Research on Social Comment Text Classification Based on BERT and Graph Neural Network

Xingyu Zhao [1, a], Jian Zheng [1, b, *]

[1] School of Computer and Information Engineering, Institute for Artificial Intelligence, Shanghai Polytechnic University, Shanghai 202306, China

[a] zxycjcjdsg@gmail.com, [b] zhengjian@sspu.edu.cn

Abstract. In the era of information data explosion, accurately classifying massive text data and mining emotional tendencies play an important role in purifying the network environment. A sentiment classification model that combines the advantages of BERT and graph neural networks is proposed to address the shortcomings of a single traditional classification model in terms of classification accuracy and efficiency. It is recommended to combine the transformer based BERT model with graph neural networks, using the text embedding technology in BERT to leverage its powerful feature extraction ability and the advantage of graph neural networks in aggregating semantic features of text structure information. This will advance the semantic information of the entire text and then pass the semantic features to the combined model for classification. At the same time, in text sequence processing, it is proposed to construct an emotional dependency tree and obtain a leading matrix based on dependency relationships, which can improve the model's understanding of text semantic and structural information. The model was tested on English datasets such as 20NG and R8, as well as the Weibo 100k dataset, with a particularly high rate of 93.7% on the Weibo 100k dataset. Comparative experiments were conducted with other models.The results indicate that compared with other segmentation models, our model has better segmentation performance and can effectively aggregate semantic information of comment texts.

Keywords: emotion text classification, BERT, graph neural network, semantic information.

1. Introduction

In today's era of explosive growth of information data, a huge amount of textual data is generated and disseminated, such as social media posts, comments, etc. In the face of this massive amount of text data, it becomes crucial to accurately categorise and understand it. These social forums and apps provide features such as social commenting and speaking[1]. Often hidden in these intricate comments is the feedback of the commenter on the subject of the comment, and mining to determine their emotional tendencies can purify community posts from bad speech and improve strategies. Negative speech, social network bullying these things have become a very acute problem, do an analysis system to detect these contents will greatly play a role in purging the cyberspace is a very good way. Sentiment classification is an important task in the field of natural language processing, aiming to classify text into sentiment categories such as positive, negative or neutral. Sentiment classification is widely useful and influential in applications such as social media analysis, opinion monitoring, and user comment analysis. With the popularity of social media and the generation of large-scale text data, sentiment classification techniques have become increasingly important. Deep learning has made significant progress in sentiment classification tasks. Using deep learning models such as Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and pre-trained language models (e.g., BERT, GPT, etc.), semantic and contextual information in text can be captured to improve the accuracy and generalisation of sentiment classification[2].

Text classification, derives rules from the training set that can classify text, and classifies unknown documents based on this rules[3]. Manual classification consumes a lot of human, material and financial resources and is inefficient, which is far from meeting the needs of obtaining classification information in various fields. Using artificial intelligence models for classification can complete the task more efficiently, reduce unnecessary waste of resources, achieve a more accurate and robust sentiment classification system, and further promote the development of social media analysis and user sentiment understanding[4].The contributions of this article are as follows:

Capture the rich sentiment expression, get the semantic features of the comments, combined with BERT can be very good to capture the semantic features and the information of the next sentence and the huge amount of data to deal with, as well as the graph neural network better steps the neighbouring nodes of the multiple aggregated information, you can get a better semantic features, and then combined with the pre-training of the BERT and thus a better classification task[5].

