A Hybrid GCN and RNN Structure Based on Attention Mechanism for Text Classification

Lingchao Gao, Jiakai Wang, Zhixian Pi, Huaixun Zhang, Xiao Yang, Peizhuo Huang and Jiasong Sun

State Grid Corporation of China Big Data Center, Beijing, China.
E. E. Department, Tsinghua University, Beijing, China.
Email: gaole2020@tom.com

Abstract. In the field of deep learning, for problems and tasks that are sensitive to time series, such as natural language processing or speech recognition, the recurrent neural network is usually more suitable. Long short-term memory (LSTM) is a representative network structure in recurrent neural network. It is time-dependent and enables a global representation of features. However, some problems such as the network parameters of LSTMs limit the applicability of their solutions. This paper proposes an improved hybrid structure of graph convolutional neural network and recurrent neural network. In the input layer, a two-dimensional convolutional neural network is used to build a text corpus map. Graphic embedding is used to preserve the global structure of the entire text graph structures. The LSTM layer and attention mechanism are used to fully implement text classification and improve the computational efficiency. The test results show that the hybrid network structure has better operation speed on the IMDb dataset.

1. Introduction

Natural language processing is a very expected interdisciplinary subject that involves computer science, linguistics and its representation, artificial intelligence, psychology and cognition, and so on. The highest goal of natural language processing is teach machines to acquire, understand, and produce methods that can communicate with humans in the form of language. Common machine translations, Chatbot and human-computer conversations are all the applications of natural language processing. These technologies and applications are trained on a large amount of network text. These applications constantly learn and build human language patterns, which are stored in the form of computer programs, which could make human-computer interaction and understanding more convenient.

In the past, the machine learning methods, such as the well-known support vector machine, required manual input of many dimensional features, and these features were often sparse to obtain effective natural language processing results. Advanced natural language processing technologies, including deep learning, are helping computers perform a variety of natural language tasks, such as making it possible to analyse a sentence on Google in less than a second [1, 2].

For the past few decades, classical machine learning techniques such as decision trees, Bayesian classifiers, and support vector machines have been used for natural language processing. A specific application example is to classify words or sentences as positive or negative by training a machine learning model. It is obvious that due to the lack of standard data sets at the time, and the emotional annotation of words or sentences is very rough, the performance drops sharply when encountering text content with different language types or different topics. This kind of dataset is lacking and different. It could be mitigated through semi-supervised machine learning such as LDA. The unsupervised
method starts from the dictionary to score the emotions of words or sentences, and finally evaluates the positive or negative emotions of the entire document. Both semi-supervised and unsupervised can deal with insufficient data and labelling.

The popular natural language processing toolkits include [3]:

- NLTK: A set of libraries and programs for English symbols and statistical natural language processing. It was developed by the University of Pennsylvania and includes graphical presentations and sample data.
- OpenNLP: A development toolkit for natural language text processing based on machine learning. It supports common tasks in natural language processing, such as tokenization, part-of-speech tagging, shallow analysis and grammatical analysis. These tasks usually require more advanced word processing services.
- FudanNLP: A Chinese natural language processing toolkit developed by the Natural Language Processing Group of Fudan University, including Chinese word segmentation, named entity recognition, dependency parsing, keyword extraction, news clustering, online learning, etc.
- CoreNLP: An important and easy-to-use toolkit in the field of natural language processing. It integrates basic tools such as tokenization, lemmatization and parsing. It also includes advanced functions such as sentiment analysis and named entity recognition. And the toolkit includes multiple languages.
- AllenNlp: AllenNLP is a deep learning-based NLP toolkit based on pytorch. It has the advantages of ultra-modularity, light weight and easy expansion [4].

