LAAN: A Linguistic-Aware Attention Network for Sentiment Analysis

Zeyang Lei
Graduate School at Shenzhen,
Tsinghua University
leizy16@mails.tsinghua.edu.cn

Yujiu Yang*
Graduate School at Shenzhen,
Tsinghua University
yang.yujiu@sz.tsinghua.edu.cn

Yi Liu
Peking University Shenzhen
Institute
eeyliu@gmail.com

ABSTRACT

Sentiment analysis of social media and comment data is an important issue in opinion monitoring. In this work, we propose a Linguistic-Aware Attention Network (LANN) to enhance the performance of convolution neural network (CNN). LANN adopts a two-stage strategy to model the sentiment-specific sentence representation. First, an interactive attention mechanism is designed to model word-level semantics. Second, to capture phrase-level linguistic structure, a dynamic semantic attention is adopted to select the crucial phrase chunks in the sentence. The experiments demonstrate that LANN has robust superiority over competitors and has reached the state-of-the-art performance.

CCS CONCEPTS

• Information systems → Sentiment analysis; • Computing methodologies → Natural language processing;

KEYWORDS

Sentiment analysis; interactive attention; dynamic semantic attention

ACM Reference Format:

Zeyang Lei, Yujiu Yang, and Yi Liu. 2018. LAAN: A Linguistic-Aware Attention Network for Sentiment Analysis. In *WWW '18 Companion: The 2018 Web Conference Companion, April 23-27, 2018, Lyon, France*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3184558.3186922>

1 INTRODUCTION

Sentiment classification aims to assign the sentiment polarity towards a given text. It has broad applications such as public opinion monitoring, product recommendation and so on [1]. Most existing studies build sentiment classifiers using deep learning approaches, such as convolutional neural network (CNN), long short-term memory (LSTM).

Despite the effectiveness of deep learning approaches, sentiment classification still remains a defect in real-world scenarios. The traditional sentiment lexicon plays an important role

in sentiment classification. However, the sentiment lexicon has not been fully exploited in recent neural network models (e.g., CNN and LSTM) for sentiment classification.

In this paper, we propose a Linguistic-Aware Attention Network (LANN), which aims to integrate the high-quality sentiment lexicon into convolution neural network to further boost the performance of sentiment classification. More specifically, our model adopts a two-step strategy to model word-level and phrase-level semantics, respectively. First, we design an interactive attention mechanism to fully utilize the mutual information between context words and sentiment words in the sentence. The interactive attention mechanism helps highlight the relevance context words associated with the sentiment of the sentence. Second, to better capture the sentiment-specific semantics structure of the sentence, we design a new dynamic semantics attention over n-gram convolution results, which is capable of selecting the crucial phrase chunks to compose the sentiment-specific sentence representation.

2 THE PROPOSED MODEL

LAAN mainly consists of two parts: the word-Level interactive attention module and the phrase-level dynamic semantic attention module.

Word-Level Interactive Attention. Suppose that a sentence x consists of n context words $W^c = [w_1^c, w_2^c, \dots, w_n^c]$ and m sentiment words $W^s = [w_1^s, w_1^s, \dots, w_m^s]$, where each w denotes a specific word. We embed each word into a low dimensional semantic space. This means, each word is mapped to its embedding $w \in \mathbb{R}^d$. Then we define the correlation matrix $\mathbf{C} = W^c \cdot (W^s)^T \in \mathbb{R}^{n \times m}$, which represents the mutual information between context words and sentiment words. Here, each element $\mathbf{C}_{i,j}$ in the correlation matrix denotes the relevant degree between i -th word w_i^c of the context words and j -th word w_j^s of the sentiment words. Then we average all values in a row of $\mathbf{C}$ ("col-wise average", resulting in the column vector of size n) and apply a softmax operation to obtain a single attention weight vector $\alpha = \text{softmax}(\frac{\sum_{h=1}^m \mathbf{C}[:,h]}{m})$ about the context words. Similarly, we average all values in a column of $\mathbf{C}$ ("row-wise average", resulting in the row vector of size m) and apply a softmax operation to obtain a single attention weight vector $\beta = \text{softmax}(\frac{\sum_{k=1}^n \mathbf{C}[k,:]}{n})$ about the sentiment words. Next, with the attention weight vector α and the context words W^c as inputs, we can obtain the correlation vector w_α^c about the context: $w_\alpha^c = \sum_{i=1}^n \alpha_i w_i^c$. w_α^c denotes the overall relevance metric of the sentiment words towards the context words, which is beneficial to highlight the

*Corresponding author

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23-27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3186922>

important words in the context for sentiment classification. Similarly, we can obtain the correlation vector w_β^s about the sentiment words: $w_\beta^s = \sum_{i=1}^m \beta_i w_i^s$. w_β^s denotes the overall relevance metric of the context words towards the sentiment words, which helps integrate contextual global information into the representation of sentiment words.

