
저자 (Authors)	박천음, 이창기 Cheoneum Park, Changki Lee
출처 (Source)	정보과학회 컴퓨팅의 실제 논문지 25(11) , 2019.11, 552-558(7 pages) KIISE Transactions on Computing Practices 25(11) , 2019.11, 552-558(7 pages)
발행처 (Publisher)	한국정보과학회 The Korean Institute of Information Scientists and Engineers
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE09233106
APA Style	박천음, 이창기 (2019). BERT 기반 Variational Inference와 RNN을 이용한 한국어 영화평 감성 분석. 정보과학회 컴퓨팅의 실제 논문지, 25(11), 552-558
이용정보 (Accessed)	연세대학교 165.***.14.104 2020/09/19 11:28 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

BERT 기반 Variational Inference와 RNN을 이용한 한국어 영화평 감성 분석 (Sentimental Analysis of Korean Movie Review using Variational Inference and RNN based on BERT)

박 천 음 [†]
(Cheoneum Park)

이 창 기 ^{††}
(Changki Lee)

요 약 최근 자연어처리 분야에서 많은 성능 향상을 보이고 있는 BERT는 양방향성을 가진 트랜스포머(transformer)를 기반으로 한 모델이다. BERT는 OOV (Out Of Vocabulary) 문제를 해결하기 위하여 BPE (Byte Pair Encoding)를 적용하며, 이를 기반으로 언어 모델을 사전 학습하고 출력 층(layer)을 추가하여 자연어처리 태스크를 fine-tuning한다. 감성 분석은 주어진 문장에 대한 잠재적 의미를 분석하고 분류하는 문제이다. 본 논문에서는 감성 분석에 BERT로부터 생성되는 토큰 표현을 이용하기 위하여 대용량 한국어 코퍼스로 언어 모델을 학습한 BERT 모델을 사용한다. 또한 문맥 정보를 인코딩하는 RNN을 BERT 함께 사용하는 방법과, RNN으로 인코딩한 hidden state에 variational inference를 이용하여 감성 분석을 수행하는 방법을 제안한다.

키워드: 감성 분석, 영화평, BERT, Transformer, variational inference, RNN

Abstract BERT is a model based on a bidirectional transformer and has demonstrated immense improvement in performance in the field of natural language processing. BERT applies BPE (Byte Pair Encoding) to solve OOV (Out Of Vocabulary) problem, and pre-trains language model based on the BPE and fine-tunes natural language processing task by adding an output layer. Sentimental analysis is a task of analyzing and classifying the potential meaning of a given sentence. In the present work, we have employed the BERT model which trains the language model with large capacity Korean Corpus to use the token representation generated from BERT for sentimental analysis. In addition, we propose a method of using BERT along with RNN, which encodes context information and a method of performing sentimental analysis using variational inference in hidden state encoded by RNNv.

Keywords: sentimental analysis, movie review, BERT, Transformer, variational inference, RNN

- 이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2013-2-00131, 휴먼 지식증강 서비스 위한 지능진화형 Wise QA 플랫폼 기술 개발)
- 이 논문은 제45회 한국소프트웨어종합학술대회에서 'RNN 기반 Variational Inference를 이용한 한국어 영화평 감성 분석'의 제목으로 발표된 논문을 확장한 것임

[†] 학생회원 : 강원대학교 컴퓨터과학
parkce3@gmail.com

^{††} 종신회원 : 강원대학교 컴퓨터과학 교수(Kangwon Nat'l Univ.)
leeck@kangwon.ac.kr
(Corresponding author임)

논문접수 : 2019년 7월 11일

(Received 11 July 2019)

논문수정 : 2019년 8월 21일

(Revised 21 August 2019)

심사완료 : 2019년 9월 5일

(Accepted 5 September 2019)

Copyright©2019 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.
정보과학회 컴퓨팅의 실제 논문지 제25권 제11호(2019. 11)

1. 서론

사람이 생각하는 사고와 그에 따른 의견, 어떤 대상에 대하여 느끼는 감정 등을 감성(sentiment)이라 한다. 감성 분석(sentimental analysis)은 특정 대상에 대한 의견을 수집하고 감성에 따라 분류하는 과정으로, 웹 상에서 상품평이나 리뷰 등을 분석하는 오피니언 마이닝(opinion mining)에 응용된다. 이러한 과정을 통하여 사용자는 여론의 생각을 파악할 수 있고, 이에 대한 방법을 마련할 수 있다. 그러나 자연어에 담긴 사람의 주관 을 파악하기는 어려운 일이다. 비슷해 보이는 문장이라도 작은 변화로 긍정, 부정 의견이 갈릴 수 있다. 예를 들어, 문장 “정말 재밌네”와 같이 화용에 따라 달라지기도 하며, 문장 전체를 이해해야 해결할 수 있다[1-5].

