Text Classification Method Based on Graph Neural Networks

Ruofei Gao*

Artificial Intelligence Department, Beijing Normal University, 100000, Beijing, China

Abstract. The goal of text classification is to assign labels to text units accurately, which is a basic task in natural language processing. This technology has shown great value in many practical application scenarios, covering spam detection, emotional tendency analysis, user intent recognition, and many other aspects. In recent years, because of the excellent performance of graph convolutional neural networks (GCNs) in processing non-European spatial data, it has become a research hotspot in text classification and has been widely adopted. This paper first introduces the background knowledge and working principle of graph convolutional neural networks. Then the text classification method based on GCNs is described. The last thing discussed is the limitations of graph convolutional neural networks, as well as the difficulties and future development directions in this field. This paper serves as a guide for researchers and practitioners in relevant fields to fully comprehend the latest advancements in text classification methods using graph convolutional neural networks, and by revealing its limitations and challenges, it also provides guidance for future research and development work, intending to promote technical progress and application expansion in this field.

1 Introduction

Natural language processing involves text classification as a basic and important branch. The main objective is to use text processing technology to analyze a given text content and accurately categorize it under one or more preset category labels. Text classification tasks are widely used, including sentiment analysis, spam recognition, subject recognition, intention recognition, automatic label generation, etc. Serving as the basis of many text-processing tasks, it is of great significance to improve the efficiency of text classification. This approach allows us to conduct data-intensive analysis, optimize the information retrieval process, and sharpen the data mining capabilities. Exploring and developing more efficient and accurate models is vital as text classification efficiency largely depends on the choice of classification models. This not only helps to speed up data processing but also fundamentally enhances the efficiency and accuracy of information utilization, laying the foundation for the wide application of artificial intelligence technology in text processing.

Text classification usually covers the key stages such as text preprocessing, text feature extraction, and model construction. Among them, text feature extraction is both challenging

* Corresponding author: 202411081062@mail.bnu.edu.cn

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

and crucial. To compute, text information must be transformed into a vectorized form while preserving semantic information. There are two main categories of text classification methods: text classification which uses machine learning and text classification which uses deep learning. Deep learning has caused a storm in Natural Language Processing (NLP) in recent years, and its rapid development has made it one of the main text classification methods. Compared with machine learning methods, deep learning methods have the advantage of their built-in network structure, enabling them to learn automatically from data and extract feature representations， greatly reducing the burden of manual complex feature engineering. Not only does this feature improve text classification efficiency, it also increases the model's generalization ability. Text classification has already been widely utilized with the use of certain typical deep learning models, which include Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). However, they can only handle Euclidean data. For non-Euclidean data such as text, the traditional method also first converts the text data into a two-dimensional vector matrix composed of language units such as words, phrases, or sentences, and then trains it with a network model. However, non-gridded data transformation frequently restricts the expressive power of neural networks, particularly text with complex grammatical structures. Given these conditions, the application of Graph Convolutional Neural Networks (GCNs) has attracted wide attention in text classification. The graph convolutional neural network interprets the text as a graph, expressing the rich relationship between the elements in the text in an audio-visual way. To get the global graph information, the graph network can use the connection relationship between the nodes. This feature could help to better express semantic relationships in text [1].

Based on the relevant research and experimental data on GCNs and text classification tasks, this paper introduces the basic principles of GCNs, summarizes the development of various GCN models in text classification, predicts its future development trend, and finally summarizes the existing limitations of GCNs, hoping to inspire future research.

2 Graph convolutional neural networks

GCN is a class of neural network models created to process data that has irregular graph structures. Nodes and edges make up graph structure data, with nodes representing entities and edges representing relationships between entities. It is a kind of complex non-Euclidean spatial data. Different from traditional neural networks, the advantage of Graph Neural Network (GNN) is its ability to use graph structure to aggregate the features of each node with the features of neighboring nodes, forming new node representations to obtain more comprehensive information. A lot of graph neural network models have been proposed over the past few years, and GCN is one of the important directions, that has been widely used in various fields.

2.1 Traditional types of GCNs

CNNs are the inspiration for GCNs, and their definition of convolutional operator and pooling operator steps are similar. Using the Laplacian matrix, GCNs transfer information and extract features by convolution operation.