Finally, with the concatenation of each context word and the correlation vector w_α^c , the sentiment-enhanced representation $W_e^c \in \mathbb{R}^{2d \times n}$ of the context words can be computed as follows:

$$W_e^c = \text{relu}([U^c W^c \oplus \{U^\alpha w_\alpha^c \otimes e_n\}]) \quad (1)$$

where U^c, U^α are projection parameters, $\oplus$ refers to the concatenation operator, $w_\alpha^c \otimes e_n = [w_\alpha^c; w_\alpha^c; \dots; w_\alpha^c]$ denotes the operator which repeatedly concatenates w_α^c for n times and relu is the semantic mapping function. Similarly, we can obtain the contextual enhanced representation $W_e^s \in \mathbb{R}^{2d \times m}$ of the sentiment words as follows:

$$W_e^s = \text{relu}([U^s W^s \oplus \{U^\beta w_\beta^s \otimes e_m\}])$$

Phrase-Level Dynamic Semantic Attention. After obtaining W_e^c and W_e^s , we adopt multi-gram convolution operation to capture different local semantic units [3]. Formally, with different window size, we induce a combined multi-gram feature map H^c for context words and H^s for sentiment words.

After performing convolution operation, in order to select those crucial sentimental phrase chunks to model the sentiment-specific sentence representation, we design a new dynamic semantic attention. The details about dynamic semantics attention are formulated as follows:

$$\tilde{A} = \text{softmax}(H^c \{H^s\}^T) \quad (2)$$

$$\tilde{T} = (H^c)^T \tilde{W} \tilde{A} \quad (3)$$

$$S = H^s \odot \tilde{T}, \bar{s} = \frac{\sum_{j=1}^m H^s[:, j]}{m} \quad (4)$$

$$\tilde{\alpha} = \text{softmax}(v^T \tanh(V H^c + \tilde{U} \bar{s} \otimes e_n)) \quad (5)$$

$$o = \sum_{i=1}^n \tilde{\alpha}_i H_i^s \quad (6)$$

where $\odot$ denotes element-wise multiplication, o is the final sentiment-specific sentence representation.

Finally, we feed the output vector o to a softmax layer to predict the sentiment distribution. The training objective is to minimize the cross-entropy error of the predicted and true class distributions.

3 EXPERIMENTS

Datasets and Sentiment Lexicon. Movie Review (MR)¹ and Stanford Sentiment Treebank (SST)² are used for evaluating our proposed model. Sentiment lexicon combines the sentiment words from both [4] and [2].

Experiment Results. In our experiments, we use classification accuracy as the evaluation metric. The experimental

results are shown in Table 1. From the results, our proposed model is effective to improve the performance of sentence-level sentiment classification. To be specific, we have the following findings: (i) Our model significantly outperforms those methods that do not fully employ linguistic knowledge, such as RNN/ RNTN, LSTM, BiLSTM, CNN on both datasets, which verifies the effectiveness of the idea that integrates sentiment linguistic knowledge into the deep neural networks. (ii) Compared with LR-LSTM and LR-Bi-LSTM which employ linguistic role of sentiment, negation and intensity words to neural networks by linguistic regularization with the similar idea, our model outperforms LR-LSTM and LR-Bi-LSTM by a significant margin on the two datasets. It demonstrates the considerable superiority of our model.

Methods	MR	SST(sent.-level)
RNN(i.e.,RAE) [5]	77.7%	43.2%
RNTN [6]	75.9%*	45.7%
LSTM	77.4%*	45.6%*
BiLSTM	79.3%*	46.5%*
Tree-LSTM [7]	80.7%	48.1%
CNN [3]	81.5%	48.0%
NSCL [8]	82.9%	47.1%
LR-LSTM [4]	81.5%	48.3%
LR-Bi-LSTM [4]	82.1%	48.6%
LAAN(our model)	83.9%	49.1%

Table 1: Evaluation results. The best result on each dataset is in bold. Results with * are obtained either with the same codes shared by the original authors or by our own implementation.

4 ACKNOWLEDGMENTS

This work was supported in part by the Research Fund for the development of strategic emerging industries by ShenZhen city (No. JCYJ20160301151844537 and No. JCYJ20160331104524983)

REFERENCES

- [1] Mauro Dragoni. 2017. A Three-Phase Approach for Exploiting Opinion Mining in Computational Advertising. *IEEE Intelligent Systems* (2017), 21–27.
- [2] Mingqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of SIGKDD*.
- [3] Yoon Kim. 2014. Convolutional neural networks for sentence classification. *Proceedings of EMNLP 2014*.
- [4] Qiao Qian, Minlie Huang, Jinhao Lei, and Xiaoyan Zhu. 2017. Linguistically Regularized LSTM for Sentiment Classification. In *Proceedings of ACL 2017*.
- [5] Richard Socher, Jeffrey Pennington, Eric H Huang, Andrew Y Ng, and Christopher D Manning. 2011. Semi-supervised recursive autoencoders for predicting sentiment distributions. In *Proceedings of EMNLP 2011*.
- [6] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of EMNLP 2013*.
- [7] Kai Sheng Tai, Richard Socher, and Christopher D. Manning. 2015. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. In *Proceedings of ACL 2015*.
- [8] Zhiyang Teng, Duy-Tin Vo, and Yue Zhang. 2016. Context-Sensitive Lexicon Features for Neural Sentiment Analysis. In *Proceedings of EMNLP 2016*.

¹<http://www.cs.cornell.edu/people/pabo/movie-review-data/>

²<https://nlp.stanford.edu/sentiment/>