기존 감성 분석 연구는 사전 구축을 통하여 긍/부정 가중치를 계산하는 규칙 기반 모델이나, 문장에 긍/부정이 포함되어 있는지 판별하는 확률 모델이 진행되었다. 딥러닝 모델인 단어 임베딩과 CNN(Convolutional Neural Network), RNN(Recurrent Neural Network) 등을 이용한 감성 분석 모델이 연구되었으며, 좋은 결과들을 보이고 있다[1-3,7-11]. [3,8,9]는 이미지 인식에서 뛰어난 성능을 보이는 CNN을 [7]의 CNN과 같이 확장하여 한국어에 적용하였다. [1,10,11]은 Naver 영화 리뷰 데이터인 NSMC(Naver Sentiment Movie Corpus)[12]에 대하여 RNN의 vanishing gradient problem을 해결한 LSTM(Long Short-Term Memory)[13]나 GRU(Gated Recurrent Units)[14]를 여러 층 쌓은 모델을 이용하였다.

BERT(Bidirectional Encoder Representation from Transformers)[15]는 다양한 자연어처리 분야에 적용되어 많은 성능 향상을 보이고 있으며, 대용량 코퍼스 언어 모델(language model)로 학습한 모델이다. 사전 학습된 BERT 모델에 출력 층(output layer)을 추가한 후에 fine-tuning 하는 방법으로 자연어처리 태스크에 적용된다. BERT는 양방향성을 가진 트랜스포머(transformer)[16]를 기반으로 하며, 각 블록(block)의 셀프 어텐션 매커니즘(self-attention mechanism)[16]으로 문맥 전체를 확인하여 언어 모델을 학습한다. 문장 내에서 임의의 단어를 마스킹(masking)하고 이를 예측하는 masked language modeling(masked LM)을 학습하며, 주어진 두 문장이 연결된 문장인지 예측하는 다음 문장 예측(next sentence prediction)을 함께 학습한다.

BERT는 출력 층(output layer)을 추가하여 다른 자연어처리 태스크에 적용한다[15]. BERT(large)[15]는 영화 리뷰 데이터 셋인 MR에 대한 성능은 89.3%, 감성 분석 데이터 셋인 SST-2에 대한 성능은 94.9%로 기존

ELMo[17]나 GPT[18]를 이용한 경우보다 4-7% 더욱 향상된 성능을 보였다. [19]에서는 한국어에 대한 BERT를 사전학습하여 의존구문분석에 적용한 연구이며, BERT의 임베딩을 기반으로 LSTM으로 인코딩한 후 biaffine 어텐션 매커니즘으로 의존구문분석을 수행하였다.

VI(Variational Inference)는 잠재 변수 z 가 manifold learning과 같이 고차원 안에 포함된 semantic space에 있다고 가정하고, 학습을 통하여 저차원 공간으로 변환시켜 의미 있는 자질을 추출한다. 이를 위하여 입력 벡터에 기반한 잠재 변수를 사전 확률 분포(가우시안 분포)로 변환한다. Variational inference network는 그라디언트 역전파를 수행하고 모델을 학습하기 위하여 reparameterization trick을 이용한다[20-22].

본 논문에서는 감성 분석을 RNN을 이용한 sequence-to-sequence(many-to-one) 문제로 정의하고, 양방향 인코더를 기반으로 대용량 코퍼스로 사전 학습한 BERT를 사용하며, 고차원의 의미 있는 자질을 저차원으로 변환하기 위하여 variational inference를 적용한 모델을 제안한다.

2. Variational Auto-Encoder

잠재 변수를 모델링하는 방법은 자연어처리 태스크에 다양하게 사용된다. 감성 분석과 같이 문장을 분류하는 문제는 $p(y|x)$ 와 같이 주어진 문장의 문맥 정보를 확인하여 출력 결과의 확률을 구하는 조건부 확률 문제로 정의된다. 본 논문에서는 문장의 의미 정보를 잠재 변수로 모델링하고 이를 학습하기 위하여 Variational Auto-Encoder(VAE)와 유사하게 variational inference를 적용한다.