2. Method

2.1 Representations of social comments

This experiment needs to process the text data into the type that the model can accept, so how to combine the techniques to characterise the comment text into the model needs is a problem to be solved. In BERT, the text is disambiguated, encoded, and combined with the self-attention mechanism of the transformers technique bi-directional encoding while considering the contextual information of the before and after positions[6]. In the pre-training phase, the Masked Language Model (MLM) task masks the words of the text and the Sentence Prediction (NSP) task is used to learn the bi-directional text representation. For representing text as a matrix with graph structure information, dependency relations are generated for words or words using sentiment dependency trees to learn from generation to semantic information input. Firstly, the dependency analysis is performed on the sentence using the dependency syntactic analysis tool to get the structure and label information of the dependency tree, secondly, the sentiment polarity is marked, and then the sentiment tree is constructed to form the nodes of the dependency tree, and then the nodes are merged immediately after that, and the nodes are passed upwards according to the sentiment polarity marking up to the root node, and then the parent-child node is decided according to the dependency relationship, and then the neighbour matrix is derived for the graph structure at this point. Of course, there are many other technical routes that can be used for characterisation, and this study proposes this method for the characterisation of social comments has the effect of effectively extracting emotional semantic information[7].

2.2 Based on BERT and graph neural network models

Graph Convolutional Neural Network (GCN) can effectively learn the graph structural information based on the graph after the above characterisation, and then carry out the propagation of information on the graph data and the aggregation of features with its neighbouring nodes, while the GAT also learns the correlation between the node attributes by the Attention mechanism, and adopts the Multi-Head Attention output for the aggregation to improve the classification ability, and the GCN is characterised by extracting the graph features of structural information, and after the above representation approach, the semantic information of the comment text can be analysed efficiently, and the addition of convolution and pooling layers can extract important information and obtain semantic information at different levels of abstraction[8]. Weights are assigned between these two modules to combine the outputs. The GCN propagation rules are shown below.

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)$$

Figure2.1 GCN propagation rules

Combining the word vectorisation and two-way attention mechanism of BERT, the role of each word in a sentence on the context word is fully considered, as well as capturing long-distance semantic information in different contexts of the same word, and the semantic information obtained from pre-training is put into the architecture of graph neural network, which goes through the hidden layer to get the classification results. In dealing with combining the two, the weight update combining in the same way can give better results[9]. BERT-base pre-training is as follows.

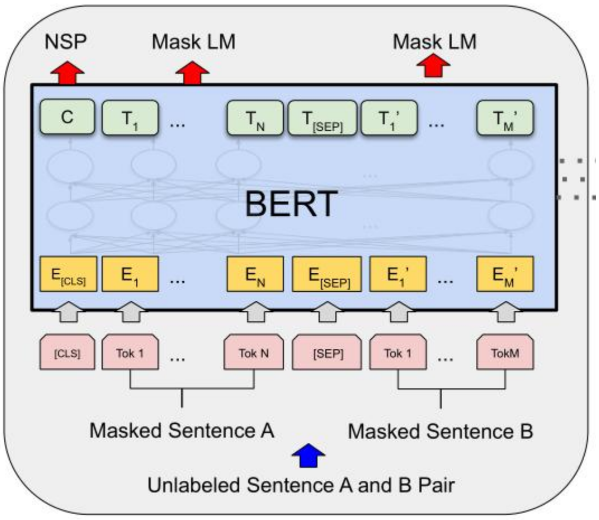

Figure 2.2 BERT-base pre-training

3. Experimental results and analysis iterature References

3.1 Data preprocessing

In this experiment, representative as well as scalable datasets are selected for experimentation, including public datasets with multiple metrics as well as representative real data for experimentation. Such as IMDB dataset, Amazon dataset, SemEval sentiment analysis dataset, MR dataset, etc., and WeiboSenti100K, this study mainly proceeds on the dataset of sentiment classification of comments on social networks.

The study will collect data about comment texts, including existing datasets as well as web crawlers to obtain comment data on major social apps or forums. The collected data will be pre-processed, including segmentation, addingtokens before the segmented sequences for overall sequence classification or sequence level tasks. A marker is added between two sentences in the sequence . To accommodate the fixed-length input of the BERT model, sequences that are too long need to be truncated, and sequences that are too short need to be filled with special markersat the end to achieve the required length. The preprocessed data will be used as input for subsequent studies by converting the segmented words into their unique identification IDs in the BERT vocabulary, as well as by ignoring deactivated words[10].

