

Research on Topic Fusion Graph Convolution Network News Text Classification Algorithm

Li Liu

School of Computer Science and Technology
Xi'an Technological University
Xi'an, China
E-mail: Lily_Liu20@163.com

Yongyong Sun

School of Computer Science and Technology
Xi'an Technological University
Xi'an, China
E-mail: yongsunjd@126.com

Abstract—In the face of a large amount of news text information, how to make a reasonable classification of news text is a hot issue of modern scholars. To solve the problem that only word co-occurrence was considered in the Text Graph Convolutional Network (Text-GCN) method to build a graph model, a news text classification algorithm which fuses themes and is based on Graph Convolution Network, is presented. Firstly, the LDA topic model is used to process the corpus to obtain the distribution of themes of the corpus. Secondly, a graph model is built to construct a global map by using the related topic words and their subject distribution in each article. Finally, the text graph is input into the Graph Convolution Network layers to compute the learning representation of combining feature in order to complete the text classification task. The experimental results show that this method can effectively realize the word level interaction of information in text. In the experiment on Chinese and English datasets, adding theme information improves the accuracy by 1% compared with the Text-GCN method.

Keywords-Text Graph Convolutional Network; Graph Convolutional Network; Latent Dirichlet Allocation Topic Model(LDA topic model); Graph Model

I. INTRODUCTION

With the rapid development of the Internet and its related fields, a large amount of data is continuously imported into the network, of which text data accounts for the vast majority. In the environment of information explosion, obtaining news information through the network has become an important way for people to understand the world and the era of science and technology. How to effectively analyze and use these data has become an urgent problem to be solved. Text classification is the work of mapping a text to one or more categories by analyzing the inherent characteristic information of the text. It is the basis of intelligent text reading. Text classification is an indispensable part of analyzing and processing a large number of text data, which is very challenging. It can serve a variety of downstream tasks, such as Machine Translation, Text Summarization, Recommendation Systems, etc., and is the focus of academic and industrial circles. Therefore, to a certain extent, accurate text classification can not only facilitate its downstream tasks, but also shorten the time for users to search text information and avoid information overload.

In the field of natural language processing (NLP), scholars have been trying to take the characters, words and sentences contained in documents as the starting point, and use their own features and associated features to integrate learning documents to achieve the final classification task. Text classification tasks can be divided into two categories: manual classification and automatic classification [1]. Manual classification methods often use professionals to interpret the content, compile a set of language rules after translation, and then classify according to this rule. This method only relies on manual processing, and there are great challenges in the face of a large number of data sets. It usually costs a lot of time and expensive cost, and the classification accuracy is easily affected by human factors. Automatic classification methods can be divided into machine learning based and depth learning based methods. Although the method based machine learning relies on a variety of calculation methods such as Bayesian, Laplace, Laplace's derivation of least square method and Markov chain to process information and can bring significant performance improvement, it is still limited by the characteristics of manual design. With the continuous development of deep learning technology, neural network provides a new way to solve complex problems in various fields, such as intelligently extracting text content and constructing text feature learning documents by using Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) or Graph Neural Network (GNN).

In recent years, further research on GNN and graph embedding has also attracted extensive attention of scholars [2-3]. Generally, in the task of text classification, the feature engineering of text information often uses sequence data for expression, and how to effectively express text information with graph data structure, that is, the

representation method of text is an important basis for the further development of graph neural network and graph embedding in the direction of NLP. This paper constructs a large heterogeneous graph, which can model the self-information and auxiliary information of the text, and connect the explicit information and potential semantic information of the text through co-occurrence. At the same time, a text graph convolution classification algorithm is proposed. This method can effectively capture the three kinds of information of documents, topics and words in heterogeneous information networks, as well as the relationship between them, enhance the semantic information contained in document nodes, and effectively alleviate the semantic deficiency caused by the graphic representation of text.

