

Cyber Violence Text Classification Model Based on Graph Convolutional Networks and Syntactic Parsing

Binting Qi, Hanming Zhai, Yunyang Bu, Zhuxuan Han, Fanliang Bu *

Department of Information Network Security, People's Public Security University of China, Beijing, China

* Corresponding author: Fanliang Bu (Email: bufanliang@sina.com)

Abstract: There are problems such as semantic sparsity and incomplete context in cyber violence texts in social media, and the current research on the refinement and classification of cyber violence texts is insufficient. This paper constructs a text classification model based on graph neural networks to improve the fine-grained classification effect of multi category cyber violence texts in social media. Introducing dependency relationships into BERT through syntactic analysis for deep semantic representation learning enhances the model's contextual understanding ability and better captures fine-grained features in sentences. Build a graph structure containing two different node types, words and documents, to propagate semantic information through GCN and strengthen the relationships between nodes. By combining large-scale pre-trained models with GCN, complementary advantages can be achieved to improve the overall performance of the model. The experimental results show that the method achieves a classification accuracy of 84.77% on publicly available cyber violence text datasets, which is 2.36% higher than the baseline model. It can effectively and accurately identify multiple types of cyber violence texts.

Keywords: Graph Convolutional Network; Cyber Violence; Text Classification; BERT; Syntactic Parsing.

1. Introduction

With the rapid development of the Internet and social media, communication on social media has become a norm. Users can freely express their opinions on the Internet. This freedom also makes cyber violence more frequent. Nowadays, cyber violence has attracted high attention worldwide and has become a major hot issue that urgently needs to be addressed. Cyber violence information refers to illegal and harmful content that is concentrated on individuals through the internet in the form of text, images, audio, video, etc., containing insults, slander, rumors, incitement to hatred, coercion, invasion of privacy, as well as accusations, ridicule, belittling and discrimination that affect physical and mental health. Despite the continuous efforts of major platforms to develop measures to combat cyber violence, its scale and complexity still make it difficult to completely solve the problem.

This study mainly explores the classification of cyber violence texts in social media. Text classification is a key task in natural language processing, aimed at assigning text content to defined categories. The cyber violence text on social platforms is usually dominated by short texts, which have sparse semantics, vague information, lack context, and often contain insulting, defamatory, threatening words as well as satirical, mocking tone features. Therefore, existing classification techniques cannot fully adapt to such texts, making it difficult to fully explore the characteristics of violent information.

This article designs a cyber violence text classification model based on graph convolutional networks and syntactic analysis, which subdivides various categories of cyber violence texts and classifies them according to the age, race, gender, religion, or other attributes of the victims. This model introduces syntactic relationship information into the BERT model by performing syntactic analysis on the text, helping the model better understand the grammar and semantic relationships within sentences and better represent the

features of cyberbullying texts. Graph convolutional network is a deep learning method suitable for processing graph structured data, which can deeply explore the semantic and internal relationships between texts by mapping them as graph structures. The joint training of BERT and GCN has improved the effectiveness of cyber violence text classification, demonstrating that through large-scale pre training, GCN can better learn the complex relationships between nodes and graph structures, enhance its representation of node features, and significantly optimize the model's generalization ability.

The main contributions and innovations of this article are summarized as follows:

(1) This article conducts fine-grained classification of cyber violence texts into multiple categories. While determining whether they belong to cyber violence texts, it is no longer limited to simple classification based on gender discrimination or racial discrimination. Based on different characteristics of cyber violence, a comprehensive classification system is constructed.

(2) By performing syntactic analysis on the text and introducing syntactic relationship information into the BERT model, the model can more accurately capture the dependencies and semantic connections within the sentence. Joint training of BERT model and GCN can better learn features in cyber violence texts, which can help improve text classification performance.

(3) The model proposed in this article was tested on a publicly available dataset of cyber violence, and its classification performance was better than the baseline model, improving the classification performance of cyber violence text in social media.

2. Related Works

2.1. Text Classification

The traditional methods for text classification include Support Vector Machine (SVM)[1], Naive Bayes (NB)[2], K-

Nearest Neighbor Algorithm (KNN)[3], Decision Tree Algorithm[4], etc. These traditional classification algorithms have high requirements for feature engineering, but in social media, the proportion of short texts is large, and the characteristics of short texts themselves are insufficient, which limits the performance of these methods in processing such data.