Question answering applications are a typical application of natural language processing technology. Question answering are users asking questions, at the same time, system uses various information and resources (such as documents or knowledge maps) trying to answer this question. If there are multiple answers, the most likely answer is selected. Dynamic memory network [5] is a neural network used to process input sequences, which is very suitable for sequence input and output problems such as question answering. The basic principle is that the problem activates the processing mechanism first, and the previous input needs to be considered during activation, and then the end-to-end dynamic memory network is used for inference, and the question output in the form of a sequence is finally output. This method has been tested on common datasets: bAbI, Stanford Sentiment Treebank and WSJ.

In terms of choosing which neural network is the most suitable for natural language processing, recurrent neural network structures represented by LSTM structure and GRU structure have been widely used in the field of natural language processing. Compared with convolutional neural networks, they are used less. The reason is that the former uses sequence as the network structure feature, while the latter uses layer as the network structure feature. Therefore, for natural language processing applications of sequence input and output, convolutional neural networks are not so ideal. Some methods [6] try to transform this sequence task into a classification task, and then can use a gated-based convolutional neural network, and still use recursive neural networks in the language model part. Both network structures can serve classification tasks, and the main competition is whether more semantic understanding functions can be added. In addition, the size of the hidden layer and batch, as well as the learning rate, have some impact on performance.

Before the emergence of the BERT model [8], pre-trained expression has been a technical bottleneck, because the previous network models used unidirectional language models, so they were greatly restricted during pre-training. The BERT model is a model of language representation proposed by Google in 2018. It uses a pre-trained two-way representation and 2 unsupervised prediction methods to train, including masking 15% of the input randomly and introducing additional output layers, etc. It further improves the level of BERT in a wide range of natural language understanding tasks: the metrics on the standard test are beyond the level of humans.

However, most of the above-mentioned natural language processing methods still have problems such as relying on data volume, excessive network structure, and long training time. This paper proposes an improved two-dimensional GCN and RNN hybrid structure for text classification. The main method is to build a corpus text map based on words and documents and their relationships, and train the graph convolution network accordingly. The purpose is to preserve the global structure of the
entire text graph in the graph embedding, further describe the rich relationship between the texts, and make reliable text classification. Test results on the IMDb movie review binary sentiment classification dataset show that this network structure does not require much data, and the amount of parameters and training speed are acceptable. The structure of this paper is as follows: The second part introduces the basic principles of graph convolutional neural networks, the third part introduces the hybrid structure of GCN and RNN, the fourth part gives the test results on the IMDb data set, and finally the conclusions and next plans.

2. Graph Convolutional Network
To implement natural language processing applications such as question answering, a knowledge base capable of storing entities and relationships needs to be constructed, but the current knowledge base construction methods usually lose the original information in a large amount of data. This only requires building the codec model in the graph model and use softmax on each node. With the graph convolution network [9] shown in Fig1, it will recover the lost relationships or entities, which can also perform the task of entity classification. Test results on some challenge datasets show that this network structure can alleviate the problem of missing information in the knowledge base.

![Figure 1. Entity Update R-GCN Model](image1)

![Figure 2. Embedding and GNN](image2)

Although deep learning has achieved significant results in some areas of artificial intelligence, non-Euclidean domain data also exists extensively and needs to be analysed. For example, physical/molecular/biological model need to be described by nodes and their relationships with each other [11]. The widely used e-commerce recommendation system, also belongs to this type. Graphical modelling of these data is a good learning method. Graph neural networks can build graph models by passing information between nodes, but graph data has strong irregularities and high complexity. The existing deep learning methods generally assume that the nodes are related, so they are not suitable for this type Modelling. On the other hand, convolution, a technique often used in deep learning, cannot be used directly in graph modelling. The purpose of the network embedding method is to preserve the overall structure and node locations. The neural network shown in Fig.2 is the embedded method
using a deep neural network. Graph convolutional networks are mainly used to [12] capture node connections to surrounding connections and architecture dependencies.