II. RELATED WORK

The text classification problem is divided into two parts: feature engineering and classifier. Feature engineering transforms text into text representation data. The traditional method is mainly manual processing, and the deep learning method takes word embedding as the core. The classifier uses Support Vector Machine (SVM), decision tree or neural network to classify the text representation data with features. The text classification method based on deep learning takes the pre-labeled samples as training data, constructs a learning model, learns and infers the internal relationship between the text and its labels, and classifies and predicts the new text set according to the learned correlation [1]. It can complete the text representation in large-scale text classification tasks. Some recent studies have shown that the effectiveness of word embedding can largely affect the accuracy of deep learning classification methods [4].

A. Text representation of sequence information

Traditional feature engineering uses word2vec and Bag-of-Word methods to realize text representation. However, the words in the text are not simply listed, but have certain rules. Kim [5] proposed TextCNN (Text Convolutional Neural Networks) model, which maps the text into a vector, then uses multiple filters to capture the local semantic information in the text as features, and finally obtains the probability distribution of classification labels through the full connection layer. Lai [6] et al. proposed Recursive Convolutional Neural Networks (RCNN), considering the context representation of words to obtain the probability representation of labels. This kind of text representation gives priority to the order or local information of the text after text data preprocessing. Although they can well capture the semantic and grammatical information in continuous sequences, they do not analyze the internal structure of the text.

B. Text representation of graphic information

In the field of natural language, text has not only the sequential expression of context, but also the internal graph structure. For example, syntactic and semantic parsing trees define the syntactic and semantic relationships between words in sentences. TextRank [7], the earliest one based on graph model, proposed to represent natural language text as a graph $G(V, E)$, V represents a group of nodes, and E represents a group of edges between nodes. Nodes can represent various types of text units, such as words, phrases or sentences in a text. Edges represent different types of relationships between any node, such as lexical or semantic relationships, context overlap relationships, and so on. Under the influence of the TextRank model, scholars take words, phrases and sentences in the document as nodes to build different graph models to effectively express text features. Yao [8] et al.

proposed the Text-GCN (Text Graph Convolutional Networks) model, which is based on the whole corpus. By constructing text maps and using GCN to jointly learn the maps, words and documents can be embedded, and then the classification results can be obtained through *softmax* function. The author uses words and documents as nodes to build a heterogeneous graph. Considering that global word co-occurrence will carry discontinuous and long-distance semantic information, the author calculates edge weights through TF-IDF, and uses word frequency and document frequency to build edges between word nodes and document nodes. The edges between two word nodes are represented by word co-occurrence. Text-GCN model transforms text classification problem into node classification problem, which improves the accuracy of classification results. However, this method ignores the relationship between context words in the text, and does not consider the fine-grained text level word interaction. Instead, all words are regarded as independent individuals without upper and lower attachments, and some document information is lost. In order to solve this problem, Zhang [9] et al. proposed to express inductive words through graph neural network, that is, texting. The author believes that each text is a separate graph, in which text level word interaction can be learned, and graph neural network is added to train the text to describe the detailed word-word relationship. Liu [10] et al. proposed TensorGCN (Tensor Graph Convolutional Networks) to describe multiple semantic, grammatical and contextual information by constructing text graph tensors. Because the above method considers the context, the model can learn the emotional preference of the text by capturing the semantic emotional information of the context, and the accuracy of the emotional text classification task has been improved. However,

when considering fine-grained word-word information, these methods construct multiple graphs for each text to cover the context information contained in the text, and then use multi-layer convolution calculation to realize text classification. The calculation and storage of multiple information graphs add a lot of storage and computational burden to the algorithm.

To sum up, in order to improve the classification effect, this paper proposes a topic fusion graph convolution neural network text classification algorithm. In this method, all words in the corpus are considered. By inserting topic information, LDA Topic Heterogeneous Graph (LDA-THG) is generated to further improve the accuracy of text classification.

