overfitting) 문제이다. 과적합 문제는 학습 데이터에 과도하게 적합한 모델로 연결 파라미터가 학습되어 실제로

모델에 활용되는 새로운 데이터가 입력되었을 때 학습 데이터에 비해 상대적으로 저조한 성능을 보이는 것을 말한다[38].

딥러닝은 인공신경망이 깊은 구조를 가질 때 발생하는 기울기 사라짐 현상, 학습속도의 저하, 과적합 문제 등의 다양한 문제들을 보다 잘 극복한 일련의 인공신경망 기법들을 총칭하는 개념이다[1-3]. 다양한 딥러닝 기법 가운데 대표적인 기법으로는 합성곱신경망(convolutional neural networks), 깊은신뢰신경망(deep belief networks), 순환신경망(recurrent neural networks) 등이 있다.

합성곱신경망은 주로 이미지 인식과 관련된 영역에서 사용되는 딥러닝 기법이다[39-40]. 합성곱신경망은 특정 필터(filter)를 사용하여 입력된 데이터의 특징지도(feature map)을 만들고 특징지도에서 일부 데이터만 선택하여 해상도를 낮추는 과정(sub-sampling)을 반복한다. 합성곱신경망은 위의 과정을 반복하여 깊은 구조의 신경망을 구성하고 필터의 크기에 대비하여 해상도가 가장 작아지게 되면 그 특징(feature)들을 완전연결된 신경망에 입력하여 결과값을 산출한다.

깊은신뢰신경망은 제한된 볼츠만 기계(Restricted Boltzmann Machine)를 적층으로 쌓은 깊은 구조의 인공신경망을 일컫는다[41]. 제한된 볼츠만 기계는 기존의 볼츠만 기계(Boltzmann Machine)에서 같은 층 사이의 연결을 제한하고 대조적 발산(contrastive divergence) 등의 기법을 통해 연산의 속도를 높인 생성모델(generative model)이다. 제한된 볼츠만 기계는 입력된 값과 확률적으로 재구성(reconstruction)된 값의 차이에 따라 연결 파라미터가 학습된다. 제한된 볼츠만 기계의 학습은 확률적인 재구성을 통해 이루어지기 때문에 깊은 구조를 취하더라도 기울기 사라

집 현상에서 자유로운 사전학습(pre-training)이 가능하다[42]. 깊은신뢰신경망은 이미지 분류, 얼굴인식, 문장의미 분석 등에 활용되고 있고 있으며[41][43-44], 한편에서는 제한된 볼츠만 기계를 적층으로 쌓은 인공신경망(active deep learning)을 통해 감성분석에 적용한 모델이 제안되기도 하였다[20].

순환신경망은 시간적인 요소를 인공신경망에 적용하고자 하는 목적으로 제안되었다[45]. 순환신경망은 각 노드를 서로 다른 시점으로 생각하여 특정 시점의 노드가 그 이전 시점의 노드의 결과값의 영향을 받도록 설계된다. 따라서 순환신경망은 연속성에서 의미를 찾을 수 있는 순차적인 데이터(sequential data)의 분석에 용이한 신경망 구조이다. 순환신경망은 구조적 특성으로 인해 음성 인식이나 언어 모델링의 영역에서 활용되는 등 자연어와 같은 순차적 데이터와 관련된 분야에서 활용되고 있다[8][46].

### 3. 감성분석에 활용되는 딥러닝 기법

#### 3.1 Word embedding & Word2vec

인공신경망은 숫자로 이루어진 입력값을 사용하므로 딥러닝 기반의 감성분석을 수행하기 위해서는 텍스트 데이터의 각 단어들을 숫자로 표현하는 과정이 선행되어야 한다. 워드임베딩(word embedding)은 텍스트 데이터를 숫자로 이뤄진 벡터 공간으로 사상(mapping)시키는 과정을 의미한다. 워드 임베딩을 통한 단어의 표현방식은 국소표현(one-hot representation)과 분산표현(distributed representation)의 두 가지가 있다.

국소표현은  $n$ 개의 단어집합이 있을 때, 그림 2-(a)와 같이 단어  $w$ 를 그에 해당하는 원소만 1인  $n$ 차원의 벡터로 표현하는 방식이다. 하지만

$$w_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} w_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \dots w_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

(a) 국소표현(one-hot representation)

$$w_1 = \begin{bmatrix} 0.3 \\ 0.9 \\ 0.8 \\ 0.5 \\ 0.1 \end{bmatrix} w_2 = \begin{bmatrix} 0.2 \\ 0.9 \\ 0.9 \\ 0.5 \\ 0.1 \end{bmatrix} \dots w_n = \begin{bmatrix} 0.8 \\ 0.1 \\ 0.2 \\ 0.1 \\ 0.9 \end{bmatrix}$$

(b) 분산표현(distributed representation)

그림 2. 단어의 국소표현과 분산표현 예

국소표현은 단어집합의 전체적인 맥락 속에서 단어들 사이의 관계를 표현하지 못하는 단점이 있다. 국소표현과는 달리 분산표현은 단어를 그림 2-(b)와 같이 실수값으로 이루어진 일련의 벡터로 표현한다. 분산표현의 장점은 분산표현을 생성하는 언어모델의 방법에 따라서 단어 사이의 관계를 나타낼 수 있다는 것이다. 그림 2-(b)의 예시가 단어 사이의 관계가 고려된 분산표현일 때,  $w_1$ 와  $w_2$ 처럼 유사한 벡터의 값으로 구성된 단어의 경우 가까운 의미를 지닌 것으로 예상할 수 있다.

최근에는 인공신경망의 구조를 통해 워드임베딩을 수행하여 단어들을 분산표현하는 방식이 주목받고 있다[46-49]. 신경망 기반의 확률적 언어 모델(neural probabilistic language model)은 입력층, 사상층, 은닉층, 출력층의 전방향 신경망과 같이 구축되어 우수한 성과를 보였다[47]. 하지만 이 언어모델의 경우 은닉층 전후의 연산이 매우 복잡한 문제가 지적되어 왔다. 신경망 구조를 차용하는 이후의 언어모델에서도 연산의 복잡성 문제는 계속해서 지적되어 왔고 이 문제를 해결하고자 많은 언어모델들이 제안되었다.