Variational Auto-Encoder는 잠재적 의미 공간의 잠재 변수 z 를 추정하는 모델로써, 그림 1의 VAE[21]과 같이 주어진 입력 벡터 x 를 확률 분포에 따른 잠재 변수 z 로 샘플링하여 압축하고 다시 z 를 이용하여 디코딩을 수행한다. 본 논문에서는 입력 문장의 인코더 벡터를 사전 확률 분포인 가우시안 분포를 따르는 잠재 변수 z 로 생

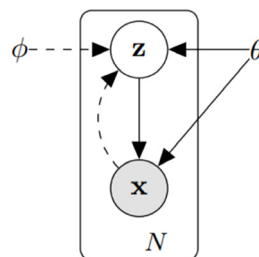


그림 1 VAE 모델
Fig. 1 VAE model

성하는 모델인 variational inference network를 사용하며, 그에 대한 조건부 확률 $p(y|z, x)$ 를 정의한다. Variational inference network는 reparameterization 방법 [2,3]으로 잠재 변수 z 와 평균 μ 와 분산 σ 를 함께 학습하며, 모델의 손실 함수(loss function)는 다음과 같다.

$$\mathcal{L} = -KL(q_\phi(z|x)||p(z)) + \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(y|z)] \quad (1)$$

여기서 $q_\phi(z|x)$ 는 입력 벡터 x 를 잠재 변수로 압축하는 인코더이며, **Feed-Forward Neural Network (FFNN)**로 구성된다. $p_\theta(y|z)$ 는 잠재 변수 z 로부터 출력 결과 y 를 만드는 디코더이다. $p(z)$ 는 사전 확률 분포인 가우시안 분포이며 평균 0, 분산 1인 정규분포를 사용한다. 왼쪽 KL항은 Kullback-Leibler 발산으로 입력 문장에 대한 벡터 x 가 가우시안 분포에 따라 z 를 생성하도록 제한하는 역할을 한다. 오른쪽 항은 cross entropy로, 샘플링된 z 에서 문장 분류 출력 결과 y 를 생성할 수 있도록 한다.

3. BERT: Bidirectional Encoder Representation from Transformer

본 논문에서는 BERT의 입력을 위하여 **입력 형태소**에 BPE를 적용하여 **토큰화(tokenization)**하며, 그림 2의 입력(Input)과 같다. BPE를 적용했을 때 형태소 단위를 구분하기 위하여 각 형태소의 마지막 토큰에는 “_”을 붙인다. 모든 입력열의 첫 번째 토큰으로는 [CLS]가 주어지며, 분류 문제를 해결할 때 해당 토큰의 벡터를 이용한다. BERT를 언어 모델로 사전 학습 할 때는 구분자 [SEP]를 사용하여 두 개의 열(sequence)을 연결하여 하나의 입력으로 만든다. 본 논문에서는 하나의 문장을 입력으로 하며, “[CLS] + 입력 문장 + [SEP]”와 같은 구조를 만들고 입력된 각 토큰들은 토큰 임베딩(token embedding)을 적용하여 벡터화한다. 세그먼트 임베딩(segment embedding)은 하나의 문장에 대한 벡터를 만들고, 토큰 길이 최대 512까지의 포지션 임베딩(position embedding)을 만들며, 위의 임베딩 벡터들에 대한 element-wise sum을 수행하여 BERT의 입력 벡터를 만든다.

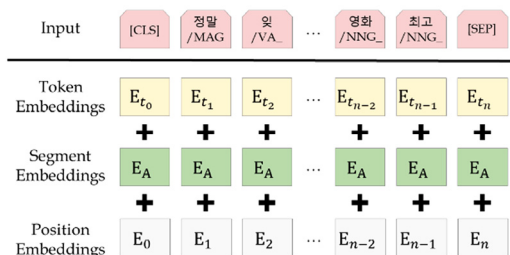


그림 2 감성 분석에 대한 BERT 입력 구조
Fig. 2 BERT input structure for sentimental analysis

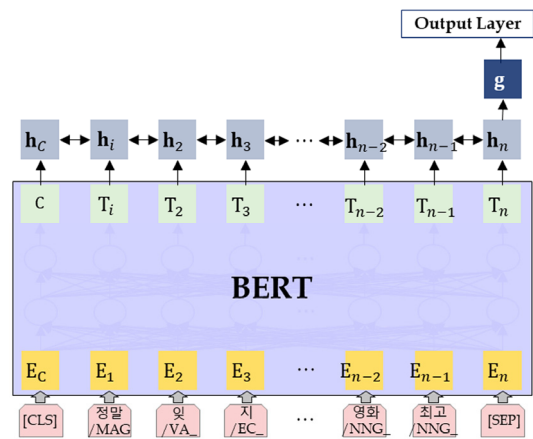


그림 3 BERT 모델 기반 RNN을 이용한 감성분석 모델
Fig. 3 Sentimental analysis model using RNN based on BERT

3.1 BERT 기반 RNN 모델

본 논문에서는 **감성 분석** 문제를 many-to-one 모델로 정의하며, BERT 모델을 기반으로 하고 RNN과 variational inference를 이용하여 **딥 러닝** 모델을 정의한다. 학습 데이터는 입력열 $X = \{x_1, x_2, \dots, x_n\}$ 와 감성 분석의 출력 $Y \in \{0, 1\}$ 과 같이 구성된다.