III. METHOD

In this paper, the text classification model based on LDA-THG is proposed. After preprocessing the text documents, the LDA Topic model (Latent Dirichlet Allocation Topic Model) is used to conduct topic training on the global corpus. Then, the probability distribution obtained by training is combined with the self-information of each text to conduct text modeling. This method constructs semantic association information among words, topics and documents, and uses graph convolutional network to process the embedded information of document nodes and their neighborhood nodes, so as to alleviate the problem of insufficient semantic information when using graph as text representation in text classification. The framework of text classification algorithm model based on LDA-THG graph convolution is shown in Figure 1.

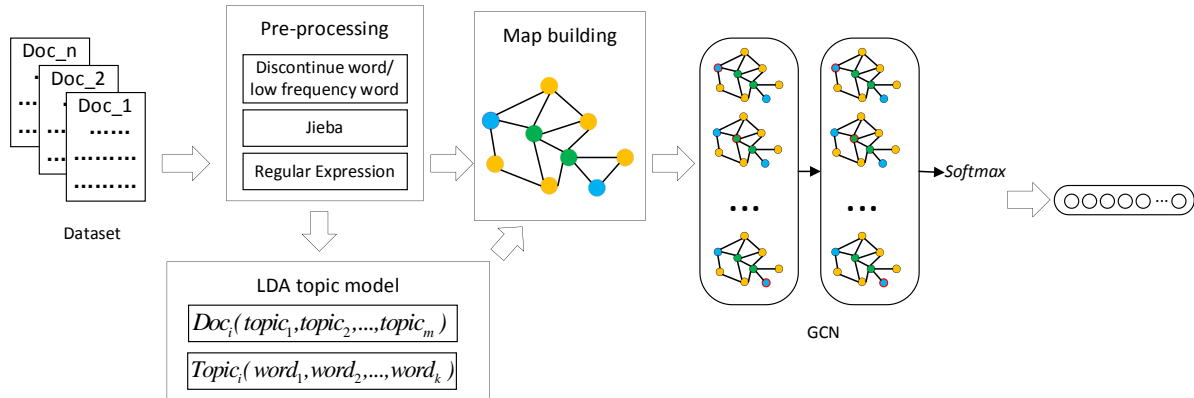


Figure 1. Framework of Graph Convolution Text Categorization Algorithm Based on LDA-THG

A. Graph Convolutional Network

Graph Convolutional Network [11] is a multi-layer neural network that transforms the traditional convolutional neural network that processes regular spatial structure data into a multi-layer neural network that processes graph data. Generally, when processing graph G , it is assumed that each node in the graph can form a connection with itself, that is $\forall v \in V$ and

$(v, v) \in E$. Let $X = (x_1^T, x_2^T, x_3^T, \dots, x_n^T)$, $X \in \mathbb{R}^{n \times m}$ be a matrix containing all nodes and their characteristics, where m is the dimension of the eigenvector, and each row $x_v \in \mathbb{R}^m$ is the eigenvector of node v . The adjacency matrix $A \in \mathbb{R}^{n \times n}$ and degree matrix $D \in \mathbb{R}^{n \times n}$ of graph G

are introduced, where $D_{ii} = \sum_j A_{ij}$. Due to self-connection and self-circulation, the diagonal element of adjacency matrix A is set to one, where $A_{ii} = 1$. When GCN only passes through one-layer of convolution, it can capture the nearest neighbor information of nodes. When multiple GCN layers are stacked, it will integrate more information.

For one-layer GCN, the new k-dimensional node characteristic matrix $H^{(1)} \in \mathbb{R}^{n \times k}$ is calculated as:

$$H^{(1)} = \sigma(AW^{(0)}) \quad (1)$$

Where, $A = D^{-1/2}AD^{-1/2}$ is the normalized symmetric adjacency matrix, $W^{(0)} \in \mathbb{R}^{m \times k}$ is the weight matrix, and $H^{(0)} = X$. σ is the activation function, such as ReLU. Similarly, higher neighborhood information can be merged by stacking more GCN layers, namely:

$$H^{(l+1)} = \sigma(AL^{(l)}W^{(l)}) \quad (2)$$

Where l represents the number of layers of the graph, $H^{(l+1)}$ is the output characteristic matrix of layer l, and $W^{(l)}$ the weight matrix of layer l.

