Document-level Sentiment Analysis Based on Domain-specific Sentiment Words

To cite this article: Chengai Sun et al 2019 J. Phys.: Conf. Ser. 1288 012052

View the article online for updates and enhancements.
Document-level Sentiment Analysis Based on Domain-specific Sentiment Words

Chengai Sun, Fang Wang and Gang Tian

College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266000, China
Email: 467341167@qq.com, gtian@sdust.edu.cn

Abstract. Aiming at the problem that the relationship between non-contiguous words in a sentence cannot be effectively captured, and the existing sentiment lexicon has poor adaptability in the field, a model for document-level sentiment analysis based on domain-specific sentiment words is constructed. We reconstructed word vectors using attention mechanism to capture the relationship between non-contiguous words in word vectors; Words are synthesized using Asymmetric Convolutional Neural Network. Sentences are synthesized by Bidirectional Gated Recurrent Neural Network based on attention mechanism to form document vector features; we used CNN to construct a domain-specific sentiment dictionary to generate emotional vector features; Document vector features and emotional vector features are combined using a linear binding layer to form document features that facilitate document classification. By comparing the performance of this method with other methods through experiments, the results show that there is a big advantage in classification accuracy, and can be widely used in various specific fields such as public health.

1. Introduction

Sentiment analysis and opinion mining are the most widely used applications in the field of language. Document-level sentiment classification is an important aspect of sentiment analysis. Tang et al [1] used the principle of synthesizing to fully consider the relationship between sentences, but did not consider the relationship between words. Yang et al [2] proposed a hierarchical attention network for document classification, extracted the core features of the document, but did not introduce the domain-specific emotional words into the model. Zhang et al [3] used a Dependency Sensitive Convolutional Neural Network to model sentences and documents, capturing the dependencies between sentences, but not considering the relationship between non-contiguous words and domain sentiment words.

In different fields, the same word may be given different emotions and meanings, which leads to the inaccuracy of emotional analysis of comments by using the general sentiment dictionary. Therefore, building a domain-specific sentiment dictionary is a necessary foundation for sentiment analysis. Tang et al [4] used the construction of sentiment dictionary as the emotional classification task of words or phrase pairs, which improved the accuracy of classification, but the expression of emotional and semantic information was insufficient. Yang et al [5] used Word2vec to train corpus to obtain an emotional dictionary, which effectively utilized the semantic information of the word, but ignored the emotional information of the word. Lin et al [6] proposed a method of constructing domain sentiment dictionary based on word vector to realize the semantic and emotional representation of emotional words, but the domain adaptability is not considered and the classification accuracy is not high.
2. The Proposed Model

A document description is processed as an input and a classification result is generated as the output. As shown in Figure 1. Firstly, a model based on the attention mechanism of emotional word vector is established to capture the emotional and semantic information in the text; at the same time, the relationship between non-continuous words is modeled, and ACNN [7] is combined with Bi-GRU based on attention mechanism to generate document representation. Then, we establish a polarity classifier through the CNN to construct a domain sentiment dictionary. Finally, the linear combination layer is used to combine the document representation with the emotional features.

![Figure 1](image1.png)

**Figure 1.** Document-level sentiment analysis model.

2.1. Document Vector Representation Model

2.1.1. Emotional Word Vector Attention Model. By adding an emotional module to include emotional information, the emotional and semantic information are trained separately, as shown in Figure 2.

![Figure 2](image2.png)

**Figure 2.** The structure of sentiment word vector model based on the Skip-gram.

The word vector matrix is expressed as: \( X = [x_1, x_2, \ldots, x_m] \), \( X \in \mathbb{R}^{d \times |V|} \), where \( d \) is the dimension of word vector and \( |V| \) is vocabulary size. The main idea of the attention mechanism is to generate a context vector. The formula for the word vector \( k_i \) containing context information is as follows:
where, \( \text{score}(x_i, x_j) \) is computed by an artificial neural network, to model word pairs \((x_i, x_j)\).

Therefore, combining the context vector \( k_i \) with the original vector \( x_i \) by tensor product to form an extended vector \( m_i = x_i \otimes k_i \), and is regarded as an input to the ACNN, is conducive to generating structured representations.

2.1.2. The Joint Model of ACNN and Bi-GRU-Att. A+B-G-Att(ACNN+Bi-GRU-Att) model consists of ACNN and Bi-GRU model based on attention mechanism, as shown in Figure 3.

