The Quality Evaluation of Online Book Reviews Based on Graph Neural Network

Can Zhang†, Wei Liu†, Qi Mu†, Deshan Zhang† and Fancheng Meng†

† Inspur Electronic Information Industry Co., Ltd., Jinan 250101, China

*Corresponding author’s e-mail: zhangcan02@inspur.com

Abstract. With the development of the Internet, the way people read has changed, the readers are more and more willing to share their reading opinions via the reading platforms. However, as the continuous increase in the number of users and reviews on the book platforms, the review data is growing exponentially. It is difficult for people to find useful information quickly in the massive reviews. Therefore, using automated methods to identify high-quality reviews among a large number of book reviews has become a hot topic in current research. This paper studies a quality evaluation method of book reviews based on graph convolution neural network, which introduces the LDA document topic generation model to add topic nodes on the basis of review and word nodes, builds the relationship between the three nodes as edges to generate heterogeneous graph. The graph is sent to a two-layer GCN model for training. In the last, this article did experimental evaluation and analysis based on the book review data of non-commercial platforms. The results show that the proposed method is superior to the machine learning method and the traditional neural network model.

1. Introduction

In recent years, the Internet has developed rapidly, and books have become more and more connected to the Internet. The rise of online bookstores, the promotion of e-books, and the establishment of book platforms have led to the rapid development of online book reviews. However, due to the large number of comments and the repetitive content, the readers have to read a large number of comments to obtain useful information [1]. Therefore, using automated methods to identify high-quality reviews among a large number of book reviews has become a hot topic in current research.

However, Graph Neural Network (GNN) has attracted wide attention from researchers due to its excellent ability to deal with irregular graph structures. Yao et al. [2] proposed the graph convolutional text classification model Text-GCN that organizes the training text, test text and words in the corpus into the same network as nodes, and then uses graph convolutional network to extract node features and classify text nodes. But the model doesn’t consider the impact of topic on the quality of the review, causing the lack of semantic information. So this paper introduces the LDA document topic generation model [3], and studies a new method based on graph neural network for evaluating the quality of book reviews. On the basis of review node and word node, topic node is added through the LDA method, and the weights between review node and topic node, topic node and word node are expressed too. So that a new heterogeneous graph network with more semantic information is conducted. Then, the graph is sent to the graph convolutional neural network for classification. This method takes into account the influence of topic on comment quality, enriches the semantic information of heterogeneous graph and learn the global structural relationship between nodes. It provides a new idea for the following research about quality evaluation of reviews, which has great significance for users to quickly obtain the...
important information in reviews quickly. In order to verify the effectiveness of this method, the experiment was conducted using SVM, logistic regression machine learning algorithm, CNN and Text-GCN on the same data set. The results show that the proposed method is superior to the machine learning method and the traditional neural network model.

2. Related Work

2.1. Text classification
Text classification was originally carried out mainly through expert rules and expert systems. In the 1990s, machine learning methods began to be used to classify text through artificial feature engineering and shallow classification models. Traditional machine learning requires human experts to analyze which feature is more important, such as shape, texture, location, and encode them as a data type. Then the machine analyzes these features in historical data to find the results caused by different combinations. Trstenjak et al. [4] used term frequency-inverse document frequency as text features, and the K-Nearest Neighbor method (KNN) as the classification function. Shima et al.[5] used latent semantic indexing (LSI) to extract low-dimensional vector features of text, and used support vector machines (SVM) to guide the extraction process. The performance of most machine learning algorithms depends on the accuracy of the extracted features.

In recent years, neural networks have gradually replaced traditional machine learning as the basic model of natural language processing tasks due to their powerful representation capabilities. Kim et al. [6] used CNN to extract local semantic features of the text, and Liu et al. [7] used LSTM to model the text. Zhou et al. [8] combined the advantages of Max Pooling in LSTM and CNN, and proposed BLSTM-2DCNN. However, the classic neural network model can only process data with Euclidean structure, and cannot be applied to data with non-Euclidean structure. For example, the nodes in an irregular network have no order, and the number of each node’s neighbor is not the same, which makes the model unable to run traditional convolution operations on the network.