B. Construction of LDA-THG

In graph-based convolution algorithm, LDA-THG constructs a large heterogeneous graph.

It is a fusion of topics and contains three types of nodes-words, documents and topics. The structure of LDA-THG diagram is shown in Figure 2. This method can explicitly model and realize topic and global word co-occurrence, making it easier to use GCN for joint learning reasoning.

In Figure 2, Doc_n represents document node, Topic_m represents the topic node extracted from all documents in the corpus through LDA model, word_k is the node of the only word in the corpus. The curves in the figure are the edges of figure G, which connect document nodes and word nodes, word nodes and topic nodes, word nodes and word nodes, document nodes and topic nodes, respectively. The number of nodes is $|V| = N + K + M$ in the text graph G, where N is the number of all documents in the corpus, K is the number of all non repeated words in the corpus, M is the number of topics extracted through LDA model.

In this paper, one-hot encoding is used to initially represent the features, that is, the feature matrix of the graph is simply set as a unit matrix, and each node in the graph is represented as a one-hot vector as the input of GCN. For the co-occurrence information of global words, a fixed size sliding window is used to collect the co-occurrence statistical information of all documents in the corpus. For topic information, the whole corpus is processed by LDA topic model to obtain topic distribution θ and word distribution φ . The distribution process of topics and words obtained by the LDA model is shown in Figure 3.

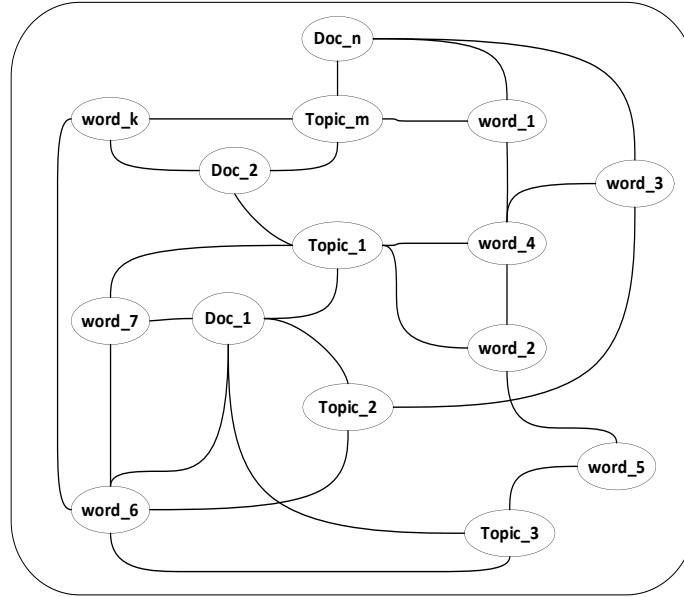


Figure 2. The structure of LDA-THG

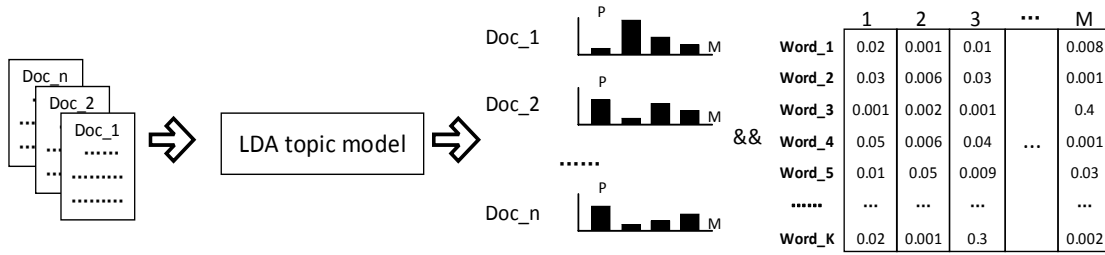


Figure 3. The process diagram of topic and word distribution obtained by LDA model