With the continuous advancement of neural network technology, more and more deep learning models have been introduced into text classification tasks, including convolutional neural networks (CNN)[5], recurrent neural networks (RNN)[6], and attention based models[7]. The emergence of BERT[8] has brought revolutionary changes to natural language processing technology, as it can generate word vectors containing contextual information, significantly improving the semantic understanding ability of models. Many researchers have studied BERT based text classification models, which have achieved better performance than the above models in text classification and multiple NLP tasks. By conducting unsupervised training on large-scale unlabeled corpora, pre trained models can deeply explore potential semantic information in language and demonstrate superior performance in specific tasks through fine-tuning.

With its powerful ability to express non Euclidean geometric relationships, Graph Neural Networks (GNNs)[9] have been applied in multiple practical scenarios. Graph Convolutional Network (GCN) [10] is a variant of GNN that can capture information of high-order neighboring nodes. GCN updates node features by multiplying the adjacency matrix and node feature matrix, and is used to perform various tasks such as node classification and graph classification. Yao[11] et al. proposed a Text Graph Convolutional Network (TextGCN) model, which constructs the entire dataset into a heterogeneous text graph containing word nodes and document nodes, and effectively extracts global word co-occurrence features. Graph Attention Network (GAT)[12] utilizes masked self-attention layers to manipulate graph structured data, and evaluates the importance of neighboring nodes to the target node through attention mechanisms when aggregating neighborhood information of graph nodes. The Heterogeneous Graph Attention Network (HGAT) with dual layer attention mechanism[13] learns the importance of different adjacent nodes and node types in a node. This model

solves the sparsity problem in semi supervised short text classification by propagating information in the graph structure and capturing the relationships between nodes. Lin[14] combined large-scale pre-trained models with GCN and used graphs to model the relationships between different samples in the entire corpus, achieving advantageous results on public datasets.

2.2. Classification of Cyber Violence Text

Due to the enormous threat posed by online violence to society, online violence detection has gradually become a research hotspot. Dinakar[15] et al. used Naive Bayes and SVM to construct a separate topic sensitive classifier for detecting cyberbullying texts. Chavan and Shylaja[16] proposed an algorithm to calculate a score representing the probability of comments offending other users, utilizing a series of features such as skip grams to detect aggressive comments targeting peers. Squicciarini[17] et al. applied decision tree classifiers to specific content features to detect cyberbullying behavior in social networks such as MySpace and Spring.me. At present, deep learning models have also become the main method for detecting online violent texts. Sweta Agrawal[18] uses four models, CNN, LSTM, BLSTM, and BLSTM, to analyze online violent texts on multiple social media platforms. Daraghmi[19] et al. compared and analyzed various machine learning and deep learning algorithms and proposed a CNN BiLSTM GRU hybrid model for detecting Arabic cyberbullying. The model has stacked word embedding function and can adapt to different data changes.

The research on cyber violence in China is also increasing. Huang Rui[20] proposed a dictionary and rule-based method for detecting cyberbullying language and established a corresponding detection system. Qiang Lan[21] constructed a Chinese cyber violent text language database and implemented the detection of cyber violent text using a character level neural network model. Wang Runyu[22] provided a comprehensive summary and analysis of research results on cyberbullying, and put forward thoughts and suggestions on intelligent intervention for cyberbullying.