2.2. Graph neural networks
In the field of natural language processing, text structure, syntax and even sentences exist in the form of graph data. In recent years, graph neural network have received widespread attention due to their excellent ability to deal with irregular graph structures, and have been successfully applied in many fields of natural language processing. In 2016, Defferrard et al. [9] first introduced graph convolutional networks to text classification tasks. Later, Kipf et al. [10] simplified the definition of spectrogram convolution and converted the text documents to document-word graphs model, which improved the computational efficiency, and have achieved good classification results on some benchmark data sets. In 2019, Yao et al. [2] proposed the graph convolutional text classification model Text-GCN on this basis. The Text-GCN model organizes the training text, test text and words in the corpus into the same network as nodes, and then uses graph convolutional network to extract node features and classify text nodes. This method has achieved the best classification effect currently on multiple open source data sets.

3. Methodology

3.1. Problem Description
This article needs to evaluate the review quality based on the text information of the book review. The classification problem that can be transformed into that as follows: It is known that a set of user evaluation information \( O = \{O_1, O_2, O_3, \ldots, O_n\} \), \( O \) represents evaluation information of every user, and a predefined category \( C = \{c_1, c_2\} \). Our aim is to find a mapping model \( F(\cdot) \), so that \( \forall O_i \in O \rightarrow C \). For the evaluation of review quality, \( c_1 \) means useful comment, and \( c_2 \) means useless comment.
3.2. Heterogeneous graph construction

The graph is a structure with powerful data representation capabilities. In the review quality classification system, the review nodes, word nodes, and topic nodes are represented as different types of nodes in the graph. Specifically, we construct a heterogeneous graph $G(V, E)$, where $V$ represents the set of all nodes, and $E$ represents the set of all edges. $V$ contains three types of nodes: review node $U = \{u_1, ..., u_m\}$, word node $I = \{i_1, ..., i_n\}$, topic node $T = \{t_1, ..., t_l\}$, $|V| = m + n + l$, each node is represented as a feature vector.

Similar to the common graph neural network method for text classification, the weight between the review node and the word node is the word frequency inverse document frequency (TF-IDF) of the word in the comment. The edges between word nodes are based on global word co-occurrence information. A fixed-size sliding window is used to slide across the entire corpus (shifting one word to the right each time) to count word co-occurrence information, and then point mutual information (PMI) is used to calculate the weight of the connection between two word nodes. When $\text{PMI} > 0$, an edge is established between two word nodes, and the PMI value of the two words is the weight of the edge.

When adding topic nodes to determine whether there are edges between review nodes and topic nodes, topic nodes and word nodes, and their weights, this paper adopts the LDA document topic generation model. The core formula can be simply expressed as:

$$P(\text{word}|\text{review}) = \sum_{\text{topic}} P(\text{word}|\text{topic}) \times P(\text{topic}|\text{review})$$

The probability graph model is shown in the Figure 1. Among them, $\theta$ represents the probability distribution of a topic in a comment, $\alpha$ represents the Dirichlet distribution parameter of the comment-topic distribution (the prior probability of $\theta$), $\phi$ is the probability distribution of words contained in a certain topic, and $\beta$ is the Dirichlet distribution parameter of topic-word distribution (prior probability of $\phi$), $W$ represents the number of words in each review, $D$ represents the number of reviews, $\theta \rightarrow z$, $\phi \rightarrow w$ all obey the multinomial (Multinomial) distribution. The probability graph of LDA is described as follows:

- According to the Dirichlet distribution $\text{Dir}(\alpha)$, the topic probability distribution $\theta$ of each comment is obtained, and according to the Dirichlet distribution $\text{Dir}(\beta)$, the probability distribution $\phi$ of the words under each topic is obtained.

- For the $i$-th word in a comment, first extract a topic $z_i$ from the polynomial distribution $\theta$ of each topic contained in the comment, and then extract a word $w_i$ from the polynomial distribution $\phi$ of the word corresponding to this topic.

- Repeat the previous step until all $D$ comments are generated.

