A Fine-grained Sentiment Analysis Method Based on Dependency Tree and Graph Attention Network

Yejin Tan¹, Wangshu Guo¹, Jiawei He¹, Jian Liu¹*, Ming Xian¹

¹College of Electronic Science and Technology, National University of Defense Technology, Changsha, Hunan, 410073, China
²ljabc730@nudt.edu.cn *Corresponding author’s e-mail: ljabc730@gmail.com

Abstract. With the rapid development of the Internet, e-commerce and social media have continued to develop and grow. Merchants tend to better understand users’ attitudes and emotional tendencies through comments. As the number of user reviews increases, the previous sentiment analysis methods have highlighted the problems of high cost and high error rate. How to use more advanced methods to analyse comments has become an urgent problem to be solved. In order to solve the above problems, in this article, we propose a fine-grained sentiment analysis method based on dependency tree and graph neural network, which can help businesses and social network platforms to identify users’ sentiment tendencies and can be subsequently used in recommendation systems and public opinion analysis systems. The experimental results show that our scheme has achieved the best results.

1. Introduction

With the rapid development of the Internet, e-commerce continues to grow and develop, and many excellent companies have emerged, such as Taobao, Jingdong. With the development of big data technology and the increase in the number of reviews, merchants tend to better understand users’ attitudes and emotional tendencies through reviews.

Traditional sentiment analysis methods based on document-level and sentence-level cannot accurately judge different aspects of sentiment in sentences. For example, in the sentence ‘The menu is limited but the dishes are excellent.’, the sentiment polarity is different for each aspect. Aspect-based sentiment analysis (ABSA) can judge the sentiment polarity of each aspect in a sentence, which is very suitable for product reviews.

Existing work on aspect-based sentiment analysis (ABSA) can be classified into two main aspects, those methods examine sentiment polarity using attention mechanisms in neural networks and those utilizing syntax structures. Attention mechanisms can be considered as an implicit method of using sentence structure, because opinion words usually seem not far away from all aspects. Such methods have made gratifying progress. For example, Wang et al.[¹] introduced an attention-based LSTM model to identify those sentiment information related to a specific aspect. Tang et al. [²] proposed a multi-hop attention Memory Network model. Moreover, to obtain long-distance opinion words for aspects, Chen et al. [³] proposed a multi-layer attention neutral network. Fan et al. [⁴] introduced a multi-grained attention network with different granularities. With the introduction of the pre-trained language model BERT[⁵], many good models have been proposed and achieved good performance. For example, Xu et al. [⁶] utilized an added corpus to train BERT and proved its effectiveness in ABSA. Sun et al. [⁷] transformed ABSA into a sentence-pair classification task by introducing some auxiliary sentences.
Other works have attempted to include grammatical information directly in ABSA. Since aspects are often assumed to be the core of this task, it is important to establish syntactic links between each target aspect and other words. In order to identify the relationship between aspects and potential opinion words, Qiu et al.\cite{8} manually defined some syntactic rules. And then Liu et al.\cite{9} obtained Partially aligned links with these syntactic rules. Afterward, a large number of neural network-based models were applied in this work. In another work, Wang et al.\cite{10} integrated RNN with conditional random fields(CRF) for the co-extraction of aspects and opinion terms. Moreover, He et al.\cite{11} introduced an attention model that incorporates syntactic information into the attention mechanism, which can better capture the semantics of the viewpoint target.

Recently, graph neural networks have achieved good results in many NLP tasks. However, the biggest problem is how to represent the data as a grid-like structure. To learn node representations from a dependency tree, Zhang et al.\cite{12} proposed a graph convolutional network(GCN) over the dependency tree and Sun et al.\cite{13} presented a convolution over a dependency tree (CDT) model. For a similar purpose, Huang et al.\cite{14} propose target-dependent graph attention network(TD-GAT) utilizing the dependency relationship between words.

In order to solve those problems, in this article, we propose a fine-grained sentiment analysis method based on dependency tree and graph neural network, which can help businesses and social network platforms to identify user sentiment tendencies, the experimental results show that our scheme has achieved the best results.

2. Our Scheme

In our scheme, there are two types of text that need to be processed, namely text and relation. The overall framework of our model is shown in Figure 1. It is mainly composed of two process: the prediction of text and the prediction of relation.
2.1. Data preprocessing

Before data processing, we first perform data preprocessing, the main work is to extract semantic relations. For an input sentence, we first use the biaffine parser[15] to extract the semantic relationship of the sentence. We represent the input as $S = [S_1, S_2, \ldots, S_n]$

$$S_i = \{"sentence": "sentence", "tokens": [x_1, x_2, \ldots, x_n], \}
$$

$$\text{"dependencies": } \{[y_1, p_{i1}, p_{i2}], [y_2, p_{i2}, p_{i2}], \ldots, [y_n, p_{ni}, p_{ni}]\}, \}
$$

$$\text{"aspect_sentiment": } \{[\text{"aspect"}, \text{"sentiment"}], \text{"from_to"}: \{[p_1, p_2]\}\}
$$

where "sentence" represents a complete sentence, in "tokens", $x_i$ represents a single word, in "dependencies", $y_i$ represents the dependency relationship of two words, $p_{i1}$ and $p_{i2}$ represent the orientation of the relationship, in "aspect_sentiment", "aspect" represents the aspect word, "sentiment" represents the sentiment polarity, in "from_to", $p_1$ and $p_2$ represent the Start and end position, respectively.