In practical applications such as knowledge graphs, social networks and traffic flow, there is a large amount of data in the form of a graph structure [13]. Graph semi-supervised learning technology is used to deal with the situation where most nodes in the data have not been labelled, where these nodes are specific categories or labels. GCN is defined as follows: The definition graph is, each node feature is taken as the input, and the adjacency matrix of the association between nodes is \( A \). The goal is to output the feature matrix, which represents the learned or labelled information of the unlabelled nodes. The formula for two-layer GCN semi-supervised node classification is shown in Eq.1:

\[
Z = f \left( X, A \right) = \text{softmax} \left( \text{ReLU} \left( \hat{A}XW^{(0)} \right) W^{(1)} \right)
\]

Where \( W^{(0)} \) is input of hidden layer with \( H \) feature maps to hidden weight matrix. \( W^{(1)} \) is the hidden output weight matrix.

3. Hybrid Structure of GCN and RNN

Recursive neural network is mainly used to deal with the problem of sequence input and output [14]. It is a scalable deep neural network. Long short-term memory network (LSTM) (shown in Fig.4) is often used in recurrent neural networks, which feature is to be effect on the long-time input. LSTM is widely used in speech modelling, text recognition, image analysis and other fields.

In general, the standard LSTM has the problem that it cannot distinguish which features are important parts of the classification. Another question is the difficulty of vanishing or exploding gradients. Attention-based LSTM [15], which is shown in Fig.5, address this problem by introducing attention to increase the weight of important parts that affect classification. Attention LSTM takes attention as a measure of weight and pays more attention to the features with greater classification impact. The method is to associate attention weight to each input vector so that the output also carries this information.
As mentioned earlier, no matter whether it is the GCN method or the RNN method, it still cannot solve the problems of large network structure and long training time, and the performance is not good under the condition of fewer labels and data. This paper proposes an improved two-dimensional graph convolutional neural network GCN and RNN hybrid structure, which is implemented on text classification tasks. To construct and train graph structure representations based on words, documents and relationships, a network model based on GCN and attention LSTM is generated, so that the long-term relationships and structure of the entire text graph can be retained in graph embedding to deeply describe textual relationships and perform reliable text categorization. As shown in Fig.6, the first layer of this hybrid architecture is the GCN structure, which included an input-output layer and multiple hidden layers. The middle and upper layers are the LSTM layer and the embedding layer, and both the GCN layer and the embedding layer carry their attention weight. Compared with attention LSTM, this structure reduces the number of parameters at the GCN layer, and uses the attention mechanism on both the GCN and LSTM layers. When processing sequence samples with a sufficient length of time, the attention models are fully used by aggregating information from past times into more abstracting output vectors.

![Figure 5. The Attention-based LSTM](image)

![Figure 6. GCN AND Attention LSTM Hybrid Architecture](image)

4. Experimental Results

In order to demonstrate the effectiveness of this method, the IMDB movie review binary sentiment classification dataset was selected [16], which is to be the benchmark for text classification. It is a widely processed and evaluated natural language processing dataset, which is mainly used for sentiment emotion classification. It contains 50,000 IMDB reviews, specifically for sentiment analysis, half for training and half for testing. There are also unlabelled documents of the same size for
unsupervised learning. The sentiment of the review is binary, which means that the IMDB rating of less than 5 means a score of 0, and the sentiment of a rating greater than means a score of 1.

Table 1. Performance of Three Structures

| Approach        | Condition_1 | Condition_2 | Condition_3 |
|-----------------|-------------|-------------|-------------|
|                 | Mean Val-acc| Cost time(s)| Mean Val-acc| Cost time(s)| Mean Val-acc| Cost time(s)|
| LSTM            | 87.5%       | 930         | 86.2%       | 1891        | 84.0%       | 4717        |
| Attention LSTM  | 88.5%       | 1260        | 87.3%       | 2175        | 86.4%       | 6394        |
| Hybrid structure| 88.0%       | 642         | 87.7%       | 1079        | 87.1%       | 2175        |

In order to compare the accuracy and time performance of the general LSTM, Attention LSTM and the proposed Hybrid structure, we tested under three conditions:

✔ Condition 1: Training data vs. testing data is equal to 9:1.
✔ Condition 2: Training data vs. testing data is equal to 4:1.
✔ Condition 3: Training data vs. testing data is equal to 1:1.