3. Cyber Violence Text Classification Model

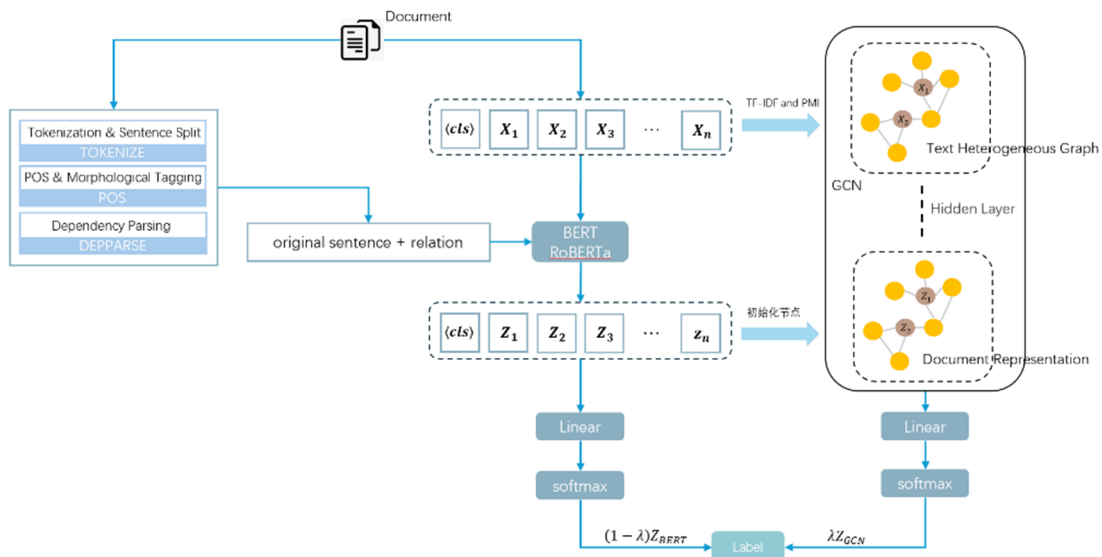


Fig 1. R-BG model structure

In the cyber violence text classification model based on graph convolutional networks and syntactic relationship analysis proposed in this article (Relation BertGCN, hereinafter referred to as R-BG model), the text is first subjected to syntactic analysis, and the extracted dependency relationship features are input together with the original text into BERT related models (BERT, RoBERTa) for training. This approach adds additional contextual information during model training, enabling the model to more accurately capture the dependency relationships and semantic connections within sentences. Then use BERT related models to generate initial representations for each document node and input them into GCN for processing. The document representation is iteratively updated in the graph structure through GCN, and the final generated document node features are fed into a softmax classifier for prediction. The structure of the R-BG model is shown in Figure 1.

3.1. Syntactic Parsing

Firstly, label the original text and group the labels into sentences. Then, based on a bidirectional long short-term memory network, assign a part of speech (POS) to each word in the sentence and analyze its universal morphological features (UFeats). In order to ensure consistency between

Universal POS (UPOS), Tree Library Specific POS (XPOS), and UFeats, the Biafine scoring mechanism proposed by Dozat and Manning was used to convert XPOS and UFeats prediction conditions into UPOS prediction conditions. Adjust the predictions of XPOS and UFeats based on UPOS predictions. Finally, neural dependency syntactic analysis was implemented based on BiLSTM's deep double affine approach, parsing the syntactic structure of each sentence. Each word in the sentence was assigned a syntactic header, which is another word or an artificial root symbol in the sentence. The model was further expanded with the linearization order of the two words and their typical linear distance to improve the accuracy of syntactic analysis.

The deep double affine neural dependency syntactic analysis structure based on BiLSTM is shown in Figure 2. The input is a vector representation obtained by concatenating words and part of speech vectors. The feature r_i is extracted using BiLSTM, and $h(\text{arc-dep})$ and $h(\text{arc-head})$ are obtained through two different MLPs. The hstack of all words yields $H(\text{arc-dep})$ and $H(\text{arc-head})$, and an additional unit vector is concatenated for $H(\text{arc-head})$. Using the intermediate matrix $U(\text{arc})$ for affine transformation, dot product each word as dep with each word as head to obtain the score matrix $S(\text{arc})$ where arc holds.