According to the result of dependency parser, we extract the semantic relationship of the sentence and represent them utilizing dependency tree, as shown in Figure 2.

![Figure 2 Construction of an ordinary dependency tree](image)

However, utilizing an ordinary dependency tree can’t express the sentiment polarity of different aspects. To solve this problem, we reshape the dependency tree based on aspect. The reshaped dependency tree is shown as Figure 3.

2.2. Data processing

In the process of data processing, there are two types of data that need to be processed, namely text and relation. For text, we use BERT for processing and get context-based prediction results $h^w = \{h_1^w, h_2^w, \ldots, h_n^w\}$.

For relation, we use the last hidden states of the pre-trained BERT for word representations and use relation-graph attention network (R-GAT)[16] to obtain the prediction results based on semantic relations.
Given a set of relation embeddings \( \{r_1, r_2, \ldots, r_n\} \), R-GAT takes them as \( \{h^0_i, h^1_i, \ldots, h^0_u\} \) and produces a new set of word features \( \{h'_1, h'_2, \ldots, h'_u\} \) as its output.

Given a relation \( r_i \) with its neighbors \( r_j \in N(i) \) and its representation \( h^{l-1}_j \), the multi-head-attention-based R-GAT can be described as:

\[
\begin{align*}
  h'_j & = \| \sigma( \sum_{j \in N(i)} \alpha_j^z W^z_j h^{l-1}_j ) \\
  \alpha_j^z & = \frac{\exp(F(h^z_i, h^z_j))}{\sum_{j \in N(i)} \exp(F(h^z_i, h^z_j))}
\end{align*}
\]

where \( \| \) denotes vector concatenation, \( W^z_j \in \mathbb{R}^{d \times d} \) is a parameter matrix at layer \( l \), \( d \) represents the dimension of relation vectors, \( Z \) is the number of attention heads and \( \sigma \) denotes the sigmoid activation function.

The weight \( \alpha_j^z \) is calculated via attention process:

\[
\alpha_j^z = \frac{\exp(F(h^z_i, h^z_j))}{\sum_{j \in N(i)} \exp(F(h^z_i, h^z_j))}
\]

where \( F \) is an attention function. In this paper, we use the scaled dot-product attention function:

\[
F(h^z_i, h^z_j) = \frac{(W^z_i h^z_i)(W^z_j h^z_j)^T}{\sqrt{d/Z}}
\]

where \( W^z_i, W^z_j \in \mathbb{R}^{d \times d} \) are parameter matrices in layer \( l \). The final representation of each word is computed by:

\[
\begin{align*}
  x_t^{l+1} & = h^B_n || h_t^{l+1} \\
  h_t^{\text{output}} & = \text{relu}(W_{l+1} x_t^{l+1} + b_t^{l+1})
\end{align*}
\]
3. Model training
During model training, the data will be transmitted forward along the neural network, and the network will calculate the probability of sentiment polarity. The calculation formula is as follows:

\[
p(a) = \text{softmax}(W\hat{h}^{\text{output}} + b)
\]

After applying R-GAT on an reshaped dependency tree, its output \(\hat{h}^{\text{output}}\) is passed through a softmax layer and mapped to probabilities of sentiment polarities. Finally, we use the standard cross-entropy loss as our objective function:

\[
L(\theta) = -\sum_{S \in D} \sum_{a \in A} \log p(a)
\]

where \(A\) represents the aspects appearing in sentence \(S\), \(D\) denotes all the sentence-aspects pairs and \(\theta\) contains all the parameters.

4. Experiment

4.1. Datasets and Implementation
In this paper, the datasets we used are from SemEval 2014 Task 4 [18], the statistics of them can be found in Table 1.

| Dataset | Positive | Neutral | Negative |
|---------|----------|---------|----------|
|         | Train    | Test    | Train    | Test    | Train    | Test    |
| Restaurant | 994     | 341     | 870  | 128 | 464 | 169 |
| Laptop  | 2164 | 728 | 807 | 196 | 637 | 196 |
| Twitter | 1561 | 173 | 3127 | 346 | 1560 | 173 |

In this paper, we use 5 heads in our scheme, and the hidden size is set as 300. The dropout is set as 0.1, and set L2 regularization term \(\lambda=10^{-5}\). The learning rate \(2 \times 10^{-5}\) is adopted for training.

