Text classification based on graph convolutional network with attention

Genmei Tong, Min Zhu*
School of Computer Science and Technology, East China Normal University, Shanghai, 200062, China
School of Data Science and Engineering, East China Normal University, Shanghai, 200062, China
51184506040@stu.ecnu.edu.cn, *Corresponding author’s e-mail: mzhu@cc.ecnu.edu.cn

Abstract. The paper proposes a text classification method based on the graph convolutional network, the traditional text classification problem can be transformed into the node classification problem in the graph, and then proposes a attention structure to increase the weight of certain words and phrases. This paper verifies that the above structure has a positive effect on text classification accuracy. In particular, the structure can handle irregular data, such as citation network. We obtain very competitive results on 5 commonly used text classification datasets and achieve state-of-the-art results on 4 datasets, which include one dataset of citation network. Experiments show that combination of both structures can significantly reduce the classification error.

1. Introduction
Text classification is an important task in natural language processing (NLP) and has broad applications in sentiment analysis, question-answer systems, and spam detection[1]. In the area of deep learning, text classification methods based on convolutional neural networks (CNN) is the most commonly used methods. Johnson and Kim were the first to use CNNs for text classification and achieved better results than traditional text classification methods. But these methods only consider the importance of phrases through convolution kernels and pooling operations of different sizes. Therefore these methods don't have the ability to understand context[2-3]. To solve the problem, Duque et al. proposed the use of deeper convolutional neural networks for text classification. Though the method has the ability to understand contextual information, there number of parameters will increase and training process will be harder[4-6]. Since CNN has limited ability to obtain long sequence information, and recursive neural network(RNN) is good at doing similar work, RNN is also used to do text classification[7-8].

With the development of image convolution in the field of image and good results achieved, some graph-based learning algorithms have been applied in the field of text, such as textGCN proposed by Yao et al., and SGC proposed by Wei et al., both of which used image convolution to classify text[9-10]. However, they did not consider the relationship between texts, but in some cases, the relationship between texts should be considered. If Articles that refer to each other, then they are likely to belong to the same domain.
In order to consider the relationship between text and text, expand the application scope of the model, this paper proposes a text classification method based on the graph convolution network and the nodes of the graph are used to represent the text. The associations between text and text are contained in the edges of the graph. At the same time, the attention mechanism is used to assign different weights to different information, to highlight key information. Finally, experiments on different datasets prove that this model is conducive to improving the accuracy of text classification.

2. Model

The whole structure of the model we proposed consists of two parts: graph convolutional network and attention mechanism. We will cover each part of the model in detail below.

2.1. Graph convolutional network

The basic process of semi-supervised text classification proposed in this paper is to first find the relationship between text and text, text and words, and learn the graph structure data. For citation network, the implied relationship between texts can be better learned, and the model structure is shown in the figure below.

![Graph convolutional network for text classification](image)

In Figure 1, the color nodes represent text, the gray nodes represent words, the straight line represents the relationship between nodes, and the curve shows the updating of node features, namely the learning process. After passing through the hidden layer in the middle, a new matrix is output to represent the category of the text. In the training process, the classification error on the marked text set is trained through gradient descent to finally achieve the purpose of classifying the unmarked text. Meanwhile, the feature representation of words is synchronized with the training update. Where, for a node, in the convolutional layer of the graph, the update rule of the node is:

\[
X^{k+1} = \delta(AX^k\Theta^k)
\]

\[
X_i^{k+1} = \delta(\sum_{j\in N_i} \frac{1}{d_{ij}} X_j^k\Theta^k)
\]

In the equation(1), \(\delta\) is the activation function relu, \(X_k\) represents the input characteristics of layer \(K\), \(\Theta^k\) is the filter parameter matrix with training at the layer \(k\). In the equation(2), \(N_i\) is all neighbor nodes of node \(i\), and \(d_{ij}\) is used to represent the relationship between node \(i\) and node \(j\).

2.2. Attention mechanism

Similar to the application of textGCN in the field of text classification, the simple average pooling operation will cause the loss of the information, which is between the features of each node and the nodes. Therefore, we propose to use the attention mechanism in the graph convolutional network to solve this problem. Specific as follows:
In the equation (3), $N$ is the information matrix of nodes and $D$ is the information matrix of edges. $V$ is the parameter matrix. Attention increases the weight of some words and phrases, which further improves the accuracy of classification.

3. Experiment and Test Results

3.1. Baselines

- char-level CNN/word-level CNN: Zhang designed a nine-layer convolutional neural network for text classification. It explores entering text at the character level and word level[11].
- VDCNN: VDCNN is a network composed of 29 convolutional layers, through which deeper text representations can be obtained. It has achieved very good experimental results on large datasets.
- DPCNN: This is a deep neural network with 15 convolutional layers. By using down sampling, the training time of the model can be reduced.
- LEAM: This model first proposes a new method of word embedding for classification tags and a novel attention mechanism.
- GCN: This model applies graph convolution to the text domain and achieves good results on multiple datasets with references between texts.

3.2. Main results

The whole process of building a model is like building blocks. We first show that each module has a positively impact on the classification accuracy. Then we put each module into a large model in a proper way. For the fairness of the experiment, the experimental parameters and conditions adopted in all the models are consistent. Five datasets were used in the experiment, as shown in the following table.

