

RNN과 attention mechanism을 이용한 감성분석

최경호^o, 이창기

강원대학교

{gangsparkle, leeck}@kangwon.ac.kr

Sentiment analysis with GRU-Attention

Kyoungcho Choi^o, Changki Lee

Kangwon National University

요 약

감성분석은 텍스트에서 작성자의 주관적 의견을 탐지하여 제시하는 자연언어처리의 응용이다. 본 논문에서는 최근 기계번역에서 우수한 성능을 보이고 있는 RNN encoder-decoder 모델을 감성분석에 적합하도록 수정한 RNN-encoder 모델을 제안하고, 또한 최근 크게 각광받고 있는 attention 메커니즘을 bidirectional RNN과 결합한 GRU-Attention 모델을 제안한다. 실험 결과, Movie review data에서 GRU-Attention 모델의 성능은 accuracy 80.41%로 GRU-encoder 보다 1.22% 우수하였다.

1. 서 론

최근 우수한 성능으로 영상처리와 음성인식 분야에서 각광받는 딥러닝은 Neural Network Language Modeling(NNLM) 등을 통한 word vector representations[1]과 Long Short Term Memory RNN(LSTM)[2]등의 sequence labeling 문제에 특화된 딥러닝 모델이 등장하면서, 자연언어처리 분야에서도 활발히 연구되고 있다[3-7]. 또한 LSTM이나 Gated Recurrent Unit(GRU)를 기반으로 한 RNN encoder-decoder 모델은 기계번역 분야에서 우수한 성능을 보이고 있다[6]. [6]에서는 딥러닝을 이용하여 context vector를 만들고, context vector로부터 LSTM을 이용하여 출력 언어 문장을 생성하는 방식으로 기계번역을 수행하였다. context vector란 다양한 길이의 문장을 정해진 크기의 vector로 요약한 것이다.

Attention mechanism[7]은 RNN의 각 시간 별 hidden layer의 출력 중 중요한 정보가 있다고 판단되는 곳에 가중치를 주고, 합을 구해 context vector를 만드는 방법이다.

본 논문에서는 LSTM과 GRU를 응용한 RNN-encoder 모델과 attention mechanism을 적용한 GRU-Attention을 제안하고, 감성분석(sentiment analysis)에서의 성능을 LSTM-encoder, GRU-encoder와 비교해 본다.

2. 관련연구

텍스트에서 특정 키워드에 대한 긍정 또는 부정 등의 작성자의 주관적 의견을 탐지하여 제시해 주는 감성분석(Sentiment analysis)이라 한다. 기존의 감성분석 연구로는 긍정 부정 등의 감성을 나타내는 단어를 감성어라 정의하고, 사전을 구축하여, 감성어의

빈도나 특징을 파악하는 형태의 연구인 사전기반의 방법 등이 있었다. 사전기반의 연구에서는 자체 제작한 데이터를 대상으로 72.62% accuracy 성능을 보였다[8]. 최근 recursive matrix-vector spaces 모델(MVRNN)로 감성분석에 딥러닝을 적용한 연구에서는 79.0%의 성능을 보였다[9].

3. GRU-encoder 모델

Gated Recurrent Unit은 reset과 update 두 개의 gate를 가진 LSTM의 변형 모델로, 아래 수식과 같이 정의 된다[10].

$$z = \sigma(W_z x_t + U_z h_{t-1} + b_z)$$

$$r = \sigma(W_r x_t + U_r h_{t-1} + b_r)$$

$$m = \text{act}(W_m x_t + U_m (h_{t-1} \odot r) + b_m)$$

$$h_t = (1 - z) \odot h_{t-1} + z \odot m \quad (1)$$

수식1에서 σ 는 sigmoid 함수를 뜻하고, x_t 는 현재 시퀀스의 입력을, h_{t-1} 는 이전 시퀀스의 출력을 뜻하고, $W_z, U_z, W_r, U_r, W_m, U_m$ 는 각 gate와 cell memory를 위한 weight matrix이다. r 은 직전상태를 유닛의 입력에 반영하는 비율을 결정하는 reset gate이고, z 는 update gate로 유닛의 출력에 직전상태를 보존하여 반영하는 정도를 결정한다. \odot 은 element-wise product를 의미한다. act 는 activation 함수를 뜻하며 본 논문에서는 tanh함수와 sigmoid 함수를 비교하여 사용하였다.

GRU-encoder는 GRU의 마지막 시퀀스의 hidden layer의 출력을 context vector로 사용하는 모델로 다음 수식 2로 간단히 나타낼 수 있다.

$$c = h_T$$

$$y = \text{softmax}(W_y c + b_y) \quad (2)$$

c 는 문장을 요약한 context vector이며, T 는 문장의 길이를 뜻하고, h_T 는 GRU의 마지막 hidden layer이다. y 는 시스템의 output layer를 뜻하고, softmax를 사용하였다. 그림1은 GRU-encoder를 이용하여 문장의 context vector(c)를 구하는 예제이다.