According to the word frequency in the document, the coexistence of words in the whole corpus, and the topic distribution and word distribution obtained by LDA processing the whole corpus, the edges between nodes are constructed. The weight of the edges between two nodes i and j are defined as:

$$A_{ij} = \begin{cases} PMI(i, j) & , i \text{ and } j \text{ are words} \\ DT(i, j) = \theta_{ij} & , i \text{ is doc, } j \text{ is topic} \\ TF_IDF(i, j) & , i \text{ is doc, } j \text{ is word} \\ TW(i, j) = \varphi_{ij} & , i \text{ is topic, } j \text{ is word} \\ 1 & , i = j \\ 0 & , \text{otherwise} \end{cases} \quad (3)$$

In formula (3), PMI[8] ratio is point level mutual information, which represents the association measure of a popular word and can be used to calculate the association weight between word node i and word node j . And the value of PMI is always greater than zero. $DT(i, j)$ is the weight of the edge between the document node and the topic node, θ_{ij} is the probability that the topic of the i text is j . $TF_IDF(i, j)$ is the weight of the edge between the document node and the

word node. It is the product of the word frequency of word j in document i and the inverse document frequency of word j . $TW(i, j)$ is the weight of the edge between the topic node and the word node, φ_{ij} is the probability that the content described by word j is related to the i -th topic.

C. Model training

After the heterogeneous graph is successfully constructed, the text graph is input into the GCN layer. As the input text graph is a heterogeneous graph with three types of nodes and interactive information between nodes, a large feature matrix $X \in \mathbb{R}^{N \times |V|}$ is used to represent the LDA-THG, namely:

$$X = X_D \oplus X_T \oplus X_W \quad (4)$$

Among them, \oplus is the connecting symbol of the matrix to splice the three matrices. $X_D \in \mathbb{R}^{N \times |V|}$, $X_T \in \mathbb{R}^{N \times |V|}$, $X_W \in \mathbb{R}^{N \times |V|}$ are the characteristic matrix representing document nodes, topic nodes and word nodes, successively. Each row in these matrices represents the eigenvector of a node, i.e. $x_i \in \mathbb{R}^{1 \times |V|}$.

In this paper, three-layer GCN is selected in the graph convolution network layer to process the input features and send them into a *softmax* classifier to predict the label of the document. The calculation process is as follows:

$$Z = \text{softmax}(A\sigma(A\sigma(AW^{(0)}W^{(1)}W^{(2)}))) \quad (5)$$

With formula (1) and (2), the weight parameters of the first, second, and third layers are, respectively, $W^{(0)}$, $W^{(1)}$, $W^{(2)}$. They are trained and optimized by the gradient descent method, and $\text{softmax}(x_i) = x_i / \sum_i x_i$. The loss function used is the cross-entropy error on all documents, and $L2$ regularization term is added to reduce the over fitting of the model. We define D_{train} to represent the set of text indices used for training, C to represent the number of label categories, Y to represent the corresponding tag indicator matrix, Θ to represent model parameters and η to represent regularization factors. The calculation method is as follows:

$$\mathcal{L} = -\sum_{i \in D_{train}} \sum_{j=1}^C Y_{ij} \cdot \log Z_{ij} + \eta \|\Theta\|_2 \quad (6)$$

In formula (4), $E_{1,2} = A\sigma(AW^{(0)}W^{(1)})$ contains the embedding of first-layer and second-layer documents, words and topics, while $E_3 = AReLU(AReLU(AW^{(0)}W^{(1)}W^{(2)}))$ contains the embedding expression of semantic information fusion of documents, words and topics without direct connection. In the heterogeneous graph with three types of nodes, three-layer GCN can transfer information between nodes up to three steps away, so as to realize the information exchange between documents and topics without direct edges in the graph. In the preliminary small sample experiment, the two-layer GCN makes no information transmission between some documents and topics and their similar unconnected nodes, while the three-layer GCN will perform three propagation steps in the forward transmission process and

effectively convolute the third-order neighborhood of each node to realize information exchange.