특히 순환신경망 언어모델(recurrent neural

networks language model)은 신경망 기반의 앞선 언어모델의 사상층 부분을 제거하여 은닉층에 이르는 연산의 복잡도를 낮췄다[46]. 순환신경망 언어모델은 입력층, 은닉층(문맥층, 출력층의 세 가지 층으로 구성되었는데 이전 단계의 은닉층과 현재단계의 은닉층이 연결되는 구조가 특징적인 부분이다. 순환신경망은 이전 단계와의 연결을 통해서 입력된 데이터에 대한 단기 기억(short term memory)을 형성할 수 있고, 결과적으로 입력된 데이터에 대한 워드임베딩의 효율성을 높일 수 있었다[49].

앞서 살핀 신경망기반 언어모델에서 연산량이 문제가 되는 가장 큰 이유는 신경망 구조의 은닉층에서 비선형함수가 연산되기 때문이다[49]. 이와 같은 문제를 해결하기 위해 기존의 신경망 기반 언어모델에서 은닉층을 제거하여 입력층, 사상층, 출력층의 얇은(shallow)신경망 구조로 구축된 언어모델인 Word2vec이 제안되었다[48-49]. Word2vec은 또한 출력층에서 활용되는 소프트맥스 함수연산에 이진트리 구조를 적용한 계층적 소프트맥스(hierarchical softmax)나 네거티브 표본화(negative sampling) 등의 방법을 적용하여 연산량을 감소시켰다[48].

Word2vec은 그림 3과 같이 두 가지 구조로 제안되었는데, 첫 번째 구조는 특정 단어를 그 단어 전후에 나타나는 단어들을 바탕으로 예측하는 CBOW(continuous bag-of-words)이며, 두 번째 구조는 CBOW와는 반대로 특정 단어를 입력으로 하여 그 단어 전후의 단어들을 예측하는 Skip-gram(continuous skip-gram)이다.

CBOW의 입력값은 학습 단어의 개수인  $N$ 차원의 국소표현된 벡터들의 집합이다. 입력 벡터는  $N \times D$ 크기의 임의의 행렬과의 연산을 통해  $D$ 차원의 사상층으로 사상되고 이 값은 다시  $D \times N$

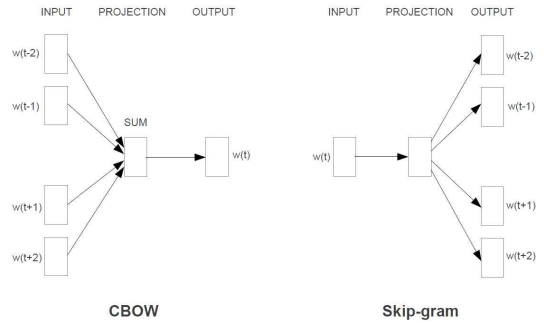


그림 3. Word2vec 구조[49]

크기의 임의의 행렬과의 연산을 통해 출력층으로 전달되어 소프트맥스 함수로 연산된다. CBOW의 학습은 학습 데이터의 단어  $w(t)$ 와 출력값의 오차를 줄이는 방향으로 진행되며, 파라미터가 수정되는 학습의 대상은 각 층의 사이에서 연산되는 행렬이다. 즉, CBOW는 학습 데이터의 특정 단어와 그 전후의 문맥과의 관계를 학습하게 되고 그 정보는 입력층과 사상층 사이의  $N \times D$ 크기의 행렬에 저장된다. 따라서 학습을 마친 뒤, 이 행렬을 활용하면 학습 데이터의 단어들 사이의 관계가 반영된 분산표현을 얻을 수 있다.

Skip-gram은 CBOW와 동일하게 각 층 사이의 행렬을 통해 연산을 수행하고 학습 데이터와 출력값과의 오차를 줄이는 방향으로 행렬이 학습되는 공통점이 있다. 다만 Skip-gram의 학습에 사용되는 단어의 수는 단어들 사이의 최대거리를 설정하여 1부터 최대거리 사이의 숫자에서 임의로 결정된다. Skip-gram도 CBOW와 마찬가지로 학습된 행렬의 값을 통해 단어들 사이의 관계가 반영된 분산표현을 얻을 수 있다[49].

특히 Word2vec은 단어들을 고차원의 벡터공간 간의 한 점으로 표현함으로써 단어들의 벡터연산에 의미를 부여하도록 하여 주목받고 있다. 식 (1)은 벡터 연산의 예시로서 “왕”을 나타내는  $\overrightarrow{King}$

에서  $\overrightarrow{Man}$  을 빼 “남자”의 의미를 제거하고,  $\overrightarrow{Woman}$  을 더해 “여자”의 의미를 부여했을 때, “여왕”을 의미하는  $\overrightarrow{Queen}$  벡터를 얻는 과정을 보여준다[49].

$$\overrightarrow{King} - \overrightarrow{Man} + \overrightarrow{Woman} = \overrightarrow{Queen} \quad (1)$$

앞서 언급한 바와 같이 딥러닝 기반의 감성분석을 수행하기 위해서는 텍스트 데이터를 신경망에 입력이 가능한 형태로 변환하는 과정이 필요하다. 국소표현된 텍스트 데이터의 활용은 단어 사이의 연관성을 활용하지 못한다는 단점이 있다. Word2vec처럼 분산표현 방식으로 단어 사이의 관계를 잘 나타내주는 언어모델을 사용함으로써 감성분석의 성능 향상을 기대할 수 있다.

### 3.2 Recurrent Neural Networks

인간의 언어 사용은 시간의 흐름에 따라 순차적으로 이루어진다. 특정 시점에 발화된 단어의 의미는 독자적으로 결정되기 보다는 그 단어를 전후로 한 맥락에 의해 결정되는 경우가 많다.