감성 분석을 위한 BERT 기반 RNN은 그림 3과 같으며, bidirectional RNN (BiRNN)을 사용하고, RNN type으로는 GRU, LSTM, SRU (simple recurrent unit) [23] 등을 사용한다.

입력된 문장의 각 형태소는 BPE로 토큰화하여 토큰 임베딩을 얻는다. BERT의 입력 표현 E_i 를 만들기 위하여 세그먼트 임베딩, 포지션 임베딩을 더한다. 그 후, 입력 표현 E_i 는 사전 학습된 트랜스포머를 거쳐 BERT의 단어 표현 T_i 를 생성하고, 이를 기반으로 BiRNN layer에서 RNN 인코딩을 수행한다. BERT의 단어 표현 T_i 에 대한 BiRNN의 수식은 다음과 같다.

$$\vec{h}_i = RNN(T_i, \vec{h}_{i-1}) \quad (1)$$

$$\vec{h}_i = RNN(T_i, \vec{h}_{i+1}) \quad (2)$$

$$h_i = [\vec{h}_i; \vec{h}_i] \quad (3)$$

\vec{h}_i 는 forward RNN의 히든 스테이트이고, \vec{h}_i 는 backward RNN의 히든 스테이트이며, 이 둘을 연결(concatenation)하여 h_i 를 만든다. 인코딩된 h_i 에서 마지막 히든 스테이트를 추출하고, 비선형연산(non-linear layer)을 수행하여 히든 스테이트 g 를 만든다. 이때 사용한 **활성 함수(activation function)**는 GELU [24]이며, 식은 다음과 같다.

$$\mathbf{g} = \text{GELU}(\text{FFNN}(\mathbf{h}_n)) \quad (4)$$

그 이후, 히든 스테이트 \mathbf{g} 는 Output layer에서 FFNN (Feed Forward Neural Network)을 거쳐 감성 분석 결과를 출력한다.

$$\mathbf{y} = \mathbf{W}_y \mathbf{g} + \mathbf{b}_y \quad (5)$$

3.2 BERT 기반 RNN을 이용한 Variational Inference 모델

본 논문에서는 BERT 모델로부터 생성한 각 입력 토큰에 대한 히든 스테이트를 RNN으로 인코딩하고, 인코더의 마지막 히든 스테이트를 variational inference network의 입력으로 하여 variational inference를 수행한다. 그 후, variational inference network에서 샘플링된 벡터는 FFNN을 거쳐 감성 분석 결과를 출력한다.

BERT 기반 RNN을 이용한 variational inference 모델은 그림 4와 같으며, BERT 모델과 RNN까지 BERT 기반 RNN 모델과 동일한 식 (1)~(4)를 따르며, 비선형 연산으로 생성된 히든 스테이트 \mathbf{g} 는 variational inference network의 입력으로 주어진다.

$$\mathbf{h}_z = \text{ReLU}(\mathbf{W}_z \mathbf{g} + \mathbf{b}_z) \quad (6)$$

$$\boldsymbol{\mu} = l_{\mu}(\mathbf{h}_z), \quad \boldsymbol{\sigma} = l_{\sigma}(\mathbf{h}_z) \quad (7)$$

$$\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon} \quad (8)$$

$$\tilde{\mathbf{h}} = \sigma(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{z} + \mathbf{b}_1) + \mathbf{b}_2) \quad (9)$$

입력된 벡터 \mathbf{g} 는 ReLU를 이용한 비선형 연산이 수행되어 벡터 \mathbf{h}_z 를 만든다. 샘플링 $\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{h})$ 는 reparameterization trick과 함께 식 (6, 7)을 기반으로 계산된다. 식 (8)에서는 가우시안 분포의 파라미터인 $\boldsymbol{\mu}$ 와 $\boldsymbol{\sigma}$ 벡터를 만들기 위하여 \mathbf{h}_z 에 FFNN $l(\cdot)$ 을 적용한다. 식 (9)에서는 정규분포 변수 $\boldsymbol{\epsilon}$ 를 정의하여 $\boldsymbol{\sigma}$ 와 element-wise multiplication을 수행하고, 이를 $\boldsymbol{\mu}$ 와 더하여 샘플링된 \mathbf{z} 를 생성한다. 마지막으로, \mathbf{z} 는 ReLU와 σ (sigmoid)로 비선형 연산을 수행하여 variational inference로 인코딩과 디코딩된 $\tilde{\mathbf{h}}$ 를 생성한다.

