Implicit sentiment analysis based on graph attention neural network

Shanliang Yang1 | Linlin Xing1 | Yongming Li2 | Zheng Chang1

1School of Computer Science and Technology, Shandong University of Technology, Zibo, China
2Institute of Scientific and Technical Information, Shandong University of Technology, Zibo, China

Abstract
Sentiment analysis is one of the crucial tasks in the field of natural language processing. Implicit sentiment suffers a significant challenge because the sentence does not include explicit emotional words and emotional expression is vague. This paper proposed a novel implicit sentiment analysis model based on graph attention convolutional neural network. A graph convolutional neural network is used to propagate semantic information. The attention mechanism is employed to compute the contribution to the emotional expression of words. In order to solve the problem of multiple attention preserving repeated information, orthogonal attention constraint was used to make different attention store different emotional information; given the uneven distribution of emotional information, score attention constraint was proposed to make the model focus on a limited number of essential words. The performance of the proposed model was verified on implicit sentiment datasets. The F value reached 88.16%, which is higher than the benchmark model in the literature. The attention mechanism is analyzed to verify the effectiveness of orthogonal constraint and score constraint.

KEYWORDS
attention constraint, attention mechanism, graph neural network, implicit sentiment analysis

1 | INTRODUCTION
Sentiment analysis is one of the crucial technologies in the field of Neural Language Processing. It analyzes the emotion, attitude, sentiment, opinion, and other factors in the document and recognizes the sentiment categories expressed in the document. With the rapid development of online social media, critical factors have an incremental impact on society, economy, and politics, such as online public opinion, netizen emotions, voter opinion, and crisis events in society. Sentiment analysis has become the core technology of online public opinion analysis and prediction and has been widely concerned and studied in academia and industry. In recent years, researchers proposed numerous sentiment analysis methods, especially with deep learning and neural network. As a result, the performance of the sentiment analysis model has improved rapidly. However, sentiment analysis still faces significant challenges, such as the performance of implicit sentiment analysis is not perfect, and the significance of the attention mechanism for sentiment information extraction is not clear.
TABLE 1  Instance table of implicit sentiment sentences

| No. | Implicit sentiment sentence                                                                 | Sentiment category |
|-----|-------------------------------------------------------------------------------------------|--------------------|
| 1   | After reading it, I am full of memories and many elements of that era.                     | Positive           |
| 2   | I cannot stand the haze. I cannot see my fingers. I do not want to go out these two days! | Negative           |
| 3   | I went to the museum when it opened, recording the history of Chengdu and even Sichuan.    | Neutral            |

Liu\(^1\) divided sentiment analysis into implicit sentiment analysis and explicit sentiment analysis. There is no strict distinguish in previous studies. However, there are apparent differences. Explicit sentiment contains sentiment words, emotional inflection words, adverbs, and others, while implicit sentiment does not contain sentiment words. So it is not easy to judge the sentiment category directly, and the semantic expression is hard to understand. Table 1 shows examples of the implicit sentiment sentence. The example sentences express positive, negative, and neutral sentiments, respectively. Although the words that expressed sentiment directly were not found in the example sentences, they can implicitly express positive or negative sentiment. For example, the sentence “After reading it, I am full of memories and many elements of that era.” clearly expressed positive sentiment.

Implicit sentiment analysis has received more attention in recent years. Neural network and attention mechanism have been utilized to building an implicit sentiment analysis model. For example, Wei\(^2\) and Zuo\(^3\) took advantage of bi-directional long short-term memory (BiLSTM) and graph convolutional network (GCN) to research implicit sentiment analysis of Chinese documents. The challenges of implicit sentiment analysis comprise the following three aspects. Firstly, without emotional words, it is more challenging to extract compelling sentiment features. Secondly, words in sentences belong to objective words or neutral words, which produce different emotions through context semantics. However, it is not easy to extract semantic features of expressing sentiment. Thirdly, implicit sentiment has related to background knowledge, which is hard to define in advance, and the expression of subjective emotional tendency is euphemistic.