A text classification task was first assigned a graph neural network by Textgcn [2]. Taking documents and words as nodes, Textgcn used text data to construct a word-text undirected weighted heterogeneous graph. The model constructs edges between nodes according to the occurrence of words in the document and the whole corpus, taking the word frequency of words in the document as the weight. Point mutual information (PMI) is employed to compute the weight between two-word nodes. The text data is converted into graph structure data by constructing a co-presence graph, which is input into the GCN neural network, learns a two-layer GCN model, and finally input into a softmax classifier to get the final classification result. Modal can better express the rich relationship between text elements using graphs to represent text information.

Wang et al. came up with the TensorGCN model, which is improved based on Textgcn and introduces a graph tensor structure to represent the relationship between nodes, and then perform operations on the tensor. Such improvements can effectively handle different types of relationships, as they can better capture complex semantics and correlations in textual data. At the same time, the tensor structure also improves the modeling ability of the model for complex text data, making the network structure of TensorGCN more flexible and able to adapt to text classification tasks of different complexity and requirements [3].

Although it is successful to use graph structure to represent text information and then use GCN to process graph structure, the computational complexity will gradually increase if the graph structure gets larger, which may lead to performance degradation and increase the training time. To solve the problem of the computational complexity of GCN, Wu et al. proposed simplified graph convolution (SGC). The core idea of SGC is to remove the nonlinear feature extraction part of GCN, remove the activation process of hidden layers, make the intermediate process more straightforward using linear transformation, and significantly decrease the fitting parameters, thus reducing the model redundancy and complex calculation, and improve the efficiency of model processing large-scale graphs [4].

The Text-Level GCN model, proposed by Huang et al., incorporates word interaction within the text. Each text is analyzed to construct a directed graph, and the node feature matrix and edge weight matrix are designed globally. The model is capable of effectively addressing the problem of high memory consumption and enhancing the generalization capability of new nodes [5]. In addition, Gao et al. proposed a hybrid convolution operation combining GCN graph convolution and CNN one-dimensional convolution, rapidly enhancing the receptive field and extracting word order data from text [6].

In response to several graph models of transduction sexual problems, literature [7] proposed a modal called induct graph convolutional network (InductGCN) to make GCNs more widely applied to text classification tasks. As real-life data sets are dynamic and multivariate, the conventional GCN model requires retraining when introducing new data points, while InductGCN can infer relationships between nodes and generalize to new data points that follow similar structural features by taking into account the topology and connection patterns of the graph. InductGCN also reduces space complexity and parameters to capture global information efficiently. As a result, the model is more scalable, flexible, and adaptive than other common GCN variants. In production environments, this advantage also makes it possible for faster deployment and more efficient applications.

To make it more sufficient for the text semantic extraction, the literature [8] proposed a model based on multiple semantic dual channel attention fusion, dual-channel attention network model (DCAT). The DCAT approach initially involves the extraction of logical semantics from text using transduction learning and the analysis of graph structures. This method effectively decodes the underlying meaning embedded within the text by leveraging the strengths of transduction learning, which is adept at capturing patterns and relationships. Additionally, by examining the graph structure, DCAT can discern the intricate connections and hierarchies present in the data, providing a comprehensive understanding of the text's logical framework.

2.2 Integrating GCNs with Natural Language Processing

In the past few years, Transformer has been commonly utilized, and a growing number of researchers have combined large-scale pre-trained language models and graph models to carry out text classification tasks. Through the pre-training stage, the language model is trained to learn rich language representations and can accurately capture semantic relations and context information between words. The graph model can effectively capture the relations and structural information between text data. The two models' respective advantages can be utilized by this combination to further optimize the performance of natural language processing tasks. BertGCN model was proposed, combining bidirectional encoder representations from transformers (Bert) with GCN [9]. The model constructs a heterogeneous graph from a data set and uses Bert to represent the document as a node. Training data and unlabeled test data representations are jointly learned by pre-training of copious amounts of raw data and propagating label impact through graph convolution.