순환신경망(recurrent neural networks)은 시간의 변화에 따라 순차적으로 진행되는 시계열 데이터를 학습하는 주요한 방법이다[45]. 순환신경망의 특정 시점의 노드는 그 시점의 입력값과 더불어 이전 시점의 은닉층의 값을 활용하여 상태를 결정한다. 즉, 순환신경망은 과거의 기억들이 계속해서 누적되어 현재 시점의 상태에 영향력을 발휘할 수 있는 모델이다. 이와 같은 특성 때문에 순환신경망은 음성이나 텍스트 데이터와 같이 인간의 언어와 관련된 데이터의 활용에 있어 강점을 지닌다.

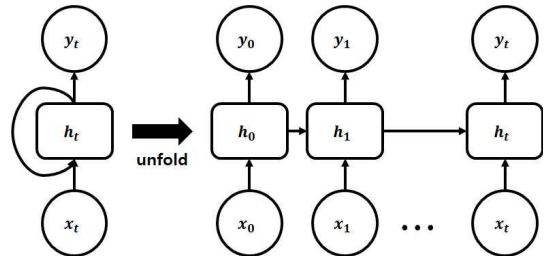


그림 4. 순환신경망의 기본구조

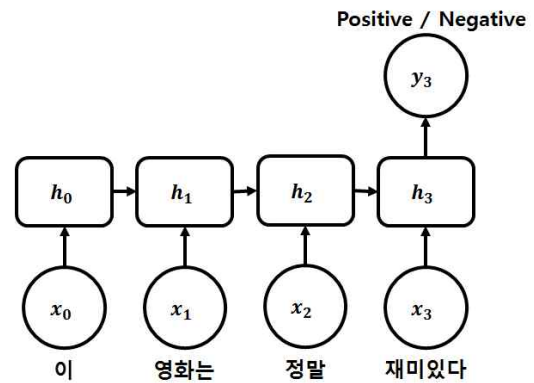


그림 5. 감성분석에 활용되는 순환신경망 예시

$$h_t = f(Wx_t + Vh_{t-1} + b) \quad (2)$$

$$y_t = g(Uh_t + c) \quad (3)$$

그림 4는 순환신경망의 기본구조를 나타낸다.  $t$ 시점의 입력값  $x_t$ 가 주어지면, 식 (2)처럼 비선형활성함수  $f$ 를 사용하여  $t$ 시점의 은닉층의 상태값  $h_t$ 가 결정된다.  $h_t$ 는  $t$ 시점에서의 입력값  $x_t$ 와 연결 파라미터  $W$ 를 곱한 값, 이전 시점 은닉층의 값  $h_{t-1}$ 와 연결 파라미터  $V$ 를 곱한 값, 마지막으로 바이어스 값인  $b$ 를 더해 결정된다. 다음으로 출력층  $y_t$ 의 값은 식 (3)처럼 은닉층의 값  $h_t$ 와 연결 파라미터  $U$ 를 곱한 값에 바이어스 값  $c$ 를 더해 산출된다. 이러한 연산과정을 시점의 변화에 따라 순차적으로 반복하면 기본적인 순환신경망이 구축된다.

순환신경망의 연산과정은 그림 5의 감성분석에 활용되는 순환신경망 예시에서도 그대로 적용된다. 그림 5에서는 “이 영화는 정말 재미있다”라는 문장이 입력값으로 제시되어 있다. 이 입력문장을 띄어쓰기 단위로 워드임베딩 한다면, 순환신경망에 입력될 네 개의 벡터값을 생성할 수 있다. 네 개의 입력값은 순환신경망의 각 시점마다 순차적으로 입력된다. 다만 기존의 순환신경망과 달리 감성분석을 위한 순환신경망은 최종 시점에 도달하기 이전의 출력값은 연산을 하지 않고 최종시점의 출력값만 연산한다.  $h_3$ 은  $h_0$ 부터 현재까지의 모든 문맥정보를 담고 있는 상태값이기 때문에  $h_3$ 을 통해 결과값  $y_3$ 을 출력함으로써 입력된 데이터에 대한 전체적인 의미가 반영된 감성분석이 가능하다.

순환신경망의 학습은 최종 시점에서의 결과값과 실제 학습 데이터의 “긍정”, “부정” 등 극성으로 범주화된 레이블과의 오차를 줄이는 방향으로 시점 간 오류역전파 기법(backpropagation through time)을 사용하여 이루어진다[50]. 위의 과정을 통해 학습된 순환신경망에 새로운 데이터가 입력되면 입력 데이터의 전체적인 의미가 담긴 결과값이 산출되고, 이를 범주화 하는 방식으로 순환신경망 기반의 감성분석이 수행된다.

위의 예시 이외에도 순환신경망은 다양한 구조로 변형되어 감성분석에 활용되고 있다. 특히 인간의 언어사용에서 특정 시점의 단어는 과거 시점 뿐 아니라 뒤에 이어지는 미래 시점의 단어들에 의해서도 의미가 결정되기도 한다. 따라서 시간에 순차적인 흐름만 반영한 기존의 순환신경망을 변형하여 뒤의 시점이 앞의 시점에 영향을 줄 수 있는 양방향 모형도 제안된 바 있다. 표 1의 우측 하단의 모형은 순환신경망이 깊은 구조로 이루어져 있으면서 각 시점 사이의 연결 파라미터가 양

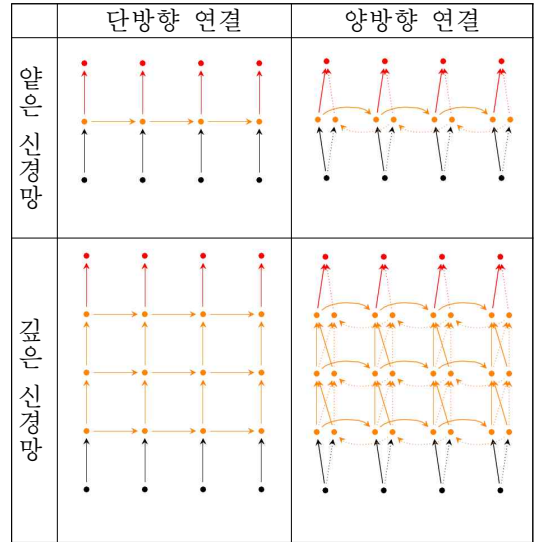


표 1. 다양한 구조의 순환신경망[13]

방향(bidirectional)으로 되어있는 깊은 구조의 양방향 순환신경망 모형이다. 실제로 이 모형은 단방향(unidirectional) 연결된 순환신경망보다 더 뛰어난 성능을 보이고 있다[13].