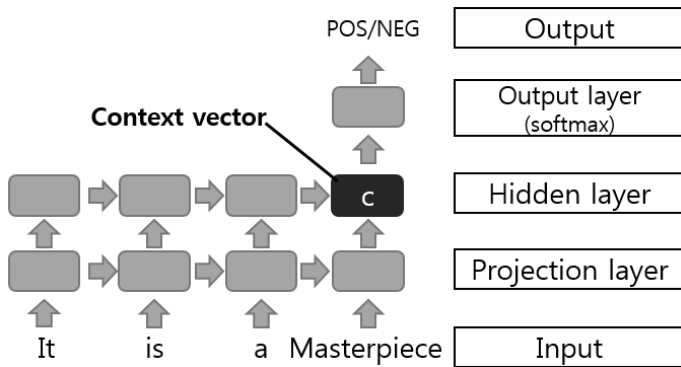


그림1. GRU-encoder로 구한 context vector

4. GRU-Attention

Attention mechanism은 각 시퀀스(t)의 hidden layer 출력의 중요도인 attention들을 구하고, 각 hidden layer 출력에 attention을 곱한 가중치 합(Weighted sum)을 구하여 Context vector를 구축한다. GRU-Attention 모델은 앞서 설명한 GRU-encoder수식의 context vector에 attention mechanism을 적용하여 다음 수식3으로 확장한 모델이다.

$$s_t = V_a^T \tanh(W_a h_t + b_a)$$

$$a_t = \frac{\exp(s_t)}{\sum_k \exp(s_k)}$$

$$c = \sum_t a_t h_t \quad (3)$$

수식3에서 V_a 는 attention을 구하기 위한 weight vector 이고, h_t 는 수식2에서 구한 각 시퀀스의 hidden layer의 출력을 나타낸 vector이며, W_a 는 attention을 구하기 위해 학습되는 weight matrix 이고, b_a 는 Bias 이다. a_t 는 attention 으로 context vector(c)를 구할 때 각 h_t 에 대한 가중치 역할을 하게 된다.

그림2는 GRU-Attention 모델에서 hidden layer의 출력과 attention으로 context vector를 구하는 예제를 그림으로 나타낸 것이다. attention은 각 h_t 에 가중치로 곱해져, 문장 내에서 결과에 영향을 크게 미치는 h_t 를 강조하게 된다.

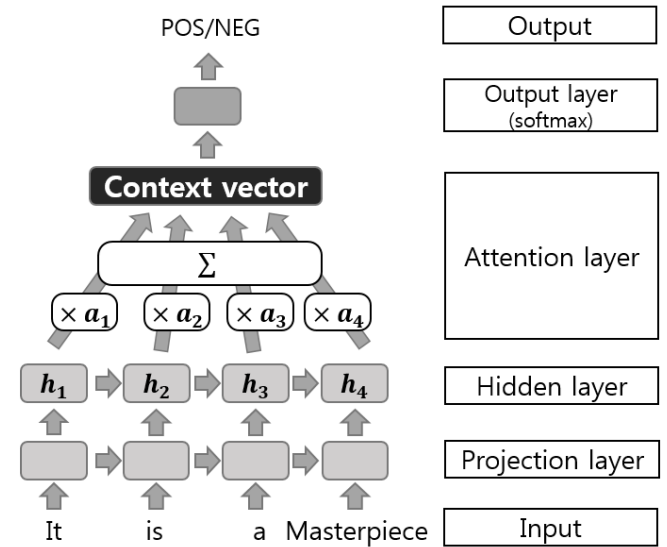


그림2. GRU-Attention 모델의 attention

5. 실험

시스템의 학습과 평가를 위하여 Cornell movie review data[11]중 sentiment polarity datasets을 사용하였다. sentiment polarity datasets은 영화에 대한 긍정적 의사를 표현한 5311개의 문장과, 부정적 의사를 표현한 5311개의 문장으로 구성되어 있다. dataset에서 긍정 문장과 부정 문장을 9592 문장 뽑아 Train set으로 사용했고, 1067문장을 test set으로 사용하였다. word embedding은 SENNA시스템에서 공개된 것을 사용하였다. dropout은 projection layer와 hidden layer에 0.2와 0.5의 dropout 확률로 사용하였다.

그림3은 GRU-Attention 모델의 최적 성능을 구하기 위해 Hidden unit의 개수를 변경해가며 성능을 측정한 것이다.

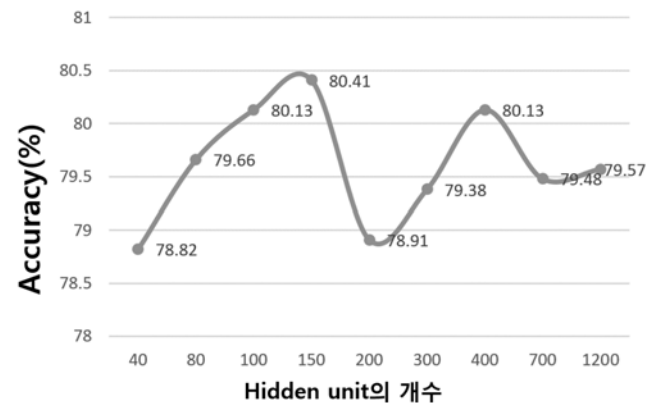


그림3. GRU-Attention의 hidden unit에 따른 성능

그림3을 보면 hidden unit의 개수가 150 이상으로 많아지면 점차 accuracy가 낮아지는 것을 알 수 있다. 이는 hidden unit의 개수가 많아지면서 학습되는 파라미터의 개수도 함께 증가해 과적합이 발생한 것으로 판단된다.