This experiment first analyses the Chinese text dataset weibo_senti_100k to understand the structure within the dataset and visualise the classification percentage. The data processing part firstly divides the document list as well as the content of the documents, 9:1 for the training set and validation set, and 2:1 for the training set and test set, and disrupts the order of the documents in the training set and test set. After that, composition operation is performed on the dataset, documents are segmented using jieba, vocabulary lists are created, and word frequencies are counted. TF-IDF features of the documents are computed using TfidfVectorizer. Extract all the tags from the metadata of the document, convert the tags to One-Hot encoding form and apply them to the training and test sets. Then proceed to construct the training set feature matrix, calculate the mean of the word vectors for each training document and construct the sparse matrix x and label matrix y, similarly for the test set and finally create a sparse matrix allx and label matrix ally that includes all the words in the training set documents and vocabularies.Finally on the graph construction, the words in the documents are divided into multiple sliding windows according to the size of the window, which are used to compute the co-occurrence of word pairs and also compute the word pair frequency :, compute the frequency of occurrence of word pairs in the window and compute the PMI (Point Mutual Information) of the word pairs and construct the adjacency matrix, using the PMI values to construct the graph adjacency matrix for the documents and vocabulary. fig. 1 is a representation of the graph construction with more adjacency edges selected (only some of the subgraphs are shown):

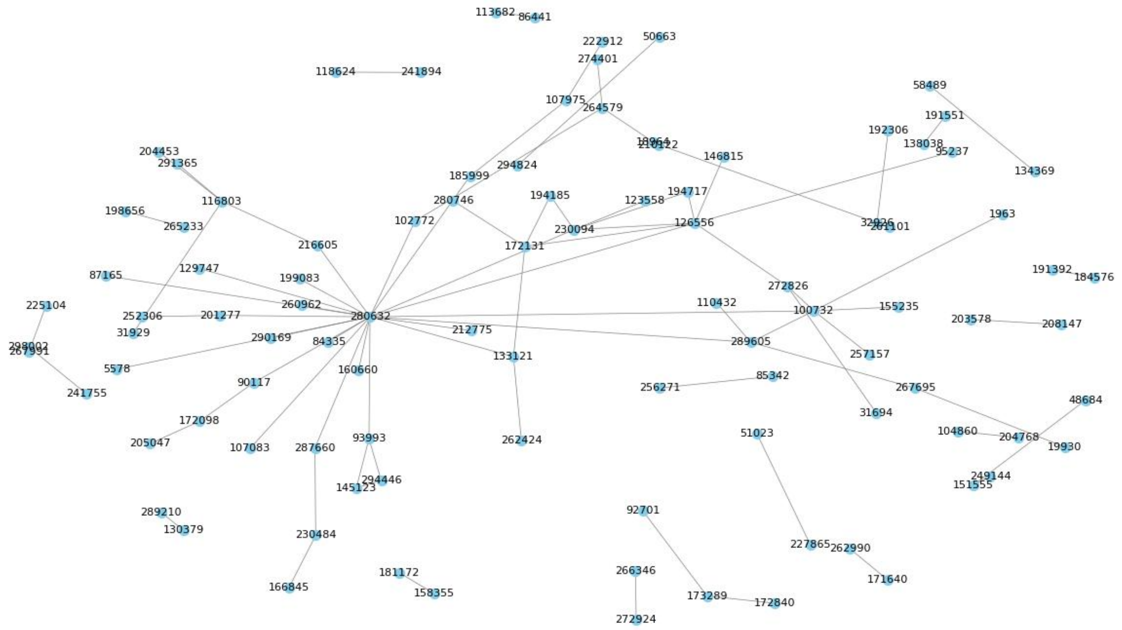

Figure3.1 subgraph

3.2 Segmentation Results

The following data compares this paper with other classical models on each model.