[12]와 [13]은 순환신경망을 활용한 감성분석의 대표적인 사례이다. [12]는 식당과 차량의 리뷰 데이터를 대상으로 순환신경망을 이용한 감성분석을 수행하여 기존의 기계학습 기반의 감성분석 기법들에 비해 우수한 성과를 거두었다. [13]은 순환신경망을 단방향과 양방향, 얕은 층과 깊은 층으로 구성하여 각각도 감성분석의 성능비교를 수행하였다. 특히 [13]은 단층에서부터 시작하여 다섯 층에 이르기까지 층의 개수에 따른 성능을 비교하였는데, 가장 많은 층을 쌓은 모델보다 중간 정도인 세 개의 층을 쌓은 모델이 가장 뛰어난 성과를 보였다.

### 3.3 Long-Short Term Memory

다양한 구조의 순환신경망이 감성분석에 적용

되고 있지만 순환신경망은 일반적으로 문장이 길어질 경우 성능이 저하된다. 긴 문장에 대한 순환신경망의 성능 저하는 학습의 과정에서 출력단의 오차를 충분히 전달하지 못하는 장기 의존성(long-term dependencies)의 문제 때문에 발생한다[40]. 이러한 장기 의존성의 문제를 극복하고자 하는 다양한 구조의 순환신경망의 모델들이 제안되었으며, 그 중 대표적인 모델로 LSTM(Long Short-Term Memory)이 있다[51].

LSTM은 기존의 순환신경망 구조에서 은닉층의 노드에 해당하는 부분들에 변형을 준 모델이다. LSTM은 시점 간 오류역전과 기법 등을 통한 순환신경망의 학습과정에서 오랜 학습시간, 불필요한 연산, 오차의 사라짐 현상 등의 문제를 해결하고자 제안되었다. 기존의 순환신경망은 그 시점의 입력값과 이전 시점에서 전달받은 값을 입력값으로 하여 활성화함수를 활성화시켜 노드의 상태를 결정한다. 이와는 다르게 LSTM은 은닉층의 노드에 분리된 메모리 공간(cell)을 설정하고 그 메모리 공간의 상태값들이 입력되거나 출력되는 길목에 특수한 게이트(gate)를 설치하여 장기 의존성 문제에 접근하였다. 그림 6은 LSTM의 기억공간(memory block)의 세부 구조를 이해하기 쉽게 묘사한 것이다. 이 그림은 하단에서부터 상단의 방향으로 LSTM 기억공간 전체의 입력(block input)과 출력(block output)이 연산되는 과정과 메모리 셀의 입력과 출력에 활용되는 세 가지 게이트인 입력 게이트(input gate), 망각 게이트(forget gate), 출력 게이트(output gate)의 활용을 보여준다[52-53].

$$z^t = g(W_z x^t + R_z y^{t-1} + b_z) \quad (4)$$

$$i^t = \sigma(W_i x^t + R_i y^{t-1} + p_i \odot c^{t-1} + b_i) \quad (5)$$

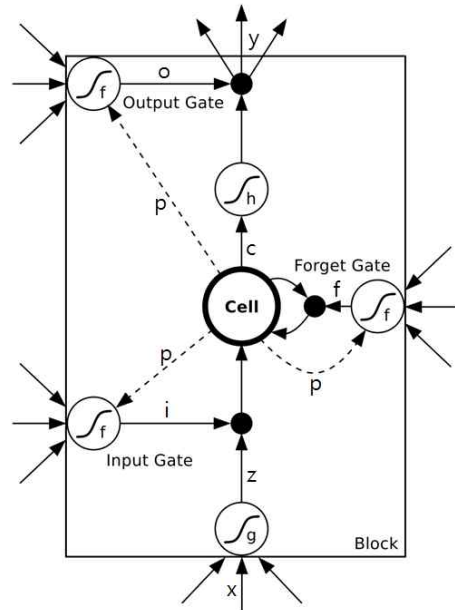


그림 6. LSTM memory block[52]

$$f^t = \sigma(W_f x^t + R_f y^{t-1} + p_f \odot c^{t-1} + b_f) \quad (6)$$

$$c^t = i^t \odot z^t + f^t \odot c^{t-1} \quad (7)$$

$$o^t = \sigma(W_o x^t + R_o y^{t-1} + p_o \odot c^t + b_o) \quad (8)$$

$$y^t = o^t \odot h(c^t) \quad (9)$$

위의 식 (4)~(9)는 그림 6의 하단에서부터 상단으로 진행되는 LSTM의 연산과정을 순차적으로 제시한 것이다[53]. 수식에서  $x^t$ ,  $y^t$ 는 시점  $t$ 의 입력과 출력을 의미하며,  $i^t$ ,  $f^t$ ,  $o^t$ 는 시점  $t$ 의 입력 게이트, 망각 게이트, 출력 게이트의 상태값이다.  $c^t$ 는 시점  $t$ 에서의 셀의 상태값이며, peep-hole connection이라 불리는  $p_i$ ,  $p_f$ ,  $p_o$ 는 셀의 내부 메모리 값과 게이트를 연결하는 가중치 벡터로서  $\odot$ 은 벡터곱(point-wise multiplication)을 나타낸다.  $b_i$ ,  $b_f$ ,  $b_o$ 는 바이어스 값이다.  $W$ 은 입력 가중치 행렬을,  $R$ 은 순환을 통한 입력에 대한 가중치 행렬을 나타낸다. LSTM은 블록 내에 독립



된 메모리 공간인 셀을 확보하고 세 게이트를 활용하여 장기 기억능력과 단기 기억능력의 균형을 맞췄다. 실제로 처음으로 제안된 LSTM 구조의 경우 1000여 시점(steps) 전까지 정보전달이 가능할 정도로 뛰어난 성능을 보였다[51].