GRU-Attention 모델은 hidden unit이 80개에서 180개 사이일 때 최적의 성능을 보였다. 또한 LSTM-encoder과 GRU-encoder모델도 Hidden unit의 개수와 dropout 확률을 같이 변경해가며 최적의 성능을 구했고, 그 결과는 표1과 같다.

표1. MR data에서 각 모델 별 성능

	act	hidden	p_drop	h_drop	acc
RAE[12]	-	-	-	-	77.70
Tree-CRF[13]	-	-	-	-	77.30
MV-RNN[9]	-	-	-	-	79.00
LSTM encoder	sigm	200	0.2	0.5	78.73
	tanh	300	0.1	0.3	79.10
GRU encoder	sigm	50	0.2	0.5	78.44
	tanh	100	0.2	0.5	79.19
GRU-attention	sigm	100	0.2	0.5	79.19
	tanh	120	0.2	0.5	80.41

표1 에서 act는 activation function, hid는 hidden unit의 개수, p_d는 projection layer의 dropout 비율, h_d는 hidden layer의 dropout 비율을 뜻하며, acc는 시스템의 accuracy 성능을 뜻한다. GRU-attention 모델이 가장 accuracy 80.41%로 가장 우수한 성능을 보였으며, 79.0%의 MV-RNN 모델을 이용한 시스템[9] 보다 우수하였다. 따라서 attention mechanism을 이용하여 구축한 context vector가 encoder모델로 구축한 context vector보다 우수함을 알 수 있다.

The(5) best(20) of(26) the(9) pierce(10)

Brosnan(3) Fames(1) Bond(0) films(0) to(0) date(9) .(12)

그림4. Hidden layer의 입력 단어와 attention 예

또한 GRU-attention 모델을 사용 하면 그림4와 같이 입력 단어에 따른 attention을 확인 할 수 있어, classification의 근거를 직접 확인하고, 활용 할 수 있다.

6. 결론

본 연구에서는 GRU-Attention 모델을 이용하여 영문 영화리뷰 데이터를 대상으로 감성분석을 수행하였다. 제안한 GRU-Attention 모델은 GRU-encoder과 LSTM-encoder 모델보다 우수한 성능을 보였다.

향후 연구로는 LSTM에 attention mechanism을 적용해보고, 한국어를 대상으로 GRU-Attention 모델을 이용하여 감성분석을 수행할 것이다.

감사의글

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임. (No.R0101-15-0062, 휴먼 지식증강 서비스를 위한 지능 진화형 WiseQA 플랫폼 기술 개발)

참 고 문 헌

- [1]Tomas Mikolov, Quoc V. Le and Ilya Sutskever, "Exploiting similarities among languages for machine translation", Technical report, arXiv, 2013.
- [2]Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory", Neural computation 9.8: 1735-1780, 1997.
- [3]Ronan Collobert et al. "Natural language processing (almost) from scratch", The Journal of Machine Learning Research 12: 2493-2537, 2011.
- [4]배장성, 이창기, 임수중, "딥 러닝을 이용한 한국어 의미역 결정", 한국정보과학회 학술발표논문집, 690-692, 2015.
- [5]이창기, "Long Short-Term Memory 기반의 Recurrent Neural Network를 이용한 개체명 인식" 한국정보과학회 학술발표논문집, 645-647, 2015.
- [6]Ilya Sutskever, Oriol Vinyals and Quoc V. Le. "Sequence to sequence learning with neural networks", Advances in neural information processing systems, 2014.
- [7]Minh-Thang Luong, Hieu Pham and Christopher D. Manning. "Effective Approaches to Attention-based Neural Machine Translation.", arXiv preprint arXiv:1508.04025, 2015.
- [8]김경민, 이지형, "어휘기능 정보를 이용한 트위터 감성분석." 한국정보과학회 학술발표논문집, 734-736. 2014.
- [9]Richard Socher 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. ACL, 2012.
- [10]Junyoung Chung et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling", arXiv preprint arXiv:1412.3555, 2014.
- [11]Movie Review Data, <http://www.cs.cornell.edu/people/pabo/movie-review-data/>
- [12]Richard Socher et al. "Semi-supervised recursive autoencoders for predicting sentiment distributions", *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 2011.
- [13]Tetsuji Nakagawa, Kentaro Inui and Sadao Kurohashi. "Dependency tree-based sentiment classification using CRFs with hidden variables" *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. ACL, 2010.