Table 1. Comparison of model analysis results

Model	20NG	R8	MR	R52	Ohsumed
BertGCN	89.3	97.8	86.0	96.6	72.8
RoBERTaGCN	89.5	97.9	89.7	96.1	72.8
BertGAT	87.4	97.8	86.5	96.5	71.2
RoBERTa-ch-GCN-GAT	89.6	98.0	89.9	96.6	73.0

From the table it can be concluded that the model in this paper performs well on the dataset, performing well on the R8 and R52 datasets with 98.0% and 96.6% respectively. Considering that bert may have captured a lot of data information in the pre-training, there may be training data understood before this, coupled with a slight enhancement compared to other base models, there is still room for the model to be improved as well as fluctuations during training, which need to be considered for adversarial training. In addition the model can capture local contextual and structural information, as well as the global information of the graph neural, to obtain rich global information, at the same time, because the construction of the graph at the same time also led to the emergence of a very large sparse matrix, which also occupies a large amount of memory resources. The model in this paper is improved a little compared with the base model, compared with other models this model is in the optimal value in each data set, indicating that the model can judge the text category better.

Next the results of the fine-tuning of the model in this paper on each dataset.

The following is the Ohsumed dataset in BERT fine-tuned training results (acc, loss):

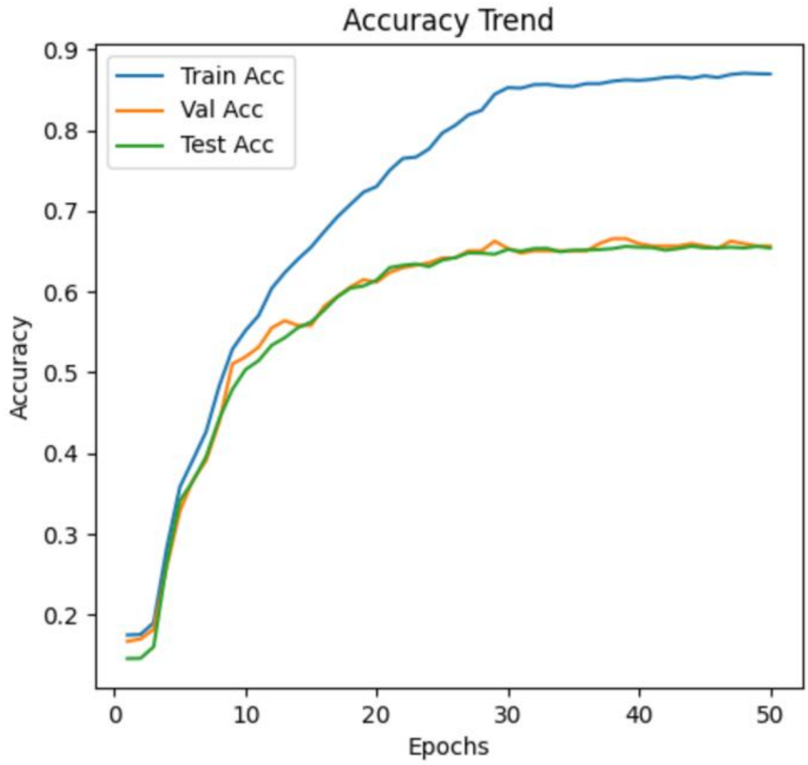

Figure3.2 Ohsumed dataset in BERT fine-tuned acc trend

The following is the Ohsumed dataset for the model training results in this paper.