IV. EXPERIMENT

A. Dataset

In order to verify the effectiveness of the model, this paper selects three data sets for experimental verification, namely, two public benchmark English text data sets 20NewsGroup and AGNews, and a self-collected Chinese dataset Ch_News. Among them, the English data set 20NewsGroup collected about 18,846 newsgroup documents, which were divided into 20 newsgroups sets with different topics, including three kinds of data information-document content, document title and document attribution tag. And there were no duplicate documents in the corpus. The English data set AGNews was published by Cornell University in 2004. It contains nearly one million news articles, which are mainly divided into four categories-World, Sports, Business and Sci/Tec. 8000 news articles were selected as corpus set in this paper. And the data volume of the four categories of news in these 8000 news articles is equal. The Chinese dataset Ch_News is made by manually compiled and tagged about 1,000 news texts from Ecms.cn and english.chinamil.com.cn. The data come from five categories, namely, culture and education, finance and economics, military equipment, biotechnology and sports.

For Chinese and English datasets, different processing is carried out according to their

language characteristics. For two English datasets, there are spaces between words to segment, and the preprocessing of removing low word frequency and stop words is directly carried out. For the Chinese dataset, according to the characteristics that there is no space to separate two characters in Chinese and the news contains a large number of domain knowledge entities, an entity dictionary is built, which contains the knowledge nouns in each domain in the dataset, and then the preprocessing of removing stop words, Chinese word segmentation and low-frequency words is carried out in turn.

Table 1 summarizes the statistics for the three data sets. Each data set takes 70% of the data as the training set, 30% as the test set, and then randomly selects 10% of the data in the training set as the verification set.

B. Comparative experiment

In order to verify the effect of the model on the text classification task, experiments are carried out based on tensorflow and keras, and the programming language is python3.6. In this paper, SVM+LDA, TextCNN and Text-GCN are used to compare with the proposed method. In the experiments of the three benchmark models, the parameters set in the original paper or the default parameters are used to reproduce. In the experiments of the three benchmark models, the parameters set in the original paper or the default parameters are used to reproduce. The three algorithms are briefly introduced as follows.

TABLE I. STATISTICAL TABLE OF DATA SETS USED IN EXPERIMENTS

Dataset	Total docs	Training docs	Test docs	Total words	Categories
20NewsGroup	18,846	13,192	5,654	42,739	20
AGNews	8,000	5,600	2,400	32,653	4
CH_News	1,080	756	324	13,324	5

SVM + LDA The method extracts features through LDA topic model and uses SVM classifier to realize text classification [13].

TextCNN The method uses convolution neural network[21], using randomly initialized word embedding vectors as input, convoluting text matrix with filters of different lengths, and then using max pooling to extract vectors for each filter.

Text-GCN[8] The model construct word co-occurrence heterogeneous graph as graph model, and use graph convolution neural network to jointly learn the embedding of words and documents. After convolution, the feature representation vector learned for each node is obtained, which implies the prediction label of each document, so as to realize classification.

C. Result comparison and analysis

Text classification is a basic and classical task with classical evaluation parameters. Considering that this model solves the problem of multiple text classification. Accuracy is selected as the evaluation index in this paper, and its calculation method is as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (7)$$

Where TP represents the number of correctly identified positive samples, FP represents the number of incorrectly identified positive samples, FN represents the number of incorrectly identified negative samples, and TN represents the number of correctly identified wrong samples.

Table 2 shows the accuracy test of text classification task for three data sets using different methods. By comparison, the method proposed in this paper has achieved better classification results on three data sets, especially

in the self-constructed Chinese data set, which proves the effectiveness of the model.