4.2. Results
We compare our scheme with the baseline models, the experiment results are shown in Table 2.

| Method       | Restaurant Accuracy | Macro-F1 | Laptop Accuracy | Macro-F1 | Twitter Accuracy | Macro-F1 |
|--------------|---------------------|----------|-----------------|----------|------------------|----------|
| LSTM+SynATT  | 80.45               | 71.26    | 72.57           | 69.13    | -                | -        |
| ASGCN        | 80.77               | 72.02    | 75.55           | 71.05    | 72.15            | 70.40    |
| CDT          | 82.30               | 74.02    | 77.19           | 72.99    | 74.66            | 73.66    |
| GAT          | 78.21               | 67.17    | 73.04           | 68.11    | 71.67            | 70.13    |
| TD-GAT       | 80.35               | 76.13    | 74.13           | 72.01    | 72.68            | 71.15    |
| ATA-E-LSTM   | 77.20               | -        | 68.70           | -        | -                | -        |
| MGAN         | 81.25               | 71.94    | 75.39           | 72.47    | 72.54            | 70.81    |
| RAM          | 80.23               | 70.80    | 74.49           | 71.35    | 69.36            | 67.30    |
| BERT         | 85.92               | 78.28    | 77.58           | 72.38    | 75.28            | 74.11    |
| Our-Scheme   | 87.23               | 80.44    | 78.35           | 74.37    | 76.38            | 74.97    |

We can find that our model outperforms most of the baseline models and the performance can be significantly improved when considering relational heads in our reshaped dependency tree structure.

This demonstrates that our model is better at encoding the semantic relationship information. These results have demonstrated the effectiveness of our scheme in capturing important syntactic structures for sentiment analysis.
5. Conclusion
In this paper, we proposed a fine-grained sentiment analysis method utilizing Biaffine syntactic parser, dependency tree and graph attention network, which can help businesses and social network platforms to identify user sentiment tendencies and can be subsequently used in recommendation systems and public opinion analysis systems. The experimental results show that our scheme has achieved the best results.

Funding: This research was funded by the National Natural Science Foundation of China under Grant No.61801489.

References
[1] Wang Y, Huang M, Zhu X, et al. Attention-based LSTM for Aspect-level Sentiment Classification[C]// Conference on Empirical Methods in Natural Language Processing. 2016. More references
[2] Tang D, Qin B, Liu T. Aspect Level Sentiment Classification with Deep Memory Network[J]. 2016.
[3] Chen P, Sun Z, Bing L, et al. Recurrent attention network on memory for aspect sentiment analysis[C]//Proceedings of the 2017 conference on empirical methods in natural language processing. 2017: 452-461.
[4] Fan F, Feng Y, Zhao D. Multi-grained attention network for aspect-level sentiment classification[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018: 3433-3442.
[5] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. 2018.
[6] Xu H, Liu B, Shu L, et al. Bert post-training for review reading comprehension and aspect-based sentiment analysis[J]. arXiv preprint arXiv:1904.02232, 2019.
[7] Sun C, Huang L, Qiu X. Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence[J]. arXiv preprint arXiv:1903.09588, 2019.
[8] Qiu G, Liu B, Bu J, et al. Opinion word expansion and target extraction through double propagation[J]. Computational linguistics, 2011, 37(1): 9-27.
[9] Liu K, Xu H L, Liu Y, et al. Opinion target extraction using partially-supervised word alignment model[C]//Twenty-Third International Joint Conference on Artificial Intelligence. 2013.
[10] Wang W, Pan S J, Dahlmeier D, et al. Recursive neural conditional random fields for aspect-based sentiment analysis[J]. arXiv preprint arXiv:1603.06679, 2016.
[11] He R, Lee W S, Ng H T, et al. Effective attention modeling for aspect-level sentiment classification[C]//Proceedings of the 27th International Conference on Computational Linguistics. 2018: 1121-1131.
[12] Zhang C, Li Q, Song D. Aspect-based Sentiment Classification with Aspect-specific Graph Convolutional Networks[J]. arXiv preprint arXiv:1909.03477, 2019.
[13] Sun K, Zhang R, Mensah S, et al. Aspect-Level Sentiment Analysis Via Convolution over Dependency Tree[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 5683-5692.
[14] Huang B, Carley K M. Syntax-aware aspect level sentiment classification with graph attention networks[J]. arXiv preprint arXiv:1909.02606, 2019.
[15] Dozat T, Manning C D. Deep biaffine attention for neural dependency parsing[J]. arXiv preprint arXiv:1611.01734, 2016.
[16] Wang K, Shen W, Yang Y, et al. Relational Graph Attention Network for Aspect-based Sentiment Analysis[J]. arXiv preprint arXiv:2004.12362, 2020.
[17] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural
information processing systems. 2017: 5998-6008.

[18] Pontiki M, Papageorgiou H, Galanis D, et al. SemEval-2014 Task 4: Aspect Based Sentiment Analysis[J]. SemEval 2014, 2014: 27.