Where data were split into training set and test set randomly. When initializing the data, first the sentences were divided into words, some non-ASCII characters were filtered, and uppercase letters to lowercase were converted. Then data was serialized and unified into unique length, and the words are mapped into indexed representations. In all three methods, the 50% Dropout method is used, and the LSTM parameters are initialized with random values. The classification accuracy and time cost of three network structures are shown in Tab.1.

From the test results of the three methods in Table 1 under three conditions, it can be seen that the Attention LSTM method has the highest accuracy under Condition 1; the hybrid structure has the highest accuracy under Condition 3. In terms of time cost, the hybrid structure has the shortest test time cost due to the use of attention mechanism parameters on multiple layers and GCN structure on the input layer.

5. Conclusion

As an effective sequence-to-sequence deep neural network, LSTM and its derivatives are widely used in natural language processing related fields. However, realistic problems such as the large scale of the network structure and high time cost still exist. This paper proposes an improved GCN and RNN hybrid structure. By establishing a corpus text map, the purpose is to reduce the number of parameters while retaining the global structure of the entire text graph in the graphics embedding, and to perform reliable text classification and shorten the operation time. The test results on the IMDB movie review dataset show that compared with two kinds of LSTM networks, the improved hybrid structure has a consistent level of classification accuracy and has a better operation speed.

6. Acknowledgments

This paper is supported by the State Grid Big Data Center Self-built Technical Project: Research on intelligent data resource catalog construction and business feature Traceability Technology (5500/2019-83001B).

7. References

[1] Hirschberg J and Manning C D. Advances in natural language processing. Science, 2015, 349(6245), pp 261-266.
[2] Young T, Hazarika D, Poria S, et al. Recent trends in deep learning based natural language processing. IEEE Computational intelligence magazine, 2018, 13(3), pp 55-75.
[3] Sun S, Luo C, Chen J. A review of natural language processing techniques for opinion mining systems. Information fusion, 2017(36), pp 10-25.

[4] Gardner M, Grus J, Neumann M, et al. AllenNLP A deep semantic natural language processing platform. arXiv preprint arXiv 1803.07640, 2018.

[5] Kumar A, Irsoy O, Ondruska P, et al. Ask me anything Dynamic memory networks for natural language processing. International conference on machine learning. 2016, pp 1378-1387.

[6] Yin W, Kann K, Yu M, et al. Comparative study of CNN and RNN for natural language processing. arXiv preprint arXiv 1702.01923, 2017.

[7] Chiu B, Crichton G, Korhonen A, et al. How to train good word embeddings for biomedical NLP. Proceedings of the 15th workshop on biomedical natural language processing. 2016, pp 166-174.

[8] Devlin J, Chang M W, Lee K, et al. Bert Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv 1810.04805, 2018.

[9] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks. European Semantic Web Conference. Springer, Cham, 2018, pp 593-607.

[10] Zhou J, Cui G, Zhang Z, et al. Graph neural networks A review of methods and applications. arXiv preprint arXiv 1812.08434, 2018.

[11] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2d knowledge graph embeddings. Thirty-Second AAAI Conference on Artificial Intelligence. 2018, pp 1811-1818.

[12] Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks. arXiv preprint arXiv 1901.00596, 2019.

[13] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv 1609.02907, 2016.

[14] Greff K, Srivastava R K, Koutník J, et al. LSTM A search space odyssey. IEEE transactions on neural networks and learning systems, 2016, 28(10), pp 2222-2232.

[15] Wang Y, Huang M, Zhao L. Attention-based LSTM for aspect-level sentiment classification. Proceedings of the 2016 conference on empirical methods in natural language processing, 2016, pp 606-615.

[16] Maas A L, Daly R E, Pham P T, et al. Learning word vectors for sentiment analysis. Proceedings of the 49th annual meeting of the association for computational linguistics Human language technologies-volume 1. Association for Computational Linguistics, 2011, pp 142-150.