![Figure 1 Schematic diagram of LDA process](image-url)
same way, an edge is established between the comment and the topic node in the comment-topic matrix, and the weight is the corresponding probability.

3.3. Graph Convolutional Neural Network Classification

After the heterogeneous graph is constructed, it will be sent to the two-layer GCN, and the eigenvalues of the first-layer matrix are updated as \( H^0 = \alpha (A \cdot X \cdot W_0) \), where \( A = D^{-\frac{1}{2}}AD^{-\frac{1}{2}} \) is the normalized adjacency matrix; \( W_0 \) is the weight matrix; \( \alpha(\cdot) \) is the activation function. The second-level domain information can be obtained as \( H^{l+1} = \alpha (A \cdot H^l \cdot W_1) \), where \( l \) represents the number of layers.

The loss function of the model adopts cross entropy loss, which is
\[
L = -\sum_{d \in y} \sum_{f=1}^{F} Y_{df} \ln Z_{df},
\]
where \( y \) is the index of all labeled nodes. The parameter matrix \( W_0, W_1 \) can be updated using the gradient descent method; \( Y_{df} \) represents the label category; \( Z_{df} \) is the predicted category; \( F \) is the feature dimension of the output layer, which is equal to the number of categories. In this article, \( F = 2 \).

3.4. Evaluation index

The evaluation indicators of the experimental results in this paper are precision rate, recall rate, accuracy rate and F value.

Precision represents the proportion of the number of samples correctly classified as positive to all samples with a positive prediction category, as shown in formula (1). TP (True Positives) refers to the number of samples whose actual category is positive and the predicted category is also positive; FP (False Positive) refers to the number of samples whose actual category is negative but the predicted category is positive.

\[
\text{Precision} = \frac{TP}{TP + FP} \quad (1)
\]

Recall rate represents the proportion of the number of positive samples that are correctly classified as positive samples, as shown in formula (2). FN (False Negative) represents the number of samples in which the actual category is positive but the predicted category is negative.

\[
\text{Recall} = \frac{TP}{TP + FN} \quad (2)
\]

Accuracy represents the proportion of the number of accurately classified samples in the overall result, as shown in formula (3).

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)
\]

The F value is a balanced consideration of precision and recall. The higher the F value, the more robust the classification effect of the model, as shown in formula (4).

\[
F = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)
\]

4. Experiment and Analysis

4.1. Data set

The online comment data used in this article is crawled using a special crawler tool and the corresponding website code. This article selects books in the Top 250 list of Douban Reading website as the research object. A total of 7,500 reviews of 15 books are collected under the three categories of humanities, science fiction, and history. After deleting the reviews repeated or only contain special characters and numbers, 4706 comments were finally retained as the research object. For each review, get the content of the review and the number of useful votes.

According to statistics, among the pre-processed comments, there are 394 comments with useful voting parameters, which are automatically determined as useful comments. For the remaining 4312 comments that lack usefulness votes, they are labeled by manual, with useful comments marked as 1, and useless comments as 0. In the end, 1500 useful and useless comments were randomly selected, 3000 comments in total.
4.2. Experiment
In the experiment, the data is divided into training set and test set according to the ratio of 7:3. The window size is 20. The text is trained for 100 epochs. If the loss function exceeds 10 epochs and does not decrease, the training is stopped. The learning rate is set to 0.02, and the dropout is set to 0.5. By calculating the perplexity of the LDA model under different number of topics, the optimal number of topics of the LDA model is finally selected as 18.

4.3. Experimental analysis
The training results are shown in the Figure 2. Before epoch=10, the model training has not reached the optimal level, and the accuracy rate continues to rise. The reason may be that for the model proposed in this article, node information needs to be propagated in the graph. When it reaches about 10 rounds, the global information of the graph is fully propagated during the training process, and the accuracy of the model fluctuates around 85%.

![Figure 2](image)

Figure 2 Schematic diagram of changes in training accuracy

In order to verify the effectiveness of this method, the experiment was conducted as SVM, logistic regression machine learning algorithm, CNN and Text-GCN on the same data set. The overall experimental results are shown in Table 1.