Variational inference로 생성한 $\tilde{\mathbf{h}}$ 는 Output Layer에서 선형 연산(FFNN)을 수행하고 감성 분석 결과를 예측한다.

$$\mathbf{y} = \mathbf{W}_y \tilde{\mathbf{h}} + \mathbf{b}_y$$

4. 실험

4.1 한국어 BERT 사전 학습

본 논문에서는 ETRI에서 사전 학습한 BERT 모델을 이용하였다. 사전 학습을 위한 데이터는 웹에서 수집한 뉴스 및 위키백과 데이터이며, 총 23.5 기가바이트이다. 형태소 분석을 위하여 AIOpen[25]에서 공개한 ETRI 언

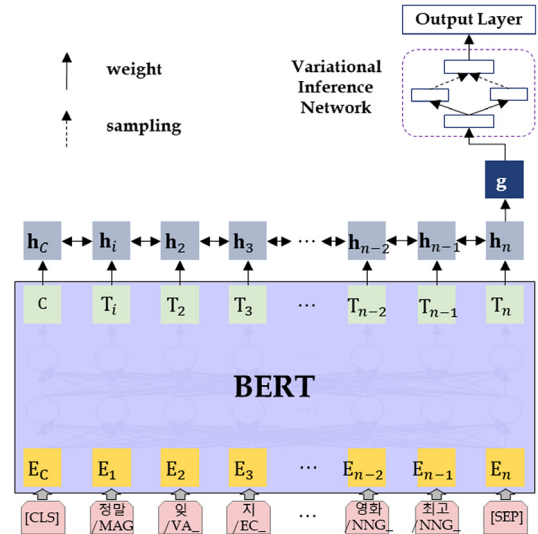


그림 4 BERT 모델 기반 RNN을 이용한 variational inference 감성 분석 모델

Fig. 4 Sentimental analysis model with variational inference using RNN based on BERT

어분석기를 사용하였다. ETRI에서 사전 학습할 때 사용한 하이퍼 파라미터는 구글에서 공개한 BERT-base (트랜스포머 블록 수: 12, 히든 레이어 차원 수: 768, 어텐션 헤드 수: 12) 옵션을 따른다. 각 히든 레이어의 활성화 함수는 GELU를 사용하고, 히든 레이어의 드랍아웃(dropout)은 0.1, 언어 모델 학습을 위한 최대 문장 길이는 512로 설정하였다.

4.2 영화평 감성 분석 실험 설정

본 논문에서는 영화평 감성 분석의 실험 데이터로 NSMC와 다음카카오 영화평 데이터[26]를 이용하였다. NSMC의 학습 데이터는 15만개, 평가 데이터는 5만개로 구성되며, 다음카카오 영화평 데이터는 학습 셋(train set) 109,066개, 개발 셋(dev set) 13,634개, 평가 셋(test set) 13,634개로 구성된다.

하이퍼 파라미터 설정은 다음과 같다. RNN의 히든 스테이트 차원 수는 768로 설정하였고, variational inference network의 차원 수 512, 샘플링은 128로 설정하였다. RNN의 활성화함수는 tanh를 사용하였고, FFNN의 활성화함수는 GeLU를 사용하였다. 모델에서 사용한 모든 드랍아웃(dropout)은 0.1로 설정하였다. 학습을 위하여 Adam을 사용하고, warmup linear를 적용하였으며, 학습율은 $5e-5$, 가중치 감소는 $1e-2$ 로 설정하였다. Variational inference를 사용할 때 kld (kullback-leibler divergence) weight는 $1e-3$ 의 값으로 설정하였다. NSMC 데이터를 사용할 경우, 문장 최대 길이는 200, 배치 크기 20으로 설

정하였고, 다음카카오 영화평 데이터를 사용할 경우, 문장 최대 길이는 400, 배치 크기는 16으로 설정하였다.

실험을 수행한 환경은 다음과 같다. Intel Core i7-4790 CPU 3.60 GHz와 32 GB RAM, GTX 1080Ti를 사용하였으며, fine-tuning을 위해 사용한 코드는 PyTorch에서 수행되는 huggingface의 pytorch-pretrained-BERT [27]를 기반으로 한다.

4.3 영화평 감성 분석 실험

표 1은 NSMC 데이터 셋에 대하여 본 논문에서 제안한 모델들과 기존 감성 분석 모델들에 대한 비교 실험을 보인다. 표 1은 NSMC 데이터 셋으로 학습하고 NSMC 데이터 셋으로 평가한 결과이다. 본 논문에서 제안한 방법들 중 BERT+FFNN은 BERT 모델을 거쳐 생성된 히든 스테이트에서 첫번째 위치의 벡터를 추출하고 FFNN을 적용하여 감성 분석을 수행한 방법이며, 이는 [15]에서 분류 문제를 해결하는 BERT와 같은 모델이다. BERT+FFNN+VI는 BERT+FFNN에서 만든 히든 스테이트를 기반으로 variational inference를 수행한 모델이다. BERT+RNN은 본 논문에서 제안한 방법 중 하나로 BERT 모델로 단어 표현을 얻고 RNN으로 인코딩하여 감성 분석을 수행하는 모델이며, 이때 rnn type은 SRU이다. BERT+RNN+VI도 본 논문에서 제안한 모델 중 하나로, BERT+RNN으로부터 생성된 히든 스테이트를 variational inference network로 넘겨 샘플링을 수행하고 감성 분석을 판별하는 모델이다. 이 모델도 SRU를 사용한다.