TABLE II. COMPARISON OF EXPERIMENTAL RESULTS

DataSets \ Model	20new sgroup	A GNews	Ch_ News
SVM+LDA	0.6827	0.7241	0.6956
CNN	0.7160	0.8059	0.6893
Text-GCN	0.8634	0.9245	0.7402
LDA-THG+GCN	0.8796	0.9310	0.7594

The main reasons why using LDA-THG and graph convolution method is superior to the other three methods are: 1) text graph can capture the self-information of document and subject related auxiliary information at the same time; 2) GCN can well capture the information between high-order neighborhood nodes, and can calculate the new characteristics of nodes as the weighted average of themselves and their third-order neighbors. And the document and topic nodes in LDA-THG integrate the potential information of each other and enrich the expression of text features.

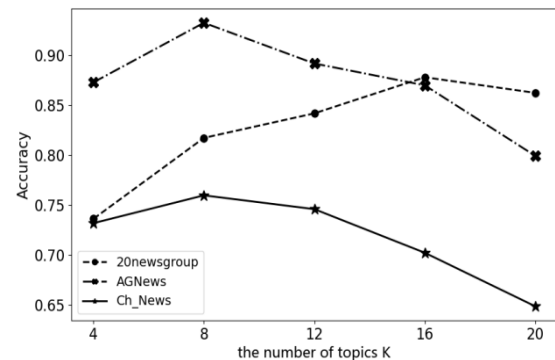


Figure 4. Influence of topic number on classification results of LDA-THG+GCN model on three data sets

The impact of different topic numbers K on the classification results in three data sets is shown in Figure 4. It can be seen that for the three datasets, the categories of the datasets themselves are 20, 4

and 5, respectively. The range of their high classification accuracy is when K is [14,20], [4,12] and [4,12], that is, when K value is within the range corresponding to the category of the data set itself, the accuracy is relatively high. Therefore, when the number of topics is maintained in a specific range, the classification accuracy is high. When the number of topics K exceeds this range, it will decrease in varying degrees. The main reason for this phenomenon is that when there are too many or too few topics, the found topics will be very rough or too detailed, both of which will cause the document to lose some relevant topic information. At the same time, the grammatical and expressive features of the text of English and Chinese lead to a great difference in the classification accuracy of AGNews and Ch_News under similar circumstances.

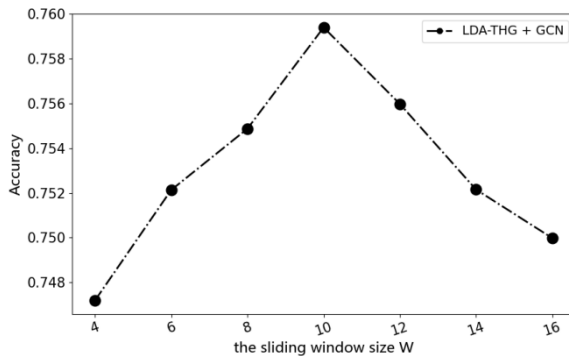


Figure 5. Influence of W on classification results of LDA-THG+GCN model on Ch_News data set

On Chinese data set Ch_News, the influence of the sliding window size W of this model on classification results is shown in Figure 5. As can be seen from the figure, with the increase of the sliding window, the accuracy rate will be improved first and then slowly decrease. Accuracy is highest when the window length is equal to 10. This shows that when the sliding window is too small, sufficient global word co-occurrence information can not be generated to represent the

probability relationship of word co-occurrence, while when the sliding window is too large, edge relations will be added between the nodes that are not close, making the information to be trained redundant. This not only increases the difficulty of training and storage, but also makes the features in the text representation fail to highlight the main features contained in the text itself.

V. CONCLUSION

In this paper, a new method of news text classification based on graph convolution neural network is proposed by combining text topics and words in text. This method uses the topic distribution and the word distribution corresponding to the topic in the LDA model to construct the weights of nodes and edges in the graph model, and generates a heterogeneous network information graph. Then, the point and edge information of the graph is integrated through the multilayer graph convolution neural network and *softmax* layer, and finally the label of the prediction document is obtained. The purpose of this method is to provide more reference knowledge for improving the performance of the model in text classification by integrating the data that helps to understand the text. Future research may further improve the performance of the model and achieve better training results by expanding the number of node layers involved in primary graph convolution, such as deep reasoning of two-layer nodes.