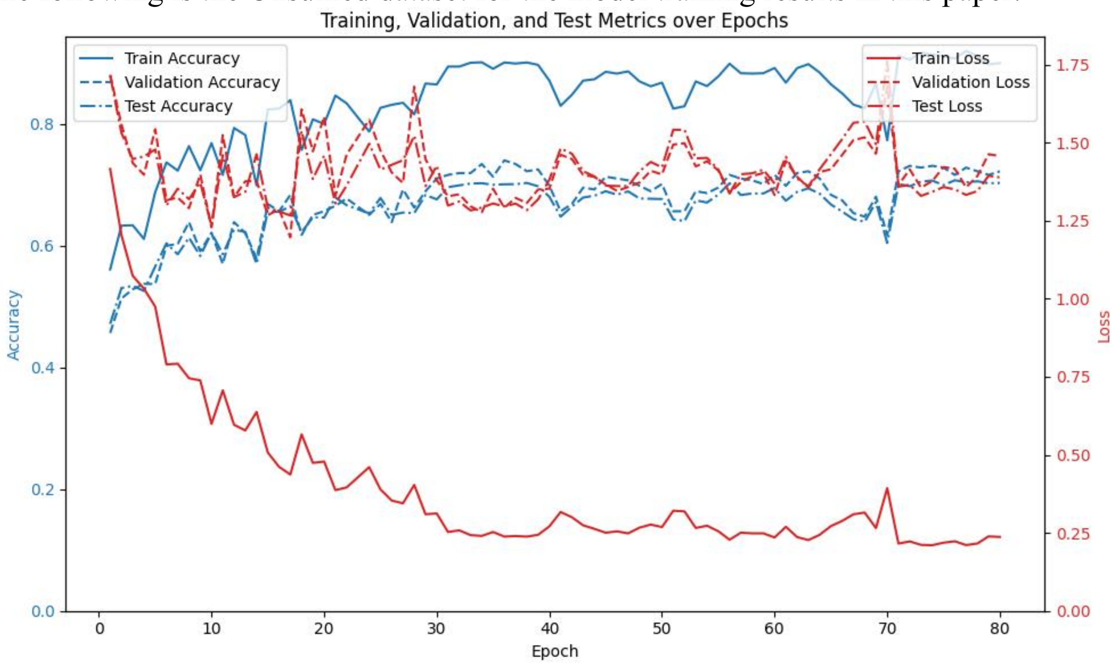

Figure 3.3 Ohsumed dataset in the current model of this paper results

The training results of our model, which integrates BERT with Graph Neural Networks (GNN), are highly promising, as demonstrated in the accompanying training effect chart. The model exhibits excellent performance with a high accuracy rate, highlighting its ability to effectively classify social comments by leveraging both the powerful language representations from BERT and the relational insights provided by GNN. Additionally, the chart shows a significant reduction in the loss function over time, indicating that the model is not only accurate but also efficiently learning to minimize errors during training. These characteristics—high precision and low loss—underscore the robustness of the model, making it well-suited for complex social comment classification tasks. The combination of BERT's contextual understanding and GNN's relational reasoning proves to be a highly effective approach, setting the foundation for superior performance in real-world applications.

4. Summary

In this study, we propose a novel social comment classification model that integrates BERT and Graph Neural Networks. The combination of BERT's powerful language understanding capabilities and GNN's ability to capture relationships between entities allows the model to better understand the context of social comments, considering both textual content and inter-comment relationships. Our experimental results show that the model achieves excellent performance in terms of classification accuracy, demonstrating its ability to effectively handle the complexities of social comment data. The results further indicate that the integration of BERT with GNN significantly improves the model's ability to identify nuanced sentiment and contextual information in social media interactions. This approach sets a new standard for leveraging deep learning techniques in social comment analysis, providing more accurate and context-aware predictions.

References

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

[2] Yao L, Mao C, Luo Y. Graph convolutional networks for text classification[C]//Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 7370-7377.

[3] Liu X, You X, Zhang X, et al. Tensor graph convolutional networks for text classification[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(05): 8409-8416.

[4] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J]. 2018.

[5] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners[J]. OpenAI blog, 2019, 1(8): 9.

[6] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. Advances in neural information processing systems, 2020, 33: 1877-1901.

[7] Achiam J, Adler S, Agarwal S, et al. Gpt-4 technical report[J]. ArXiv preprint ArXiv:2303.08774, 2023.

[8] Du Z, Qian Y, Liu X, et al. Glm: General language model pretraining with autoregressive blank infilling[J]. ArXiv preprint ArXiv:2103.10360, 2021.

[9] Zhao W X, Zhou K, Li J, et al. A survey of large language models[J]. ArXiv preprint ArXiv:2303.18223, 2023.

[10] Yang J, Jin H, Tang R, et al. Harnessing the power of llms in practice: A survey on chatgpt and beyond[J]. arXiv preprint arXiv:2304.13712, 2023.