실험 결과, 본 논문에서 제안한 모델 중 BERT+RNN이 90.51%의 정확도(accuracy)로 본 논문에서 제안한 다른 모델들에 비하여 가장 좋은 성능을 보였다. BERT에 RNN을 적용한 경우, BERT에 기반한 히든 스테이트를 인코딩하여 생성된 문맥 정보를 이용하기 때문에 성능 향상에 도움이 된 것으로 생각된다.

표 1 NSMC에 대한 영화평 감성 분석 비교 실험
Table 1 Comparing sentimental analysis of movie review for NSMC (% , test)

Model (NSMC)	Acc
Multi-channel CNN [9] (baseline)	86.27
CNN + bi-gram (CNN) [3]	86.59
Joint model between VAE and CNN [8]	86.74
Parallel stacked BiLSTM [11]	87.07
Parallel stacked BiLSTM w/ ensemble [11]	87.75
RNN-based VI Model [28]	88.38
RNN Model 2 + bi-gram (RNN) + ELMo [10]	89.24
BERT + FFNN (ours)	90.06
BERT + FFNN + VI (ours)	90.34
BERT + RNN (ours)	90.51
BERT + RNN + VI (ours)	90.26

기존 연구들과 비교한 경우, 베이스라인(baseline) 모델인 Multi-channel CNN을 이용한 [9]와 CNN으로 입력열과 bi-gram을 인코딩한 [3], toward controlled generation of text 모델과 유사하게 문장을 생성하고 VAE와 CNN을 이용하여 감성 분석을 수행한 [8]에 비하여 각각 4.24%p, 3.92%p, 3.77%p 더 좋은 성능을 보였다. 또한 본 논문에서 제안한 방법은 Parallel stacked BiLSTM [11] 모델과 이를 ensemble한 방법보다 각각 3.44%p, 2.76%p 더 좋은 성능을 보였다. 본 논문의 선행 연구 방법인 RNN-based VI Model [28]은 88.38%로 본 논문에서 제안한 방법이 2.13%p 향상된 성능을 보였고, RNN 모델을 기반으로 bi-gram을 RNN으로 인코딩하고 ELMo를 사용한 [10] 보다 1.27%p 더 좋은 성능을 보였다.

표 2는 다음카카오 데이터에 대하여 본 논문에서 제안한 모델과 선행 연구와의 비교 실험을 수행한 것이다. 표 2는 다음 카카오 데이터 셋으로 학습하고 다음 카카오 데이터 셋으로 평가한 결과이다.

실험 결과, 본 논문에서 제안한 모델들 모두 선행 연구 CNN + Bi-gram (CNN) (baseline)보다 더 좋은 성능을 보였다. BERT+FFNN은 개발 셋에서 92.64%, 평가 셋에서 92.67%의 성능을 보였고, 여기에 variational inference를 이용한 경우, 개발 셋은 0.15%p 하락했지만, 평가 셋에서 92.71%로 0.04%p 더 좋은 성능을 보였다. BERT+RNN 모델의 경우에는 다른 모델에 비하여 개발 셋 92.85%, 평가 셋 93.00%로 가장 좋은 성능을 보였으며, 이때 사용한 rnn type은 LSTM이다. BERT+RNN+variational inference 모델의 경우에는 BERT+RNN 보다 다소 하락한 (개발 셋) 92.83%, (평가 셋) 92.91%를 보였으며, 이때 사용한 rnn type은 GRU이다.

표 3은 NSMC와 다음카카오 데이터 셋에 대하여 서로 교차하여 평가한 실험을 보인다. NSMC→DaumKakao는 NSMC 데이터 셋에 대하여 표 1에서 가장 좋은 성능을 보인 모델인 BERT + RNN으로 학습하고, 다음카카오 데이터 셋으로 평가한 결과이다. DaumKakao→NSMC는 다음카카오 데이터 셋에 대하여 표 2에서 가장 좋은 성능을 보인 모델인 BERT + RNN으로 학습하고, NSMC 데이터 셋으로 평가한 결과이다.