(1) Synthesizing sentences through ACNN

Suppose a sentence includes \( n \) words \( \{w_1, w_2, ..., w_n\} \). Using \( l \) as the width of the convolution filter to encode the semantic information. In order to make up for the defect that the traditional CNN only extracts local features, we follow the idea of ACNN [7], which can reduce the number of parameters. The input vector is convoluted by the filter to obtain the corresponding feature map, and the feature map is segmented by the channel to obtain the sentence vector feature.

(2) Document composition through Bi-GRU-Att model

The sentence is encoded by using Bidirectional GRU(Bi-GRU) [8], and the calculation process is as shown in (4)-(5):

\[
\bar{f}_t = \text{GRU} \left( o_t \right) 
\]
The Bidirectional GRU contains the forward GRU $\tilde{f}$ and a backward GRU $\tilde{f}$. Combining $\tilde{f}$ with $\tilde{f}$ to get historical and future information from both directions, $f_i = [\tilde{f}_i, \tilde{f}_i]$.

In order to emphasize sentences that are important for classification, self-attention mechanism is introduced, and the calculation process is as in formula (6)-(8):

$$y_i = \tanh(W_q f_i + B) \quad (6)$$

$$\alpha_i = \frac{\exp(y_i, r_i)}{\sum_{i=1}^{n} \exp(y_i, r_i)} \quad (7)$$

$$h^* = \sum \alpha_i f_i \quad (8)$$

Where, $B$ represents the offset, $W_q$ represents the weight, $\alpha_i$ is the attention feature matrix, $r_i$ is the sentence-level context vector, $h^*$ is the document vector feature.

2.2. Construction of the Domain-Specific Sentiment Dictionary

Based on the semantic knowledge base and corpus, CNN is used to construct a domain-specific sentiment dictionary. The construction process is shown in Figure 4.

![Figure 4. Construction of domain-specific sentiment dictionary.](image)

(1) The training of emotional word vector

The preprocessed domain data is input into the sentiment word vector model, and the dimension of word vector is set to 200. The semantic information is trained by the combination of hierarchical softmax and negative sampling, the emotional softmax layer trains emotional information, and the random error and backpropagation are used to pass the training error to each word vector.

(2) CNN polarity classifier

Unsupervised and supervised learning models require manual intervention, which has an adverse effect on the classification results, while deep learning techniques can automatically extract features. It
is known that the training corpus is emotional words, and the output is emotional polarity. CNN is used to extract local features and important feature information.

(3) Construction of domain sentiment dictionary
Firstly, word frequency statistics are performed on domain-specific corpora, and the general sentiment dictionary is deduplicated and fused to obtain an emotional word set. Secondly, the most influential sentiment in SentiNetWord is selected as the seed word. Then, the n words that are most similar to the seed word are calculated as the candidate sentiment words through the semantic similarity. Finally, the candidate sentiment words are input into the CNN model, producing the emotional polarity, which added to a specific domain dictionary.

2.3. Linear Binding Layer
Using the principal component analysis method to determine the weight. Firstly, the linear binding layer is used to combine the document vector features $h^*$ with the emotional vector features $e$, obtained by the specific domain sentiment dictionary to form a document classification feature $O$. Then, the softmax is used to receive the document tensor to produce the emotional category $y$, and the calculation formula of $O$ and $Y$ is as shown in (9)-(10).

$$ O = [h^* \oplus e] $$

$$ y = \text{softmax}(W_oO + b) $$

Where, $\oplus$ is splicing operation, $W_o$ is weight, $b$ is offset.

3. Experiment and Result Analysis
The entire model is trained end-to-end with stochastic gradient descent to determine the parameters of the neural networks, where the loss function is the cross entropy error of supervised sentiment classification.

3.1. Data Sets
The data set we use is the IMDB [9] and Yelp data sets [1]. We use 80% of the data for training, 10% for validation, and the remaining 10% for test.

3.2. Experimental Parameter Setting
In the experiment, all word vectors are initialized by Skip-gram, and other unregistered words are initialized by uniform distribution $U \sim (-0.01, 0.01)$. The parameters used by ACNN are shown in Table 1.

| parameter | Parameter description         | value |
|-----------|-------------------------------|-------|
| f         | Window size                   | 3,4,5 |
| n         | Number of convolution kernels | 64    |
| d         | Dropout                       | 0.5   |
| m         | Mini_batch                    | 32    |

3.3. Experimental Results and Analysis
We use accuracy ($A$) and $MSE$ as evaluation metrics. $MSE$ Measures the error between the predicted emotion value and the real emotion value. The calculation formula is as follows (11), the experimental results are shown in Table 2.
Table 2. Classification results based on different classification methods.