According to the above challenges, put forward two hypotheses: firstly, words with language environment affect each other and interact to produce different sentiments; secondly, words in sentences have different importance to sentiment expression. In order to express the mutual influence between words and the differences in sentiment that words express in a sentence, this paper proposes a sentiment analysis model named graph attention convolutional neural network (GACNN), which employs the graph attention neural network. The graph model is used to describe the relationship between words and the relationship between words and sentences. The attention mechanism calculates the importance of the word to sentiment expression. In order to extract more useful information from the sentiment sentence, the attention mechanism was improved by introducing constraint conditions. Firstly, multiple head attentions storage repetitive and similar information in different heads. Inspired by the orthogonal attention model in the literature,\(^1\) the orthogonal attention constraint was used to store different sentiment information to ensure the difference between head attentions. Secondly, there is part of words in sentences which are very important for sentiment expression. Inspired by the constraints of the number of deleted words and reserved words in the literature,\(^4\) the score attention constraint was used to make the attention weight pay attention to several essential words.

The main contributions of this paper are summarized as follows:

- We propose a novel graph attention neural network model for implicit sentiment classification, which extracts context semantic information by constructing a semantic graph with word node and sentence node, and two attention constraint methods are used to capture more critical features for sentiment expression.
- Orthogonal attention constraint is used for preserving the differences among the attentions’ representation in the multi-head attention layer. Score attention constraint is adopted for taking attention to little important words for sentiment expression to meet the hypothetic condition.
- We conduct experiments to evaluate the performance of the proposed model for implicit sentiment analysis on the evaluation dataset and carry out a comparative analysis with baseline models. The experimental results demonstrate that the proposed model can achieve substantial improvements over state-of-the-art baselines.
The rest of the paper is organized as follows: In Section 2, introduce the related research about explicit sentiment analysis, implicit sentiment analysis, and graph neural network (GNN). Section 3 elaborates on the proposed sentiment analysis model based on the graph attention neural network. In Section 4, we conduct implicit sentiment analysis and parameter analysis experiments. Discuss the experiment results, and obtain some meaningful conclusions. Finally, Section 5 summarizes the paper and gives some future research work.

2 | RELATED WORK

In recent years, sentiment analysis has become one hot research direction in natural language processing. According to the solution method, existing works on sentiment analysis can be divided into the rule-based approach, machine learning method, and deep learning method. According to whether explicit sentiment words emerge in the expression, the literature divided sentiment analysis into explicit sentiment analysis and implicit sentiment analysis.1 With the development of the field, implicit sentiment analysis has received more and more attention. This section analyzes the related work of sentiment analysis from several aspects, such as sentiment analysis, attention mechanism, and GNN method. Firstly, we introduce sentiment analysis with state-of-art techniques in the literature. Then, introduce the attention mechanism in sentiment analysis that plays a vital role in extracting features from samples. Eventually, elaborate the methods based on GNNs, and the research motivation of this paper is expounded based on previous research.

2.1 | Sentiment analysis

Literature researching sentiment analysis is abundant. Most literature investigates sentiment from a certain number of aspects, such as word represent method, external knowledge, and the classification model type. We summarize the research from these aspects.

Word representation is essential and fundamental for sentiment analysis. Most literature adopted word embedding as word representation, which uses a real number vector to encode a word, including semantic, syntactic, and sentiment information. Kayvan et al.5 investigated the semantic relations in the document and proposed a sentence-level graph-based text representation to extract the latent and continuous features. Peng et al.6 studied sentiment information for word embedding and proposed an adversarial learning method for training sentiment word embedding.

External knowledge is critical for comprehending the sentiment information contained in sentences or documents, so we must consider external knowledge in the study of sentiment analysis. Chen and Huang7 researched aspect-level sentiment classification by incorporating external knowledge into a neural network and proposed a novel framework for model aspect-opinion pair identification. Vo et al.8 studied a knowledge representation approach that centers on aspect rating and weighting and proposed a novel model that utilizes semantic and syntactic components to capture semantic and sentimental information. Sentiment words present the most sentiment information in text, so a sentiment dictionary is critical for explicit sentiment analysis. Li et al.9 proposed a lexicon integrated two-channel CNN-BiLSTM model, which introduced a novel padding operation named sentiment padding, emphasizing the importance of sentiment word to extract sentiment information. Abdi et al.10 constructed a unified feature considering word embedding, sentiment knowledge, sentiment shift rules, statistical and linguistic knowledge simultaneously.