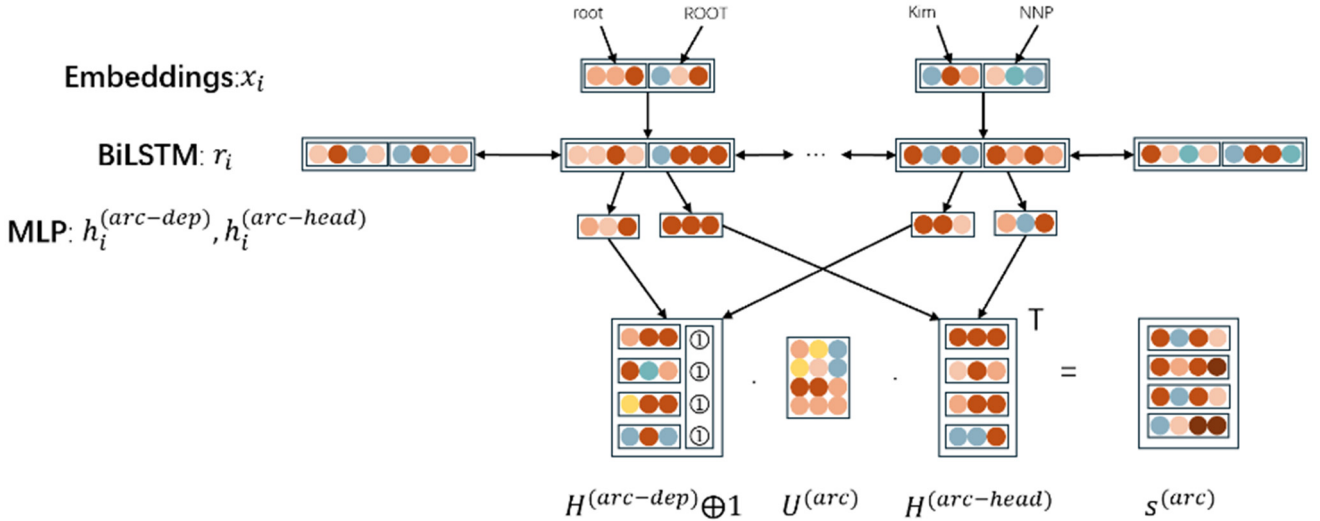


Fig 2. Syntactic parsing structure

3.2. Graph Convolutional Network Model for Cyber Violence Text

A heterogeneous graph $G = (V, E)$ was constructed based on TextGCN [12], which includes word nodes and document nodes. The graph can visually represent the co-occurrence relationship of global words and adapt to graph convolutional networks. It can explicitly model global word co-occurrence and adapt well to graph convolutional networks. V represents the set of nodes in the graph, and E represents the set of edges in the graph. By analyzing the occurrence of words in a document, edges between document nodes are established, and their weights are determined by the word frequency inverse document frequency (TF-IDF) of the words. Word frequency represents the number of times a word appears in a document, while inverse document frequency is the reciprocal of the number of documents containing that word. The larger the TF-IDF value of a word, the higher its

importance in the document. Based on the co-occurrence of words in the entire corpus, edges are constructed between word nodes. The weights between two-word nodes are calculated using Point Mutual Information (PMI), which measures word associations. Based on the co-occurrence pattern of words in the corpus, edges are generated between word nodes, and their weights reflect the strength of word associations by calculating PMI. The calculation formula is as follows:

$$PMI(i, j) = \log \frac{p(i, j)}{p(i)p(j)} \quad (1)$$

When the PMI value is positive, it indicates that the word has high semantic relevance in the corpus, while a negative PMI value indicates low or non-existent semantic relevance. Therefore, only add edges between two-word nodes when the PMI value is greater than 0. Define the weight of the edge between nodes i and j using the following formula:

$$A_{ij} = \begin{cases} PMI(i, j), i, j \text{ is words} \\ TF - IDF, i \text{ is document}, j \text{ is word} \\ 1, i = j \text{ is words} \\ 0, \text{ others} \end{cases} \quad (2)$$

TextGCN uses the identity matrix $X = I_{n_{doc} + n_{word}}$ as the node feature initialization, where n_{doc} and n_{word} represent the number of document nodes and word nodes, respectively. In the R-BG model, document embeddings are generated using models such as BERT or RoBERTa, and B uses them as input representations for document nodes, denoted as $X_{doc} \in \mathbb{R}^{n_{doc} \times d}$, with an embedding dimension of d . The specific form of the initial node feature matrix is as follows:

$$X = \begin{pmatrix} X_{doc} \\ 0 \end{pmatrix}_{(n_{doc} + n_{word}) \times d} \quad (3)$$