Table.1 Datasets information. Here Critics refers to film reviews, QA refers to question answering. #s denotes the number of sentences and c is the number of target classes.

| datasets | tasks    | train #s | test #s | c  |
|----------|----------|----------|---------|----|
| MR       | Critics  | 130k     | 10k     | 2  |
| AG       | News     | 120k     | 7.6k    | 4  |
| DBP      | Ontology | 560k     | 70k     | 14 |
| Yah.A    | QA       | 1.4M     | 60k     | 10 |
| Cora     | Papers   | 1.5M     | 70k     | 7  |

Our graph convolutional network (CoGCN): We apply graph convolution to text classification and consider the correlation information between text and text on the basis of graph convolution.

Table.2 Error rate(%) on five date sets for CoGCN

| Models | MR | AG | DBP | Yah.A | Cora |
|--------|----|----|-----|-------|------|
| CoGCN  | 8.13 | 6.37 | 0.87 | 23.79 | 20.64 |

Attention mechanism: The attention mechanisms in both nodes and edges are described above. We applied the attention mechanism to the nodes of the graph and the edges of the graph respectively. In the subsequent experiments, we applied the attention mechanism to both nodes and edges. The results are as follows.
Table 3 Error rate(%) on five data sets for ATTs

| Models  | MR   | AG   | DBP  | Yah.A | Cora |
|---------|------|------|------|-------|------|
| ATT-N   | 7.99 | 7.67 | 0.97 | 24.99 | 21.92|
| ATT-E   | 8.01 | 7.82 | 1.01 | 25.03 | 21.75|
| ATT-E&N | 7.95 | 7.48 | 0.94 | 24.76 | 21.21|

In a previous paper, we experimentally verified that different components of the whole model improve the text classification accuracy. So we combined all the structures into a more complete model. For comparison purposes, table 4 lists the error rates of different models on the test set. We use bold font to highlight the best experimental results. We see that on all five datasets, the model achieves the best results on four of them. Very competitive results are also obtained on the remaining one data set.

Table 4 Error rate(%) of the combined model on five datasets

| Models        | MR   | AG   | DBP  | Yah.A | Cora |
|---------------|------|------|------|-------|------|
| word-level CNN| 13.54| 8.55 | 1.37 | 28.84 | 29.81|
| char-level CNN| 13.23| 9.51 | 1.55 | 28.8  | 27.78|
| VDCNN         | 10.89| 8.67 | 1.29 | 26.57 | 27.69|
| DPCNN         | 9.88 | 6.87 | 0.88 | 23.9  | 25.67|
| LEAM          | 9.72 | 7.55 | 0.98 | 22.58 | 23.78|
| CoGCN-ATT     | 7.93 | 6.32 | 0.82 | 22.9  | 18.74|

Although each component did not perform well in the text categorization task, we combined all the structures to achieve very competitive results. Our model can address a wider range of data types, and at the same time consider the relationship between text and text, increasing the accuracy of classification.

4. Conclusion

In this paper, we propose a new graph convolutional network and a attention mechanism. After verifying that these two methods are helpful to reduce the error rate of text classification, we combine these two methods to build a new text classification model. We proved that this new model, which combines the two structures proposed in this paper, can significantly reduce the error rate of text classification. Compared with other deep CNN models, in this paper, convolution is generalized to the graph structure, and the relationship between texts is correlated, while the accuracy of classification is improved by using the attention mechanism, which proves that combining different structures in an appropriate way can be more effective than repeating a single structure.

References

[1] Wang, G.; Li, C.; Wang, W.; Zhang, Y.; Shen, D.; Zhang, X.; Henao, R.; and Carin, L. 2018. Joint embedding of words and labels for text classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2321–2331. Association for Computational Linguistics.

[2] Johnson, R., and Zhang, T. 2014. Effective use of word order for text categorization with convolutional neural networks. arXiv preprint arXiv:1412.1058

[3] Kim, Y. 2014. Convolutional neural networks for sentence classification. In EMNLP, 1746–1751.

[4] Johnson, R., and Zhang, T. 2017. Deep pyramid convolutional neural networks for text categorization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), volume 1, 562–570.

[5] Conneau, A.; Schwenk, H.; Barrault, L.; and Lecun, Y. 2017. Very deep convolutional networks
for text classification. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, 1107–1116. Association for Computational Linguistics.

[6] Andréa B. Duque, Luã Lázaro J. Santos, David Macêdo, Cleber Zanchettin. 2019. arXiv preprint arXiv: 1901.09821.

[7] Lai, S.; Xu, L.; Liu, K.; and Zhao, J. 2015. Recurrent convolutional neural networks for text classification. In AAAI, 2267–2273.

[8] Zhou, P.; Qi, Z.; Zheng, S.; Xu, J.; Bao, H.; and Xu, B. 2016. Text classification improved by integrating bidirectional lstm with two-dimensional max pooling. In COLING, 3485–3495.

[9] Yao, L.; Mao, C.; and Luo, Y. 2019. Graph Convolutional Networks for Text Classification. In AAAI, 7370-7377.

[10] Wu, F.; Zhang, T.; Amauri Holanda de Souza Jr.; Fifty, C., Yu T., Kilian Q. W. 2019. arXiv preprint arXiv: 1902.07153.

[11] Zhang, X.; Zhao, J.; and LeCun, Y. 2015. Character-level convolutional networks for text classification. In NIPS, 649–657.