The type of model employed in the sentiment analysis task has a noticeable influence on classification results. Traditional machine learning and deep learning have been researched, such as naive Bayes, Support Vector Machine, Convolutional Neural Network, Recurrent Neural Network, GNN. Han et al.11 considered the vocabulary and the latent semantic information and proposed a support vector machine method utilizing a Fisher kernel function. Chen et al.12 proposed one divide-and-conquer approach to classify sentences into three types and recognize the sentiment category using Convolutional Neural Network. Akhtar et al.13 proposed a stacked ensemble method, in which three deep learning models were developed and combined the outputs using a perceiver network. Cambria et al.14 integrated top-down learning and bottom-up learning via an ensemble of symbolic and subsymbolic AI tools for polarity detection of text.
2.2 | Implicit sentiment analysis

In the absence of precise sentiment words, implicit sentiment analysis is more complex than explicit sentiment analysis to recognize the sentiment categories of sentences or documents. Recent research focuses on semantic understanding of sentences deeply and integrating external knowledge, and the model of GNN also attracted considerable attention. Fang et al.\textsuperscript{15} utilized a feature-based method to analyze implicit sentiment and recognize aspect level sentiment and opinion level sentiment by mining the potential sentiment model. Cho et al.\textsuperscript{16} proposed a knowledge-based word ambiguity resolution method to reduce the impact of word ambiguity on implicit sentiment analysis. Zuo et al.\textsuperscript{3} proposed a novel context-specific heterogeneous GCN that simultaneously integrates context semantic background and phrase dependency information. Liao et al.\textsuperscript{17} researched fact-implied implicit sentiment recognition of sentences and proposed a multi-level semantic fusion method integrating sentiment target representation, structure embedded representation, and semantic context representation. Wang et al.\textsuperscript{18} proposed a novel model named hierarchical knowledge enhancement and multi-pooling, which integrates the knowledge information from different levels by hierarchical knowledge representation learning.

2.3 | Attention mechanism

The attention mechanism is the critical component in neural network models and plays a vital role in feature extraction for sentiment analysis. Liu et al.\textsuperscript{19} proposed a co-attention mechanism to capture the semantic correlations and generate more excellent feature representation. Li et al.\textsuperscript{20} proposed a lexicon-enhanced network by combing the sentiment lexicon and attention mechanism and incorporating sentiment linguistic knowledge into the deep learning methods. Meb et al.\textsuperscript{21} proposed an attention-based model named ABCDM (An attention-based bidirectional CNN-RNN deep model) considering features with different importance, which extract temporal information by utilizing long short term memory (LSTM) and gated recurrent unit layers. Su et al.\textsuperscript{22} improved the attention mechanism, which extracted context words with active or misleading influence on the prediction based on their attention weights. Gan et al.\textsuperscript{23} proposed a sparse attention model to relieve the defect of CNN and LSTM model, which is composed of a sparse attention layer, multichannel embedding layer, and convolution module.

2.4 | GNN method

GNNs can express associated information and capture the structural relationship in data. The GNN model obtained outstanding performance in social analysis, biological information, and computer vision tasks. The literature provides a systematic summary of the GNN model.\textsuperscript{24,25} In recent years, a large amount of novel GNN models have been proposed, such as GCN, Graph Attention Network (GAT), GaAN (Gated Attention Network), HetGNN (Heterogeneous GNN), Heterogeneity Attention Network (HAN). Velickovi et al.\textsuperscript{26} proposed the graph attention neural network model named GAT, which utilizes a self-attention mechanism to solve the defects of the graph convolution model. This model gives different weights to neighbor nodes so that the model can solve the induction and conduction problems. Zhang et al.\textsuperscript{27} proposed a gated attention graph model named GaAN, which utilizes a gating attention mechanism on large-scale spatiotemporal maps, and the convolutional subgraph is adopted to control the importance of each attention. Zhang et al.\textsuperscript{28} proposed a heterogeneous GNN model named HetGNN, which considers heterogeneous structure information and heterogeneous content information. Wang et al.\textsuperscript{29} proposed a heterogeneous graph attention neural network model called HAN, which considers heterogeneous graphs with different types of nodes and edges, and takes advantage of node level and semantic layer attention to learn important nodes and meta paths. GNN has obtained great success in various tasks, such as text classification and sentiment classification.