Experiments show that compared with SVM, logistic regression machine learning algorithm, CNN and other models, the accuracy of this method has been improved to varying degrees. The main reason is that the method in this paper transforms the comment information into a graph structure, not only using the text information of the review nodes, word nodes, but also connecting the reviews with topic, making better use of the neighbor information between the nodes and alleviating the sparseness of the data. It is worth noting that the detection effect of the model proposed in this paper is better than that of Text-GCN, which proves the effectiveness of the model proposed in this paper to introduce topics in constructing text graphs. After adding thematic factors, the semantic information represented by the heterogeneous graph is better enriched.

| Model          | Accuracy (%) |
|----------------|--------------|
| SVM            | 84.6         |
| Logistic regression | 81.2        |
| CNN            | 85.6         |
| Text-GCN       | 87.7         |
| This article   | 90.0         |

Moreover, as shown in Table 2, the review quality evaluation model of this article has a smaller gap between the precision rate, recall rate and F value of the two categories in the sub-category experimental results, and the overall classification effect is better.
Table 2  Comparison of experimental results by category

| Model      | Useful review | Useless review |
|------------|---------------|----------------|
|            | P (%)  | R (%)  | F (%)  | P (%)  | R (%)  | F (%)  |
| SVM        | 87.4   | 89.4   | 88.4   | 67.9   | 63.7   | 65.7   |
| Logistic regression | 81.9   | 94.0   | 87.5   | 73.2   | 43.9   | 54.9   |
| CNN        | 82.1   | 87.2   | 84.6   | 87.1   | 83.0   | 85.0   |
| Text-GCN   | 88.4   | 86.2   | 87.3   | 86.0   | 88.3   | 87.1   |
| This article | 89.2   | 88.7   | 88.9   | 88.2   | 91.5   | 89.8   |

5. Conclusion
On the basis of GNN and LDA, this paper converts the quality evaluation of review into text classification, constructs a heterogeneous graph network based on review, word and topic, and carries out the experiment that achieves good results. The research of this model adds the topic information to construct the heterogeneous graph network, fully embodies the influence of the topic on the quality evaluation of reviews. It provides new ideas for the research about quality evaluation of reviews, helps to distinguish the quality of large-scale reviews, which has great significance for users to obtain the important information in reviews quickly. But there is one more problem, when a new test review occurs, the structure of the graph cannot be modified. In other words, the model cannot dynamically process the new sample. Relevant exploration will be carried out in the future research.

References
[1] Qu, J.S. (2015) Mining Research of Online Chinese Book Reviews Based on Content Classification [D], (Doctoral dissertation, Nankai University).
[2] Yao, L., Mao, C., Luo, Y. (2018). Graph convolutional networks for text classification. arXiv: Computation and Language.
[3] Blei, D. M., Ng, A. Y., Jordan, M. I. (2003). Latent dirichlet allocation. J. Journal of Machine Learning Research, 3(Jan):993-1022.
[4] Trstenjak, B., Mikac, S., Donko, D. (2014). KNN with TF-IDF based Framework for Text Categorization. J. Procedia Engineering, 69(1): 1356-1364.
[5] Shima, K., Todoriki, M., Suzuki, A. (2004). SVM-based feature selection of latent semantic features. J. Pattern Recognition Letters, 25(9), 1051-1057.
[6] Kim, Y. (2014). Convolutional neural networks for sentence classification. In: Empirical Methods in Natural Language Processing. pp. 1746-1751.
[7] Liu, P., Qiu, X., Huang, X., et al. (2016). Recurrent neural network for text classification with multi-task learning. arXiv: Computation and Language.
[8] Zhou, P., Qi, Z., Zheng, S., Xu, J., Bo, X. (2016). Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. In: International Conference on Computational Linguistics. pp. 3485-3495.
[9] Defferrard, M., Bresson, X., Vandergheynst, P., et al. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. In: Neural Information Processing Systems. pp. 3844-3852.
[10] Kipf, T.N., Welling, M. (2016). Semi-supervised classification with graph convolutional networks.