The GCN model receives the initial feature matrix X as input and achieves information propagation through multiple iterations during training and testing. For the i -th GCN layer (denoted as $L^{(i)}$), the calculation formula for its output feature matrix is:

$$L^{(i)} = \rho(\tilde{A}L^{(i-1)}W^{(i)}) \quad (4)$$

In the formula, ρ represents the activation function, $\tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ represents the normalized adjacency matrix, and $W^{(i)} \in \mathbb{R}^{d_{i-1} \times d_i}$ is the weight matrix of that layer, $L^{(0)} = X$ is the input feature matrix. Generate the final feature representation of the document through GCN and input it into the softmax layer for category prediction:

$$Z_{GCN} = \text{softmax}(g(X, A)) \quad (5)$$

g represents the GCN model. During the training process, the cross entropy loss function on the labeled document nodes is used to jointly optimize the parameters of BERT and GCN. The calculation formula is as follows:

$$\mathcal{L} = - \sum_{d \in y_D} \sum_{f=1}^F Y_{df} \ln Z_{df} \quad (6)$$

Among them, y_D represents a set of labeled document indexes, F is the dimension of output features, and Y is the label index matrix used to represent the structural format of multi label data.

Through research, it has been found that interpolation prediction can lead to faster convergence and better performance of the model. Therefore, embedding the document directly into a fully connected layer containing softmax is used to form an auxiliary classifier:

$$Z_{BERT} = \text{softmax}(WX) \quad (7)$$

The ultimate training objective is the linear interpolation of predictions from GCN and BERT, calculated using the following formula:

$$Z = \lambda Z_{GCN} + (1 - \lambda) Z_{BERT} \quad (8)$$

Among them, λ controls the weight of two objectives. $\lambda=0$ indicates the use of only BERT related models that introduce syntactic relationships, while $\lambda=1$ indicates the use of only GCN models. When $\lambda \in (0, 1)$, the predictions of both models can be balanced, and the overall model can be better

optimized. Directly manipulate the input of GCN through Z_{BERT} to ensure that the optimization direction of the input is consistent with the goal. This method effectively alleviates the common problems of gradient vanishing and excessive smoothing in multi-layer GCN models, thereby significantly enhancing the performance of the model.

4. Experiment

4.1. Dataset

The dataset used in the experiment is the cyberbullying tweets dataset, which was constructed by Wang [23] et al. in their study of a graph convolutional network method for fine-grained cyberbullying detection. Wang et al. identified and collected 16 open-source datasets on cybullying through literature review, and ultimately selected six Twitter datasets for further analysis. After using the Twitter API to retrieve tweet content, they manually classified tweets and expanded the sample in a semi supervised manner through dynamic query expansion methods. Finally, they randomly selected not-cyberbullying texts and cyberbullying tweets targeting age, race, gender, religion, and other types, forming a balanced dataset of approximately 48000 texts. The text on age discrimination and racism in the dataset may differ from what is expressed in existing cyberbullying detection. The data labels are divided into six types: not_cyberbullying, age, ethnicity, gender, religion, and other_cyberbullying.

Before validating the model proposed in this article on this dataset, a simple preprocessing was performed to remove links, mentions (@username), forwarding flags "RT", and punctuation marks from tweets. The processed effective dataset contains a total of 47242 texts, with a maximum text length of 322, a minimum text length of 3, an average text length of 25, and 6 types of labels. The specific number of texts is shown in Table 1.

Table 1. The number of texts in each category of the cyberbullying tweets dataset

Text Category	Effective Text Quantity
not_cyberbullying	7802
age	7990
ethnicity	7954
gender	7918
religion	7997
other_cyberbullying	7581

4.2. Experimental Setup

For BERT and RoBERTa, use the output features of [CLS] tokens as document embeddings, and then use a feedforward layer to output the final prediction. Use BERT or RoBERTa and a two-layer GCN to achieve joint training. The learning rate of GCN module is initialized to 1e-3, and for fine-tuned BERT or RoBERTa modules, the learning rate is initialized to 1e-5.

4.3. Comparative Experimental Model

This article selects FastText, TextCNN, TextRNN, BERT, RoBERTa and other benchmark models for comparative experiments. Specifically, as follows:

FastText: A simple and efficient text classification method

that considers the average value of word/n-gram embeddings as document embeddings, and then inputs the document embeddings into a linear classifier.

TextCNN: A text classification model based on convolutional neural networks. It extracts text features through convolution operations and then uses max pooling operations to extract the most important features.

TextRNN: Use RNN to calculate the sequential representation of text. A model for text classification tasks using RNN class recurrent neural networks.

BERT: A powerful pre-trained model that can utilize the knowledge obtained from pre training for classification.

RoBERTa: It is a further improvement of the BERT model, with stronger understanding and representation capabilities.