In the task of text classification, the GNN is used to construct a graph network by using structural relationship information of phrases in the document, such as word co-occurrence, syntactic relationship, context relationship.\textsuperscript{30,31} Yao et al.\textsuperscript{30} introduced a novel text classification model based on graph convolution neural network, which constructs the phrase relationship graph using word co-occurrence relationship and word-text relationship, and then GCN network is trained on the experiment dataset. Liu et al.\textsuperscript{31} constructed text graph tensor, which describes semantics, grammar, context, and other information. Two algorithms of intra graph propagation and inter graph propagation were used to disseminate information.
In sentiment classification, a GNN is coming into use in an amount of literature. Chen et al.\textsuperscript{32} utilized graph convolution neural network and attention mechanism to complete the task of aspect-based sentiment classification. Xiao et al.\textsuperscript{33} proposed a novel method called AEGCN based on graph attention neural network to accomplish the target sentiment classification task. Zhao et al.\textsuperscript{34} used graph convolution neural network to capture the semantic dependence between multiple aspects, and used attention mechanism to encode aspects and context in the proposed model. Literature\textsuperscript{5,35} research sentiment representation learning, which utilizes GNN to model text semantic relations and train the model to obtain sentiment word vector representation containing semantic information and syntactic information.

The attention mechanism plays a crucial role in the neural network model.\textsuperscript{36} In the above literature, self-attention, multi-head attention, and other traditional attention models are used, lacking in-depth research on attention mechanisms. In this paper, a constrained attention mechanism is proposed. Based on the graph attention neural network model, orthogonal attention constraint and score attention constraint are used to improve the effectiveness and interpretability of the attention mechanism.

\section{METHOD}
This section introduces the detail of the graph convolution neural network model GCN and graph attention neural network model GAT. Then, the implicit sentiment analysis model GACNN is proposed based on the graph attention neural network.

\subsection{GCN network}
Graph convolution network called GCN is a kind of GNN, which utilizes the features of neighbor nodes to update the features of center nodes. Finally, it obtains the node representation with context semantics. Let $G(V, E)$ denote a graph, including node set $V$ and edge set $E$. We utilized $X \in \mathbb{R}^{n \times m}$ to represent the node feature, in which $n$ denotes the number of nodes and $m$ denotes the dimension of the feature vector. $A(A_{ij}) \in \mathbb{R}^{n \times n}$ denotes the adjacency matrix of graph $G$, and the diagonal element of matrix $A$ is set to 1, indicating that the nodes in graph $G$ are self-connected. $D(D_{ij}) \in \mathbb{R}^{n \times n}$ denotes the degree matrix of graph $G$, which is used to calculate the Laplace matrix. The elements of $D$ are calculated by the formula as $D_{ii} = \sum_j A_{ij}$. The calculation of a single layer graph convolution network is defined as the following formula.

$$L(1) = \rho(\hat{A}XW_0 + b)$$

In the formula, $\hat{A}$ is the Laplace matrix calculated by $\hat{A} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$. $W_0 \in \mathbb{R}^{m \times k}$ is the linear transformation weight matrix, in which $k$ is the dimension of the feature vector after graph convolution operation. The symbol $b$ is the offset vector, and $\rho$ is a nonlinear function as activation function, such as ReLU. A single GCN layer can integrate information of neighbor nodes through one convolution operation. Multiple GCN layers can integrate information of distant nodes. The calculation formula of layer $j$ is as follows.

$$L(j) = \rho(\hat{A}L^{j-1}W_{j-1} + b^j)$$

\subsection{GAT network}
Velicković et al.\textsuperscript{26} proposed graph attention neural network GAT, which uses a self-attention mechanism in the convolutional process and enhances graph the neural network’s information propagation and representation ability. Given the input of the graph as $h = \{h_1, h_2, ... , h_n\}, h \in \mathbb{R}^m$, where $n$ is the number of nodes in the graph and $m$ is the dimension of the node feature vector. The output of the GAT network is $h' = \{h'_1, h'_2, ... , h'_n\}, h' \in \mathbb{R}^{m'}$, where $m'$ is the dimension of the output node feature vector. The calculation method of $h'$ is shown as follows.