[14]는 이미지나 비디오를 설명하는 문장들로 구성된 데이터 집합에 양방향 LSTM, 다층 LSTM, 트리구조의 LSTM 등의 다양한 LSTM을 적용하여 감성분석에 대한 성능비교를 수행하였다. [15]는 문장 내에 존재하는 감성분석 대상(target)을 파악하고 이들을 중심으로 감성분석을 수행하기 위해 대상 단어를 중심으로 전후 두 개의 LSTM을 연결하거나 대상 단어의 벡터를 모든 입력 데이터에 추가하는 기법 등을 제안하였다. 이러한 방법으로 트위터에서 추출한 데이터셋을 감성분석하여 71.5%의 정확도를 기록하며 서포트 벡터 머신이나 재귀신경망 등의 기법에 비해 뛰어난 성능을 기록하였다.

### 3.4 Recursive Neural Networks

앞서 살핀 순환신경망 계열의 딥러닝 기법 이외에도 재귀신경망(recursive neural networks)과 같은 딥러닝 기법이 감성분석에 활용되고 있다. 재귀신경망은 입력 데이터에 대해 합성함수(composite function)를 재귀적으로 반복하여 계층적인 구조를 형성하는 트리구조를 생성함으로써 데이터에 대한 전체적인 의미를 파악하는 방법이다.

재귀신경망은 기본적으로 그림 7과 같이 두 개의 자식노드와 한 개의 부모노드가 하나의 단위로 구성된다. 재귀신경망은 이 기본구조 속에서 두 개의 자식노드의 값을 합성함수를 통해 합쳐 부모노드에 전달하는 연산을 재귀적으로 수행하여 루트노드에 이르면 입력벡터의 모든 의미를 대표하

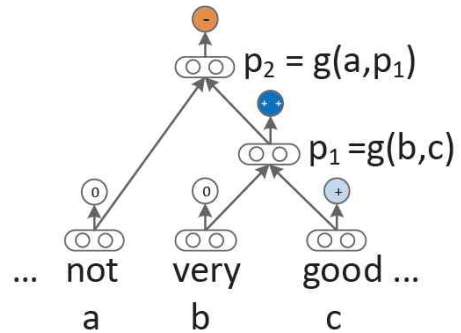


그림 7. 재귀신경망의 기본구조 예시[17]

는 하나의 결과값을 얻을 수 있게 된다.

$$p_1 = f\left(W \begin{bmatrix} b \\ c \end{bmatrix}\right), \quad p_2 = f\left(W \begin{bmatrix} a \\ p_1 \end{bmatrix}\right) \quad (10)$$

재귀신경망의 수행과정을 간략히 설명하면, 먼저 입력 데이터를 트리구조의 말단노드에 입력시킨다. 이후 합성함수를 사용하여 식 (10)과 같은 연산을 루트노드에 이를 때까지 재귀적으로 반복한다. 그림 7은 “not very good”이라는 입력값에 대하여 말단노드부터 “+”, “++”, “-”의 순으로 클래스가 달라져 최종적으로 부정을 의미하는 “-”인 클래스로 결과값이 산출되는 예시를 보여준다. 말단노드의 의미에서부터 합성을 통해 얻은 루트노드의 결과값은 주어진 입력 데이터의 전체적인 의미를 나타내기 때문에 이러한 과정은 입력 텍스트의 전체적인 의미를 파악하고자 하는 감성분석의 수행과정과 유사하다.

재귀신경망 또한 다양한 구조 변형을 통해 성능향상을 이루었다. 기존의 재귀신경망에 행렬연산을 추가해 파라미터 정보를 유지해 나가는 MV-RNN(matrix-vector recursive neural networks)이 제안되어 감성분석에서 향상된 성능을 보였으며[54], MV-RNN의 과도한 연산량의 문제를 해결하고 입력 벡터의 의미를 더욱 정확하게

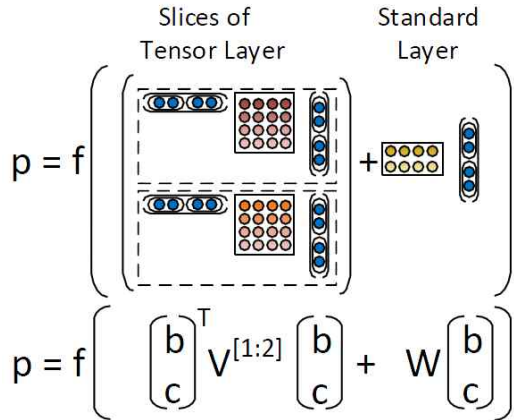


그림 8. Recursive Tensor Neural Networks[17]

합성할 수 있는 단일한 합성함수를 도입한 RNTN(recursive neural tensor network)이 제안되었다[17].

RNTN은 그림 8의 예시와 같이 입력 벡터와 직접적으로 연결된 tensor를 구성하여 행렬 연산을 수행한다. 이 연산의 결과를 모두 합하고 기존의 층에서의 연산결과를 다시 한 번 합해 최종적인 결과값을 구한다. RNTN은 합성함수가 사용될 때마다 매번 다른 행렬연산을 수행한 MV-RNN과는 달리 단어 벡터의 특성을 요약한 tensor를 사용함으로써 연산량을 감소시키는 동시에 입력 벡터의 특징을 보다 많이 반영하여 감성분석 모델의 성능을 향상시켰다.