REFERENCES

- [1] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, J. Gao, "Deep learning--based text classification: a comprehensive review," *ACM Computing Surveys (CSUR)*, vol. 54(3), April, 2022, pp. 1-40, doi:10.1145/3439726.
- [2] H. Cai, VW Zheng, KCC Chang, "A comprehensive survey of graph embedding: Problems, techniques, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30(9), Sept. 2018, pp 1616-1637. doi:10.1109/TKDE. 2018.2807452.

- [3] PW Battaglia, JB Hamrick, V. Bapst, et al. "Relational inductive biases, deep learning, and graph networks," arXiv preprint, 2018. doi:10.48550/arXiv.1806.01261.
- [4] D. Shen, G. Wang, W. Wang, et al. "Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms," arXiv preprint, 2018, doi:10.48550/arXiv.1805.09843.
- [5] Chen Y, "Convolutional neural network for sentence classification," University of Waterloo, August 2015, <http://hdl.handle.net/10012/9592>.
- [6] S. Lai, L. Xu, K Liu, J. Zhao, "Recurrent convolutional neural networks for text classification," Twenty-ninth AAAI conference on artificial intelligence, 2015.
- [7] R. Mihalcea, P. Tarau, "Textrank: Bringing order into text", Proceedings of the 2004 conference on empirical methods in natural language processing, 2004, pp 404-411.
- [8] L. Yao, C. Mao, Y. Luo, "Graph convolutional networks for text classification," Proceedings of the AAAI conference on artificial intelligence, vol. 33(01), July 2019, pp 7370-7377, doi:10.1609/aaai.v33i01.33017370.
- [9] Y. Zhang, X. Yu, Z. Cui, et al. "Every document owns its structure: Inductive text classification via graph neural networks," arXiv preprint, April 2020, doi:10.48550/arXiv.2004.13826.
- [10] X. Liu, X. You, X. Zhang, J. Wu, P. Lv, "Tensor graph convolutional networks for text classification," Proceedings of the AAAI conference on artificial intelligence, vol. 34(05), April 2020, pp 8409-8416, doi:10.1609/aaai.v34i05.6359.
- [11] TN Kipf, M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint, September 2016, doi:10.48550/arXiv.1609.02907.
- [12] X. Zhang, J. Zhao, Y. LeCun, "Character-level convolutional networks for text classification," Advances in neural information processing systems, 2015.
- [13] DM Blei, AY Ng, MI Jordan, "Latent dirichlet allocation," Journal of machine Learning research, January 2003, pp 993-1022.
- [14] R. Wang, Z. Li, J. Cao, T. Chen, L. Wang, "Convolutional recurrent neural networks for text classification," 2019 International Joint Conference on Neural Networks (IJCNN), IEEE, September 2019, pp 1-6, doi:10.1109/IJCNN.2019.8852406.
- [15] Y. KIM, "Convolutional neural networks for sentence classification," 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp 1746-1751.
- [16] Q. Li, H. Peng, J. Li, et al. "A survey on text classification: From shallow to deep learning," arXiv preprint, August 2020, doi:10.48550/arXiv.2008.00364.
- [17] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, D. Brown, "Text classification algorithms: A survey," Information, vol 10(4), October 2019, pp 150, doi:10.3390/info10040150.
- [18] M. Malekzadeh, P. Hajibabae, M. Heidari, S. Zad, O. Uzuner, J. H Jones, "Review of graph neural network in text classification," 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE, December 2021, pp 0084-0091, doi:10.1109/UEMCON53757.2021.9666633.
- [19] S. Ghosh, S. Maji, MS. Desarkar, "Graph Neural Network Enhanced Language Models for Efficient Multilingual Text Classification," arXiv preprint, March 2022, doi:10.48550/arXiv.2203.02912.