표 2 다음카카오에 대한 영화평 감성 분석 비교 실험(%)
Table 2 Comparing sentimental analysis of movie review for DaumKakao (%)

Model (DaumKakao)	Acc (dev)	Acc (te)
CNN + Bi-gram (CNN) [3] (baseline)	91.24	91.04
BERT + FFNN (ours)	92.64	92.67
BERT + FFNN + VI (ours)	92.49	92.71
BERT + RNN (ours)	92.85	93.00
BERT + RNN + VI (ours)	92.83	92.91

표 3 NSMC와 다음카카오 데이터 셋 교차 평가(%)
Table 3 Cross evaluation for NSMC and DaumKakao (%)

Model (BERT + RNN)	Acc (te)
NSMC→DaumKakao	91.83
DaumKakao→NSMC	87.02

실험 결과, NSMC→DaumKakao의 경우에는 91.83%의 정확도로 다음카카오 데이터 셋으로 학습하고 평가한 경우(BERT + RNN: 93.00%)에 비하여 1.17%p 낮은 성능을 보였지만, 기존 방법인 CNN+Bi-gram (CNN)에 비하여 0.79%p 더 좋은 성능을 보였다. DaumKakao→NSMC의 경우, 87.02%로 NCSM 데이터 셋으로 학습하고 평가한 경우(BERT + RNN: 90.51%)에 비하여 3.49%p 더 낮은 성능을 보였지만, 베이스라인 모델인 Multi-channel CNN에 비하여 0.45%p 더 좋은 성능을 보였다. 표 3에 따라, NSMC 데이터 셋으로 먼저 학습하여 다른 데이터 셋에 적용하는 방법이 다음카카오 데이터 셋으로 먼저 학습하는 방법보다 성능 하락이 낮아 다른 데이터에 적용할 때 더 의미 있음을 알 수 있다.

5. 결론

본 논문에서는 영화평 감성 분석 데이터인 NSMC (Naver Sentiment Movie Corpus)와 다음카카오 데이터 셋을 이용하여 BERT기반 variational inference network와 RNN을 사용하는 모델을 제안하였다. 실험 결과, 본 논문에서 제안한 방법들 중 BERT와 RNN을 이용한 방법이 NSMC에서 90.51%, 다음카카오 평가 셋에서 93.00%로 가장 좋은 성능을 보였다.

본 논문에서는 BERT 모델을 이용하여 동적인 단어 표현을 얻어 감성 분석을 수행할 때 기존 다른 연구 방법들에 비하여 성능이 더욱 향상되는 것을 알 수 있다. [15]에서는 BERT 모델로부터 히든 스테이트를 얻어 FFNN을 이용하여 분류를 수행하였는데(표 1, 2에서 BERT+FFNN 모델과 같다), 본 논문에서는 BERT 모델과 함께 variational inference 또는 RNN을 사용하여 분류를 수행하였다. BERT 모델의 히든 스테이트를 variational inference로 샘플링 한 경우, [15]와 같은 방법으로 감성 분석을 수행할 때보다 성능이 더 향상되었다. BERT 모델과 함께 RNN을 사용한 경우에는 variational inference를 사용한 경우보다 더 향상되는 것을 보였으며, RNN으로 인코딩 후 variational inference를 적용했을 때보다도 더 좋은 성능을 보였다. 이는 BERT 모델에서 생성된 히든 스테이트를 RNN으로 양방향 인코딩하여 문맥 정보를 만든 것이 감성 분석을 해결하는데 더 도움이 되는 것으로 사료된다.

기존 연구들[3,10,28]에서 bi-gram 자질을 사용하는

것은 성능 향상에 도움이 됐지만, 본 논문에서 제안한 방법은 bi-gram 자질을 사용하지 않고 더 좋은 성능을 보였다. 이는 본 논문에서 제안한 방법이 감성 분석에 의미 있는 것을 알 수 있다.

본 논문에서는 BERT 모델을 기반으로 할 때 학습된 모델이 같은 도메인의 다른 데이터 셋에 대하여 평가 가능한지 실험을 통해 확인하였다. 그 결과, 본 논문에서 제안한 모델을 이용하여 동일한 데이터 셋으로 학습과 평가를 수행한 경우보다 다소 낮은 성능을 보였지만, 베이스라인으로 설정한 모델들에 비하여 더 좋은 성능을 보여 학습한 모델을 다른 데이터 셋에 적용 가능함을 알 수 있다.

향후 연구로는 BERT를 이용하여 자동으로 문장을 생성하고 그를 기반으로 데이터 셋을 증축하는 실험과 NSMC, 다음카카오 두 데이터 셋을 이용한 전이학습(transfer learning)에 대한 실험을 할 예정이다.