[17]은 Stanford Sentiment Treebank[55]에 재귀신경망을 적용하여 감성분석을 수행한 대표적인 연구이다. Stanford Sentiment Treebank는 영화 리뷰 문장 11,855개에 대해 구문분석을 수행하여 감성 정보가 부여된 트리 구조를 제공한다. [17]에서는 RNTN을 감성분석에 적용하여 다섯 개의 클래스 분류, 두 개의 극성분류 등의 감성분석에서 나이브 베이저안 분류기나 서포트 벡터 머신과 같은 기존의 기계학습 방법 뿐 아니라 처

음에 제안된 재귀신경망의 기록까지도 뛰어넘는 성능을 보였다. 또한 긍정의 부정, 부정의 부정(negated positive, negated negative) 형태로 되어있는 문장에 대한 감성분석을 수행했으며, 이러한 문장들에 대해 기존의 재귀신경망이 30~40%의 정확도를 기록한 것과 대조적으로 RNTN은 70~80%의 정확도를 보이기도 하였다.

### 3.5 Convolutional Neural Networks

합성곱신경망은 앞서 살핀 바와 같이 주로 이미지 인식분야에서 활용되어 왔으나 품사부착, 구문 분석, 개체명 인식, 의미역 형성, 언어모델, 의미적 관련어 등 자연어처리에도 적용된 바 있다[56]. 합성곱신경망의 가장 큰 장점은 입력값을 그대로 학습하지 않고 특징을 추출하는 과정이 있다는 것이다. 합성곱신경망은 필터를 통해 입력 데이터에 대해서 가장 작은 부분까지 나누면서 특징을 저장하며, 이렇게 추출된 특징들을 이용하여 최종적으로 분류작업을 수행하기 때문에 강력한 성능을 발휘하고 있다[6].

이러한 합성곱신경망은 감성분석 분야에서도 적용되어 다양한 성과를 거두고 있다[18-19]. 합성곱신경망을 통한 감성분석 과정은 기존의 이미지 분류 기법과 유사하다. 먼저 합성곱신경망의 입력 데이터는 이미지 데이터와 유사하게 2차원 행렬로 구성된다. 이 과정에서 한 문장 속의 단어들을 워드 임베딩 하면서 하나의 차원을 늘이는 방식을 사용한다. 입력 데이터가 준비되면 미리 준비한 필터(filter) 혹은 커널(kernel)을 통해 데이터를 읽어 내려가며 입력 데이터와 커널과의 합성곱(convolution)을 통해 특징지도(feature map)을 제작한다. 일반적으로 특징을 추출 할 때에는 가장 강한 자극을 추출(max-pooling)한다. 합성곱신경망은 필터를 통해 입력 데이터를 읽고,

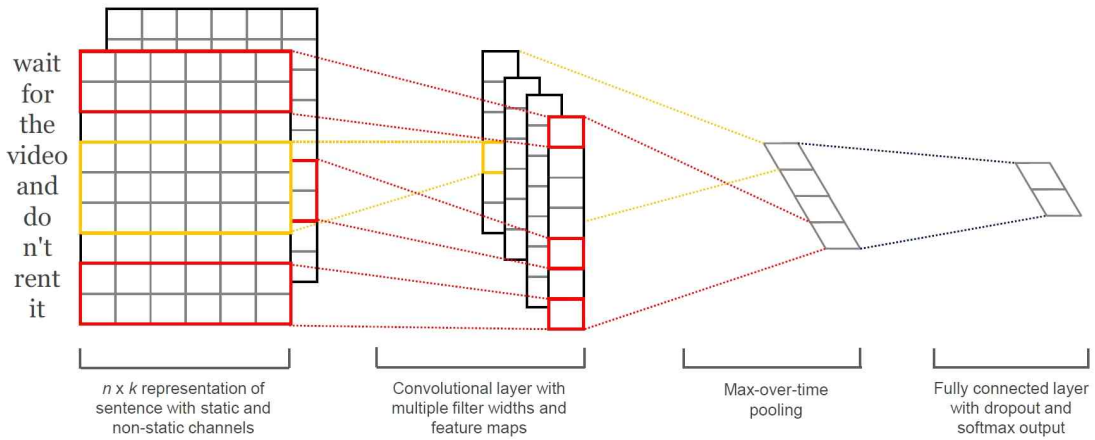


그림 9. 감성분석을 위한 합성곱신경망(Convolution Neural Networks)[18]

특징지도를 만드는 위의 과정을 반복하여 더 이상 필터를 통해 해상도를 줄일 수 없을 때까지 수행한다. 위의 과정의 결과로 마지막 단계에서는 완전 연결된(fully-connected) 신경망을 통해 최종적인 분류결과를 얻는다. 최종 결과값의 계산에는 일반적으로 소프트맥스(softmax) 함수가 사용된다.

그림 9는 [18]에 제시된 감성분석을 수행하는 합성곱신경망 구조를 나타낸다. 그림의 신경망에서는 좌측에서 우측으로 총 네 가지 단계를 거쳐 감성분석이 이루어진다. 첫 번째 단계에서는 학습 과정에서 형태가 유지되는 static, 학습에 따라 변화하는 non-static 두 채널의  $k$  차원  $n$  개 단어 벡터가 입력된다. 두 번째 단계에서는 크기가 다른 다수의 필터를 이용하여 특징지도를 생성하고, 세 번째 단계에서는 각 특징지도에서 최대값을 갖는 하나의 특징을 선택하여 특징 집합을 만든다. 마지막 단계에서는 입력 데이터에서 추출한 특징들을 바탕으로 최종적인 감성값을 얻는다.

[18]은 이미지 영역에서 활발히 사용되던 합성곱신경망을 감성분석에 적용할 수 있는 구조를 제시한 대표적인 연구로서, 총 일곱 개의 데이터

셋과 열여덟 개의 감성분석 모델을 통해 성능 비교를 수행하였으며, 다수의 데이터에서 합성곱신경망이 뛰어난 성능을 나타냄을 보였다. [19]는 길이가 짧은 텍스트 데이터를 감성분석 할 때, 그 안의 문맥정보가 많지 않아 감성분석의 어려운 상황을 극복하고자 입력 데이터의 단위를 조금 더 세분화한 모델을 제시하였다. 이 연구에서는 문자 단위를 기준으로 입력값을 필터링하는 합성곱신경망을 제안하여 영화 리뷰와 트위터 문장을 대상으로 한 감성분석을 수행해 80%가 넘는 뛰어난 정확도를 보였다.