| Method          | dataset | IMDB     | Yelp 2013 | Yelp 2014 | Yelp 2015 |
|-----------------|---------|----------|-----------|-----------|-----------|
| Conv-GRNN       |         | 0.425/2.71 | 0.637/0.56 | 0.655/0.51 | 0.660/0.50 |
| LSTM-GRNN       |         | 0.453/3.00  | 0.651/0.50  | 0.671/0.48  | 0.676/0.49  |
| HN-ATT          |         | 0.494/2.32  | 0.682/0.48  | 0.705/0.42  | 0.710/0.47  |
| A+B-G-ATT       |         | 0.531/2.34  | 0.703/0.32  | 0.728/0.34  | 0.741/0.36  |
| A+B-G-ATT+DSL   |         | 0.558/1.62  | 0.734/0.27  | 0.762/0.24  | 0.776/0.28  |

It is found from Table 2 that the A+B-G-Att and A+B-G-Att+DSL models achieve higher accuracy and lower error rate than other neural network models. The LSTM-GRNN model has higher accuracy than CNN-GRNN because the recurrent neural network is good at processing sequence data, while CNN is good at extracting local features. The accuracy of HN-ATT model with attention mechanism is higher than that of Conv-GRNN and LSTM-GRNN models, which captures important features of document sentiment classification.

Comparing DSL, A+B-G-Att and A+B-G-Att+DSL, the results are shown in Figure 5. It can be seen that the accuracy of DSL is lower than the other two. It can only extract the emotional features of words, ignoring the semantic relationship between words and between sentences. Compared with DSL, the A+B-G-Att model achieves a high accuracy rate, which reflects the role of deep learning technology in automatically extracting features. However, since the domain sentiment words are not introduced, the accuracy is lower than A+B-G-Att+DSL. A+B-G-Att+DSL is based on the A+B-G-Att model. It introduces domain-specific sentiment words, and the experimental results achieve higher classification accuracy and lower error rate.

In order to further verify the role of the domain sentiment dictionary in sentiment analysis, based on the A+B-G-Att model, the SentiWordNet and the domain sentiment dictionary were introduced respectively. The comparison results are shown in Figure 6.
4. Conclusion
A document-level sentiment analysis model based on domain-specific sentiment words is constructed. This model captures the relationship between non-contiguous words in a sentence and introduces domain sentiment words, which makes the extracted features domain and emotional, enriching the classification characteristics of the document. The proposed model is compared with the neural network model to prove the validity of the proposed model, which can be widely applied to various specific fields such as public health. In the future, we will design a method of more reasonable candidate sentiment word extraction, and measure the similarity of emotional words in a deeper level.

5. Acknowledgments
The work is supported by the National Basic Research Program of China under grant No.2014CB340404, the National Natural Science Foundation of China under grant No.61702305 Scientific Research Foundation of Shandong University of Science and Technology for Recruited Talents under the grant No.2016RCJ045.

6. References
[1] Tang D, Qin B and Liu T. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification[C]// Conference on Empirical Methods in Natural Language Processing. 2015:1422-1432.
[2] Yang Z, Yang D, Dyer C, He X, Smola A and Hovy E.Hierarchical Attention Networks for Document Classification[C]// Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016:1480-1489.
[3] Zhang R, Lee H, Radev D, Dependency Sensitive Convolutional Neural Networks for Modeling Sentences and Documents[J].Proceedings of NAACL-HLT 2016:1512-1521.
[4] Tang D Y, Wei F R, Qin B, Zhou M and Liu T. Building large-scale twitter-specific sentiment lexicon:a representation learning approach. In:Proceedings of the 25th International Conference on Computational Linguistics. Dublin, Ireland:COLING, 2014. 172-182.
[5] Yang X, Zhang Z, Wang L, Zhang Y, Ma Q,Wu J and Zhang Y.Automatic Construction and Optimization of Sentiment Lexicon Based on Word2Vec [J].Computer Science, 2017, 44(1): 42-47.
[6] Lin J, Zhou Y,Yang A and Chen J.Building of domain sentiment lexicon based on word2vec[J]. Journal of Shandong University (Engineering Science), 2018, 48(03):40-47.
[7] Liang D, Zhang Y. AC-BLSTM: Asymmetric Convolutional Bidirectional LSTM Networks for Text Classification [J]. 2017.
[8] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
[9] Qiming Diao, Minghui Qiu,Chao-Yuan Wu,Alexander j Smola,Jing Jiang,and Chong Wang 2014 In proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining vol2, pages 193202.ACM.