4.4. Evaluation Indicators

Text classification tasks are usually divided into binary classification and multi classification tasks, each with different evaluation criteria that help understand the accuracy, stability, and applicability of the model in specific tasks. Common evaluation metrics include accuracy, precision, recall, and F1 score. In the binary classification task, the following are the calculation formulas for each indicator:

$$Accuracy = \frac{TP}{TP + FP + FN + FP} \quad (9)$$

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) represent the true examples, true negative examples, false positive examples, and false negative examples in the confusion matrix.

The online violence text studied in this article belongs to a multi classification task, therefore accuracy, weighted average accuracy, weighted average recall, and weighted average F1 score are used as indicators for evaluating the model. Firstly, calculate the indicators for each category, then calculate the weights of each indicator based on the sample size, and finally calculate the weighted average. The specific

calculation formula is as follows:

$$Accuracy = \frac{\sum_{i=1}^N TP_i}{sum} \quad (13)$$

$$Precision_{weight} = \sum_{i=1}^N W_i P_i \quad (14)$$

$$Recall_{weight} = \sum_{i=1}^N W_i R_i \quad (15)$$

$$F_{1weight} = \sum_{i=1}^N W_i F_{1i} \quad (16)$$

Among them, N represents the number of categories, and in this study, N=6. TP_i refers to the number of correctly predicted samples in class i, and sum is the total number of samples. W_i, P_i, R_i, and F_{1i} respectively represent the weight, accuracy, recall, and F1 value of the i category.

4.5. Result

From Table 2, it can be seen that among all baseline models, BERT and RoBERTa outperform other baseline models due to their large pre training foundation. RoBERTa uses larger and richer text data and longer text segments during the training process, which enables it to more accurately understand and capture contextual relationships in text. At the same time, it adopts stricter masking strategies and deeper network structures, resulting in higher classification performance than BERT, achieving the best results in the baseline model with a classification accuracy of 82.41% and a weighted average F1 value of 82.30%. The R-RoBERTaGCN model proposed in this article outperforms the baseline model in classifying online violent text datasets, with an accuracy improvement of 2.36% and an F1 score improvement of 1.83% compared to the best performing baseline model. The specific results are shown in Table 2.

The results in Table 2 demonstrate the effectiveness of the proposed method in fine-grained classification of online violent texts, mainly due to the following reasons: 1. The graph convolution model can capture document word relationships and global word relationships, enabling effective propagation of document label information through word nodes in the graph. 2. Introducing syntactic relationships into BERT related models enables the model to more accurately capture the dependencies and semantic connections within sentences, while significantly improving classification performance by integrating the advantages of pre trained models and graph models.

Table 2. Comparative experimental results

Model	Accuracy (%)	Weighted Average Precision (%)	Weighted Average Recall (%)	Weighted F1 (%)
FastText	76.52	76.36	76.31	76.44
TextCNN	78.12	77.76	77.82	78.06
TextRNN	80.52	80.24	80.28	80.04
BERT	81.51	82.07	81.01	81.96
RoBERTa	82.41	82.46	82.33	82.30
R-BERTGCN	84.16	84.23	83.92	84.02
R-RoBERTaGCN	84.77	84.65	84.56	84.13

4.6. Ablation Experiment

In order to investigate the impact of each module on the model, ablation experiments were conducted. The results of the ablation experiment are shown in Table 3.

From the results in Table 3, it can be seen that when the text is not subjected to syntactic analysis, directly inputting the text into the BERT related model for text representation and combining it with GCN training results in a decrease in the

classification performance of online violent text compared to the complete model. This is because compared to directly inputting the text into the BERT related model, sending the syntactic analysis results together with the original sentence into the BERT related model can explicitly express the sentence structure and dependency relationships, making it easier for the model to understand the logical relationships of the sentence. By explicitly annotating dependency relationships, the model can better capture fine-grained features in sentences, thereby improving classification performance. In addition, the annotation of syntactic dependency relationships can help the model eliminate ambiguity in sentences, and the model can more clearly know which words are interdependent. In addition, it can be seen from the table that when the RoBERT module or GCN module is removed for training, the text classification performance decreases. Due to the limited number of nodes and edges for shorter texts and the limited ability of graphic structures, the impact of removing the GCN module is relatively small. However, RoBERT, with its large pre training features, obtains better text embeddings and helps the model achieve better classification performance.