#### 4. 결 론

본 논문에서는 딥러닝 기법을 기반으로 한 감성분석의 연구동향에 대해 살펴보았다. 기존의 감성분석은 어휘기반 자원을 통해 감성사전을 구축하거나, 기분류된 텍스트 데이터를 학습 데이터로 하여 기계학습에서 사용되는 다양한 분류기법들을 적용해 수행되었다. 최근에는 인공지능 분야에서 많은 주목을 받고 있는 딥러닝 기법이 감성분석에 적용되어 의미 있는 성과를 거두고 있다.

딥러닝 기법을 감성분석에 적용하기 위해서는 텍스트 데이터를 숫자의 형태로 변환하는 과정이 필요하며, 그에 따른 다양한 워드임베딩 기법에 대해서 살펴보았다. 특히 고차원 벡터를 통해 단어를 표현하여 단어 사이의 관계를 보다 효율적으로 표현할 수 있는 Word2vec과 같은 언어모델이 딥러닝 기반의 감성분석에 활용되는 양상을 확인하였다.

본 논문에서 다룬 감성분석에 활용되는 딥러닝 기법은 크게 세 가지 계열로 구분할 수 있다. 첫째는 순환신경망 계열의 딥러닝 기법인 순환신경망과 LSTM이고, 둘째는 트리구조와 합성함수를 통해 입력된 텍스트 데이터의 전체적인 의미를 찾는 재귀신경망이다. 셋째는 필터를 통해 입력된 데이터의 특성을 추출하는 특징 추출의 단계를 통해 감성분석의 정확도를 높인 합성곱신경망이다. 세 종류의 딥러닝 기법들은 각각 다양한 형태로 감성분석에 활용되고 있으며, 지속적으로 구조적 변형과 발전을 통해 성능을 향상시키고자 하는 연구가 계속되고 있다.

빅데이터의 성장과 더불어 블로그, 사회연결망 서비스, 상품구매 리뷰, 뉴스 댓글 등 사회 각 분야에서 생산되는 텍스트 데이터의 분석에 대한 사회적 요구는 지속적으로 커질 것으로 예상된다. 이런 사회적 요구를 바탕으로 향후의 감성분석 연구는 계속 활발하게 진행될 것이며, 딥러닝 모델의 발전에 따라 기존의 딥러닝 기반의 감성분석 방법을 최적화 시키거나 새로운 딥러닝 기법을 감성분석에 접목시키는 연구가 지속될 것으로 생각된다. 다른 한 편으로 감독학습 외에 비감독 학습, 혹은 반감독 학습의 방법론들을 감성분석에 적용하는 연구도 활발히 수행될 것으로 예상된다.

## 참 고 문 헌

- [1] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning.", *Nature* 521.7553, (2015): 436-444.
- [2] Bengio, Yoshua, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives.", *IEEE transactions on pattern analysis and machine intelligence* 35.8, (2013): 1798-1828.
- [3] Schmidhuber, Jürgen. "Deep learning in neural networks: An overview.", *Neural Networks* 61, (2015): 85-117.
- [4] Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search.", *Nature* 529.7587, (2016): 484-489.
- [5] Taigman, Yaniv, et al. "Deepface: Closing the gap to human-level performance in face verification.", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2014).
- [6] Russakovsky, Olga, et al. "Imagenet large scale visual recognition challenge.", *International Journal of Computer Vision* 115.3, (2015): 211-252.
- [7] Hinton, Geoffrey, et al. "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups.", *IEEE Signal Processing Magazine* 29.6, (2012): 82-97.
- [8] Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks.", *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, (2013).
- [9] Weston, Jason, et al. "Towards ai-complete question answering: A set of prerequisite toy tasks.", *arXiv preprint arXiv:1502.05698*, (2015).
- [10] Bahdanau, Dzmitry, Kyunghyun Cho, and

- Yoshua Bengio. "Neural machine translation by jointly learning to align and translate.", arXiv preprint arXiv:1409.0473, (2014).
- [11] Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation.", arXiv preprint arXiv:1406.1078, (2014).
- [12] Tarasov, D. S. "Deep recurrent neural networks for multiple language aspect-based sentiment analysis of user reviews.", *Proceedings of International Conference of Computational Linguistics and Intellectual Technologies Dialog-2015*. Vol. 2, (2015).
- [13] Irsoy, Ozan, and Claire Cardie. "Opinion Mining with Deep Recurrent Neural Networks.", *EMNLP*, (2014).
- [14] Tai, Kai Sheng, Richard Socher, and Christopher D. Manning. "Improved semantic representations from tree-structured long short-term memory networks.", arXiv preprint arXiv:1503.00075, (2015).
- [15] Tang, Duyu, et al. "Target-Dependent Sentiment Classification with Long Short Term Memory.", arXiv preprint arXiv:1512.01100, (2015).
- [16] Tang, Duyu, Bing Qin, and Ting Liu. "Document modeling with gated recurrent neural network for sentiment classification.", *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, (2015).
- [17] Socher, Richard, et al. "Recursive deep models for semantic compositionality over a sentiment treebank.", *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*. Vol. 1631, (2013).
- [18] Kim, Yoon. "Convolutional neural networks for sentence classification.", arXiv preprint arXiv:1408.5882, (2014).
- [19] dos Santos, Cícero Nogueira, and Maira Gatti. "Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts.", *COLING*, (2014).
- [20] Zhou, Shusen, Qingcai Chen, and Xiaolong Wang. "Active deep learning method for semi-supervised sentiment classification.", *Neurocomputing* 120, (2013): 536-546.
- [21] Lai, Siwei, et al. "Recurrent Convolutional Neural Networks for Text Classification.", *AAAI*, (2015).
- [22] Liu, Bing. "Sentiment analysis and opinion mining.", *Synthesis lectures on human language technologies* 5.1, (2012): 1-167.
- [23] Wang, Hao, et al. "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle.", *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, (2012).
- [24] Nakov, Preslav, et al. "SemEval-2016 task 4: Sentiment analysis in Twitter.", *Proceedings of the 10th international workshop on semantic evaluation (SemEval 2016)*, San Diego, US (forthcoming), (2016).
- [25] Ding, Xiaowen, Bing Liu, and Philip S. Yu. "A holistic lexicon-based approach to opinion mining.", *Proceedings of the 2008 international conference on web search and data mining*. ACM, (2008).
- [26] Taboada, Maite, et al. "Lexicon-based methods for sentiment analysis.", *Computational linguistics* 37.2, (2011): 267-307.
- [27] Tan, Songbo, et al. "Adapting naive bayes to domain adaptation for sentiment analysis.", *European Conference on Information Retrieval*. Springer Berlin Heidelberg, (2009).
- [28] Mullen, Tony, and Nigel Collier. "Sentiment Analysis using Support Vector Machines with Diverse Information Sources.", *EMNLP*. Vol. 4, (2004).
- [29] Dang, Shilpa, and Peerzada Hamid Ahmad. "Text Mining: Techniques and its