Reference [10] proposes the BertGACN model based on BertGCN, which is a variant of BertGCN. Combining GCN with Graph Attention Network (GAT), GACN is a two-tower structure model. This model combines the advantages of the two. While the GCN module is capable of effectively learning graph structure information, the GAT can use the attention mechanism to learn the links between node properties. Based on the idea of stacking ensemble learning, this paper presents an algorithm that uses the BertGACN-stacking multi-base model framework. Bert-Boosting and multiple graph neural networks are integrated training to form five base classifiers for text classification, improving text classification ability and scene adaptability. Documents and words are represented as nodes and nodes respectively in a heterogeneous graph created. The graph convolution network module, graph attention network module, and bidirectional encoder representation of the Bert module are trained together to effectively learn the graph's structure information and the associations between its nodes, improving the text classification capacity.

The paper [11] introduces an innovative text classification model that integrates BERT with a hypergraph convolutional network (referred to as IBHC). This model is designed to harness the strengths of both BERT and hypergraph convolution to achieve a more nuanced understanding of text data. The BERT model is employed to extract local semantic information from the text, capturing the intrinsic meaning of individual words and phrases. Concurrently, the construction of text hypergraphs allows the model to identify broader associations between different textual elements, enhancing the model's ability to understand the contextual relationships within the text. Furthermore, the hypergraph convolutional network is utilized to extract global structural features from the text. This component of the model provides a macroscopic view of the text, enabling it to recognize patterns and structures that extend beyond individual words or phrases. By integrating these global features with the local semantic insights provided by BERT, the model gains a more holistic perspective on the text. To refine this comprehensive representation, the model employs an attention mechanism that facilitates interaction between the local and global features. This interaction allows the model to prioritize relevant information and to construct a more accurate and detailed representation of the text. As a result, the IBHC model offers a robust framework for text classification that leverages both the depth of semantic understanding provided by BERT and the breadth of structural analysis afforded by the hypergraph convolutional network.

Furthermore, it has been demonstrated through various studies that the performance of the modal can be improved by incorporating a variety of embedding techniques. Literature [12] presents a model that combines graph and BERT embeddings to reframe text classification as a node classification task. The model initially constructs a heterogeneous graph that encompasses the entire corpus, incorporating nodes representing words and documents, as well as edges that connect these nodes. Secondly, the two-level attention mechanism is integrated into GCN, the constructed heterogeneous graph is introduced into the network model, and the graph with global information is embedded by training. Then the text is trained by pre-training model BERT in the form of sequence, and the word embedding with local information can be obtained, which can effectively tackle the challenge posed by words with multiple meanings. Then the text embedding representation and graph embedding representation of BERT are fused to obtain the word embedding representation containing both global and local information. Then, the correlation degree between each word and the text potential vector is calculated by the text potential vector, and the attention weight is given to the word. Finally, the text embedding representation is obtained, to realize the text classification. Experimental results on four widely used public data sets show that the proposed model achieves better classification results than the baseline model. Compared with the model without the introduction of an attention mechanism, this model can take into account the importance of different types of nodes to a node and the importance of the same type of neighbor nodes to the current node. At the same time, the BERT pre-training model is used to obtain embedments containing contextual information and solve the problem of polysemy of the word.

3 Limitations and future prospects

3.1 Network depth

Traditional deep learning achieves good performance by constructing multi-layer or even deep network structures, and the deepening of the network layer significantly enhances the representation learning ability of the model. One significant limitation of GNN, however, is that the network hierarchy is usually shallow, with most not exceeding three layers. A network with excessive depth is prone to over-smoothness, causing node representations to converge and consequently degrading the model's performance.

Therefore, future research can focus on breaking through the hierarchical limitations of graph neural networks and exploring ways to design GNN models with deeper layers. This includes the study of new network architectures and training strategies to address the problem of over-smoothing, and better fusion of graph structure information and node features to enhance the representation and generalization performance.

3.2 Expandability

At present, the research of GCN is mainly limited to the processing of small-scale graphs, and the scalability of the model has become a serious constraint for large-scale graphs that are frequently processed in industrial applications. For instance, platforms like YouTube and TikTok, which are social networking sites, can possess a vast network structure with a large number of nodes and edges. In this large-scale data environment, text classification not only takes a long time but also costs a lot of computation. Therefore, future research can focus on extending existing GNN models to the capability to handle large-scale graphs at an industrial level. This includes exploring efficient data processing techniques and parallel computing methods to reduce computational complexity on large-scale graphs; At the same time, new network architectures and algorithms can be studied to optimize the scalability and stability of the models. In addition, how to achieve fast and accurate information dissemination and update on large-scale graphs, and how to reduce computing resource consumption while maintaining model performance, can also be an important direction of research.