References

- [1] Jun Young Youn, Jung Ju Park, Do Won Kim, Tae Hong Min, and Jae Sung Lee, "Sentiment Analysis for Korean Product Review Using Stacked Bi-LSTM-CRF Model," *Proc. KIISE for HCLT*, pp. 633-635, 2018.
- [2] Jang Hwan-seok, Jang Eun-young, and Jeong Kwang-yong, "The method of analyzing sentiment keywords using Word2Vec," *KIISE winter*, pp. 661-663, 2017.
- [3] Geonyeong Kim and Changki Lee, "Korean Movie Review Sentiment Analysis Using Convolutional Neural Network," *KCC 2016*, pp. 747-749, 2016.
- [4] X. Glorot, A. Bordes and Y. Bengio, "Domain adaptation for large-scale sentiment classification: a deep learning approach," *ICML/11 Proceedings of the 28th International Conference on International Conference on Machine Learning*, pp. 513-520, 2011.
- [5] M. Bautin, L. Vijayarenu and S. Skiena, "International Sentiment Analysis for News and Blogs," *ICWSM*, 2008.
- [6] H. Jang and H. Shin, Language-Specific Sentiment Analysis in Morphologically Rich Languages, *COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pp. 498-506, 2010.
- [7] Y. Kim, Convolutional Neural Networks for Sentence Classification, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014*, pp. 1746-1751, 2014.
- [8] Geon-Yeong Kim and Chang-Ki Lee, "Korean Text Generation and Sentiment Analysis Using Model Combined VAE and CNN," *Proc. KIISE for HCLT*, pp. 430-433, 2018.
- [9] Min Kim, Jeunghyun Byun, Chunghee Lee, and

- Yeonsoo Lee, "Multi-channel CNN for Korean Sentiment Analysis," *Proc. KIISE for HCLT*, pp. 79-83, 2018.
- [10] Cheoneum Park, Geonyeong Kim, Hyunsun Hwang, and Changki Lee, "Contextualized Embedding-based Korean Movie Review Sentiment Analysis," *Proc. KIISE for HCLT*, pp. 75-78, 2018.
- [11] Yeongtaek Oh, Mintae Kim, and Wooju Kim, "Korean Movie-review Sentiment Analysis Using Parallel Stacked Bidirectional LSTM Model," *KCC 2018*, pp. 823-825, 2018.
- [12] <https://github.com/e9t/nsmc>
- [13] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory," *Neural computation* 9.8, pp. 1735-1780, 1997.
- [14] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio, "Learning phrase representation using RNN encoder-decoder for statistical machine translation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 14*, 2014.
- [15] J. Devlin, et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [16] A. Vaswani, et al. Attention Is All You Need. *Neural Information Processing Systems (NIPS)*, pp. 5998-6008, 2017.
- [17] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, Luke Zettlemoyer, "Deep contextualized word representations," *arXiv:1802.05365*, 2018.
- [18] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever, "Improving Language Understanding by Generative Pre-Training," <https://blog.openai.com/language-unsupervised/>, 2018.
- [19] Cheoneum Park, Changki Lee, Jun-Ho Lim, and Hyunki Kim, "Korean Dependency Parsing with BERT," *KCC 2019*, 2019.
- [20] Yishu Miao, Lei Yo, and Phil Blunsom. Neural Variational inference for Text Processing. *arXiv preprint arXiv:1511.06038*, 2016.
- [21] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [22] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra, Stochastic backpropagation and approximate inference in deep generative models, *In Proc. of ICML*, pp. 1278-1286, 2014.
- [23] Tao Lei, Yu Zhang, Sida I. Wang, Hui Dai, and Yoav Artzi, "Simple Recurrent Units for Highly Parallelizable Recurrence," *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, pp. 4470-4481, 2018.
- [24] D. Hendrycks and K. Gimpel, Gaussian Error Linear Units (GELUs). *arXiv preprint arXiv:1606.08415v3*, 2018.
- [25] <http://aiopen.etri.re.kr/>
- [26] DaumKakao. movie review page, http://movie.daum.net/review/netizen_point/movieNetizenPoint
- [27] <https://github.com/huggingface/pytorch-pretrained-BERT>
- [28] Cheoneum Park and Changki Lee, "Korean Movie Review Sentimental Analysis using RNN-based Variational Inference," *KSC 2018*, 2018.



박 천 음

2014년 강원대학교 학사. 2016년 강원대학교 석사. 2016년 3월~현재 강원대학교 컴퓨터과학과 박사과정. 관심분야는 자연어처리, 질의응답, 정보추출, 딥러닝



이 창 기

1999년 KAIST 학사. 2001년 POSTECH 석사. 2004년 POSTECH 박사. 2004년~2012년 2월 ETRI 선임연구원. 2012년 3월~현재 강원대학교 컴퓨터과학과 부교수. 관심분야는 자연어처리, 정보추출, 딥러닝