- Application.", *International Journal of Engineering & Technology Innovations*, ISSN (Online), (2014): 2348-0866.
- [30] Dang, Dr Shilpa, and Peerzada Hamid Ahmad. "A Review of Text Mining Techniques Associated with Various Application Areas.", *International Journal of Science and Research (IJSR)* 4.2, (2015): 2461-2466.
- [31] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey.", *Ain Shams Engineering Journal* 5.4, (2014): 1093-1113.
- [32] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.", *LREC*. Vol. 10, (2010).
- [33] An, Jungkook, and Hee-Woong Kim. "Building a Korean Sentiment Lexicon Using Collective Intelligence.", *Journal of Intelligence and Information Systems* 21.2, (2015): 49-67.
- [34] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques.", *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*. Association for Computational Linguistics, (2002).
- [35] Blitzer, John, Mark Dredze, and Fernando Pereira. "Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification.", *ACL*. Vol. 7, (2007).
- [36] Glorot, Xavier, Antoine Bordes, and Yoshua Bengio. "Domain adaptation for large-scale sentiment classification: A deep learning approach.", *Proceedings of the 28th International Conference on Machine Learning(ICML-11)*, (2011).
- [37] Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning internal representations by error propagation.", *No. ICS-8506. CALIFORNIA UNIV SAN DIEGO LA JOLLA INST FOR COGNITIVE SCIENCE*, (1985).
- [38] Hawkins, Douglas M. "The problem of overfitting.", *Journal of chemical information and computer sciences* 44.1, (2004): 1-12.
- [39] LeCun, Yann, et al. "Gradient-based learning applied to document recognition.", *Proceedings of the IEEE* 86.11, (1998): 2278-2324.
- [40] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks.", *Advances in neural information processing systems*, (2012).
- [41] Hinton, Geoffrey E., Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets.", *Neural computation* 18.7, (2006): 1527-1554.
- [42] Hinton, Geoffrey E. "A practical guide to training restricted boltzmann machines.", *Neural Networks: Tricks of the Trade*. Springer Berlin Heidelberg, (2012): 599-619.
- [43] Liu, Ping, et al. "Facial expression recognition via a boosted deep belief network.", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2014).
- [44] Deoras, Anoop, and Ruhi Sarikaya. "Deep belief network based semantic taggers for spoken language understanding.", *INTERSPEECH*, (2013).
- [45] Elman, Jeffrey L. "Finding structure in time.", *Cognitive science* 14.2, (1990): 179-211.
- [46] Mikolov, Tomas, et al. "Recurrent neural network based language model.", *Interspeech* Vol. 2, (2010).
- [47] Bengio, Yoshua, et al. "A neural probabilistic language model.", *journal of machine learning research* 3.Feb, (2003): 1137-1155.
- [48] Mikolov, T., and J. Dean. "Distributed repre-

sentations of words and phrases and their compositionality.", *Advances in neural information processing systems*, (2013).

[49] Mikolov, Tomas, et al. "Efficient estimation of word representations in vector space.", arXiv preprint arXiv:1301.3781, (2013).

[50] Werbos, Paul J. "Backpropagation through time: what it does and how to do it.", *Proceedings of the IEEE* 78.10, (1990): 1550-1560.

[51] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory.", *Neural computation* 9.8, (1997): 1735-1780.

[52] Graves, Alex. "Long Short-Term Memory.", *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer Berlin Heidelberg, (2012): 37-45.

[53] Greff, Klaus, et al. "LSTM: A search space odyssey.", arXiv preprint arXiv:1503.04069, (2015).

[54] Socher, Richard, et al. "Semantic compositionality through recursive matrix-vector spaces.", *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, (2012).

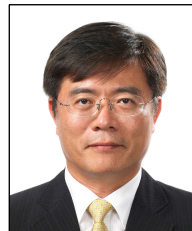
[55]<http://nlp.stanford.edu/sentiment/treebank.html>

[56] Collobert, Ronan, and Jason Weston. "A unified architecture for natural language processing: Deep neural networks with multitask learning.", *Proceedings of the 25th international conference on Machine learning*. ACM, (2008).



서 상 현

- 2016년: 동국대학교 국어국문학과, 문학사
  - 2016년~현재: 동국대학교 컴퓨터공학과, 석사과정
  - 관심분야: 자연어처리, 기계학습, 딥러닝응용
- 
- 



김 준 태

- 1986년: 서울대학교 제어계측공학과, 공학사
  - 1990년: 미국 University of Southern California, 전기공학, 공학석사
  - 1993년: 미국 University of Southern California, 컴퓨터공학, 공학박사
  - 1995년~현재: 동국대학교 컴퓨터공학과 교수
  - 관심분야: 인공지능, 기계학습, 데이터마이닝
- 
-