3.3 Interpretability

Deep learning has long suffered from a lack of interpretability as it is so similar to the black box, and GNN, as an important branch of deep learning, also faces this challenge. It is important to apply the GNN model to real-world scenarios, most of which require a high degree of interpretability; however, only a few studies have been devoted to solving the interpretability problem of GNN. Therefore, improving the explainability of graph data has become a research direction that needs to be further explored. Especially when dealing with large-scale corpora, complex semantic relations often lead to the construction of heterogeneous graphs that become extremely large. In this case, using the full batch training method means that the intermediate state of all nodes needs to be stored in memory, which brings a huge memory consumption problem. Not only does this restrict the model's scalability, but it also leads to increased computational expense.

In response to this problem, future research work can focus on how to efficiently process large text data while reducing the complexity of the model. On the one hand, the GNN architecture can be designed more compact and efficient to reduce unnecessary computing overhead and memory usage. On the other hand, innovative training approaches and optimization techniques can be investigated to enhance the model's training effectiveness and interpretability. In addition, domain knowledge and prior information can be combined to introduce more interpretable elements into the GNN model, thus further improving the accuracy of text classification and significantly reducing memory consumption and computing costs. These efforts will provide strong support for the wide promotion and in-depth development of GNN in practical applications.

4 Conclusions

Within the realm of text classification, GCNs encapsulate textual data in the form of a graph, which facilitates a more intuitive representation of the intricate relationships among textual elements. Capitalizing on the connectivity between nodes, the graph-based network architecture is adept at retaining comprehensive graph-wide information, thereby enhancing the efficacy of text categorization tasks. This paper summarizes the development of GCN in recent years, analyzes the advantages of each model, analyzes the limitations of GCN, and looks forward to the development of GCN in text classification. In general, GCN technology has shown a rapid development trend in recent years, especially the increasingly significant integration trend with the Transformer model, which has injected new vitality into GCN and expanded its application range. Researchers are actively exploring more potential value of GCN and exploring its integration path with different technologies, to promote the joint optimization of text classification tasks and GCN models and contribute new forces to the development of NLP.

References

1. T. Yingying, W. Junli, Z. Chaobo, Review of text classification methods based on graph convolutional network, Computer Science, **49**(8), 205-216 (2022)
2. K. Church, P. Hanks, Word association norms, mutual information, and lexicography, Comput. Linguist. **16**(1), 22-29 (1990)
3. K. Wang, S.-C. Han, J. Poon, Inductgcn: Inductive graph convolutional networks for text classification, in Proceedings of the 26th International Conference on Pattern Recognition (ICPR), IEEE, (2022), 1243-1249
4. F. Wu, A. Souza, T. Zhang, et al., Simplifying graph convolutional networks, in Proceedings of the International Conference on Machine Learning, PMLR, (2019), 6861-6871
5. L. Huang, D. Ma, S. Li, et al., Text level graph neural network for text classification. arXiv preprint, arXiv:1910.02356 (2019)

6. H. Gao, Y. Chen, J. Jis, Learning graph pooling and convolutional operations for text representation, in Proceedings of the World Wide Web Conference, (2019), 2743-2749
7. Y. Dai, L. Shou, M. Gong, X. Xia, Z. Kang, Z. Xu, D. Jiang, Graph fusion network for text classification, Knowl. Based Syst. **236**, 107659 (2022)
8. K. Dong, Y. Liu, F. Xu, et al., DCAT: Combining multisemantic dual-channel attention fusion for text classification, IEEE Intell. Syst. **38**(4), 10-19 (2023)
9. Y. Lin, Y. Meng, X. Sun, et al., BertGCN: Transductive text classification by combining GCN and BERT, arXiv:2105.05727 (2021)
10. X. Xu, Z. Xu, Z. Ling, et al., Comprehensive implementation of TextCNN for enhanced collaboration between natural language processing and system recommendation, in Proceedings of the International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2024), SPIE 13180, (2024), 1527-1532
11. H. Chang, L. Xiaozhong, Fusion of graph embedding and BERT embedded text categorization model, Learn. Tianjin Univ. Sci. Technol. 1-9 (2024)
12. L. Hou, J. Liu, X. Yu, J. Du, Review of graph neural networks, Computer Science, **51**(6), 282-298 (2024)