Table 3. Results of ablation experiment

Model	Accuracy (%)	Weighted F1(%)
Remove RoBERT	79.64	79.81
Remove GCN	82.95	82.93
Remove Syntactic Parsing	82.51	82.34
our model	84.77	84.13

5. Summary

This article proposes a graph convolutional network classification model R-BG for cyber violence text. By introducing dependency relationships, BERT related models can better understand the grammar and semantic structure in sentences, thereby more accurately representing features in cyber violence text. The proposed model combines the advantages of BERT related models and Graph Convolutional Networks (GCN) for joint training, significantly improving the classification performance of cyber violence text. At the same time, the model can effectively identify and distinguish different categories of cyber violence text, providing scientific basis for formulating response strategies and helping to maintain a healthy online environment.

References

- [1] Drucker H, Wu D, Vapnik V N. Support vector machines for spam categorization[J]. IEEE Transactions on Neural networks, 1999, 10(5): 1048-1054.
- [2] McCallum A, Nigam K. A comparison of event models for naive bayes text classification[C]//AAAI-98 workshop on learning for text categorization. 1998, 752(1): 41-48.
- [3] Cover T, Hart P. Nearest neighbor pattern classification[J]. IEEE transactions on information theory, 1967, 13(1): 21-27.
- [4] Kotsiantis S B. Decision trees: a recent overview[J]. Artificial Intelligence Review, 2013, 39: 261-283.
- [5] Kim Y. Convolutional Neural Networks for Sentence Classification[C]//Proceedings of Conference on Empirical Methods in Natural Language Processing, 2014.
- [6] Tang D, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification[C]//Proceedings of the 2015 conference on empirical methods in natural language processing. 2015: 1422-1432.
- [7] Vaswani A. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017.
- [8] Devlin J. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arxiv preprint arxiv: 1810.04805, 2018.
- [9] Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks[J]. IEEE transactions on neural networks and learning systems, 2020, 32(1): 4-24.
- [10] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. arxiv preprint arxiv: 1609.02907, 2016.
- [11] 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.
- [12] Velickovic P, Cucurull G, Casanova A, et al. Graph attention networks [J]. stat, 2017, 1050(20): 10-48550.
- [13] Linmei H, Yang T, Shi C, et al. Heterogeneous graph attention networks for semi-supervised short text classification[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: 4821-4830.
- [14] Lin Y, Meng Y, Sun X, et al. Bertgen: Transductive text classification by combining gen and bert[J]. arxiv preprint arxiv: 2105.05727, 2021.
- [15] Dinakar K, Reichart R, Lieberman H. Modeling the detection of textual cyberbullying[C] //Proceedings of the International AAAI Conference on Web and Social Media. 2011, 5(3): 11-17.
- [16] Chavan V S, Shylaja S S. Machine learning approach for detection of cyber-aggressive comments by peers on social media network[C]//2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE, 2015: 2354-2358.
- [17] Squicciarini A, Rajtmajer S, Liu Y, et al. Identification and characterization of cyberbullying dynamics in an online social network [C]//Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015. 2015: 280-285.
- [18] Agrawal S, Awekar A. Deep learning for detecting cyberbullying across multiple social media platforms[C] // European conference on information retrieval. Cham: Springer International Publishing, 2018: 141-153.
- [19] Daraghmi E Y, Qadan S, Daraghmi Y, et al. From Text to Insight: An Integrated CNN-BiLSTM-GRU Model for Arabic Cyberbullying Detection[J]. IEEE Access, 2024.
- [20] Huang R. Implementation of a Network Violent Language Detection System[D]. WuHan: School of information and communication engineering, Huazhong University of Science and Technology, 2016.
- [21] Qiang L. Research on Cyberbullying Language Based on Social Networks[D]. TaiYuan: North University of China, 2020
- [22] Wang R, Zhang P, Gu N, et al. Intelligent Intervention Methods for Cyberbullying in Online Communities: Landscapes, Challenges and Prospects[J]. Journal of Chinese Computer Systems, 2024, 45(10): 2312-2326.
- [23] Wang J, Fu K, Lu C T. Sosnet: A graph convolutional network approach to fine-grained cyberbullying detection[C] //2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020: 1699-1708.