$$h'_i = \sigma \left( \sum_{j \in \mathcal{N}_i} a_{ij}W h_j \right)$$
In the formula, $\sigma$ is a nonlinear activation function, $W$ is the linear transformation weight matrix, $\mathcal{N}_i$ is the neighbor node collection of node $i$, and $a_{ij}$ is the weight of node $j$, indicating the importance of node $j$ to node $i$. The graph attention mechanism only calculates the weight of neighbor nodes of node $i$ to integrate information of neighbor nodes. The calculation method of $a_{ij}$ is shown as follows.

$$a_{ij} = \frac{\exp(\text{LeakyReLU}(a^T [Wh_i || Wh_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(a^T [Wh_i || Wh_k]))}$$  \hspace{1cm} (4)$$

In the formula, $\mathcal{N}_i$ is the neighbor node collection of node $i$, $W \in \mathbb{R}^{m' \times m}$ is the weight matrix of feature vector of nodes, the symbol $||$ denotes concatenating two vector, $a^T \in \mathbb{R}^{2nd}$ is the weight of attention mechanism, $\text{LeakyReLU}$ is nonlinear activation function.

A multi-head attention mechanism is employed in the GAT model to make the learning process more stable. The calculation method of multi-head attention is as follows, in which $K$ is the number of attention heads.

$$h'_i = \sigma \left( \frac{1}{K} \sum_{k=1}^{K} \sum_{j \in \mathcal{N}_i} a_{ij}^k W^k h_j \right)$$  \hspace{1cm} (5)$$

There is only the feature information of neighbor nodes could be integrated with the single GAT network layer. However, if the feature information of distant nodes is integrated, multiple GAT network layers should be utilized.

### 3.3 Implicit sentiment analysis model

In order to improve the sentiment classification accuracy, we proposed a novel model based on attention mechanism and graph convolutional neural network named GACNN, which utilizes GCN to construct a representation learning model of implicit sentiment sentence and extract distinct sentence feature for sentiment classification using an attention mechanism. The flow chart of the proposed model is shown in Figure 1. Firstly, we take advantage of the public dataset as the input for sentiment analysis. Preprocess the data by the word segmentation tool and statistical analysis method. Then, utilize pointwise mutual information (PMI) and term frequency and inverse document frequency (TF-IDF) to calculate the weight of words, indicating word-word relationship and word-sentence relationship. Next, construct the semantic graph using PMI and TF-IDF values, composed of sentence nodes and word nodes. Finally, build the architecture of the GACNN model and train the model for sentiment category prediction. Take the semantic graph as the input of the GACNN network, and execute forward computation to acquire the semantic representation of the sentence node. Calculate the probability distribution of the sentence node on each sentiment category by soft-max function, and the sentence sentiment classification label is obtained. Attention mechanism and attention constraint are adopted in the proposed model to guarantee the accuracy of the weight. This part introduces the detailed structure of the proposed model and the principle of implicit sentiment classification.

#### 3.3.1 Input data

Input data are the set of implicit sentiment sentences, which was represented as $S(s_1, s_2, \ldots, s_n)$. The feature vector of the sentence was obtained by the segmentation algorithm. The symbol of $s_i \in \mathbb{R}^d$ represents the feature vector of the sentiment sentence.
3.3.2 Semantic graph

The semantic graph is composed of sentence nodes and word nodes, represented as $G(V, E)$, in which $V \subseteq \{S, W\}$ is the node-set of the semantic graph, $S$ is the node-set of sentiment sentences, $W$ is the node-set of characteristic words, $E$ is the edge set connecting word nodes and linking word node and sentence node. According to the value of PMI, the edge weight between words is assigned. According to the value of TF-IDF, the edge weight between word and sentence is determined. Semantic graph architecture is showed in Figure 2, in which $s_i$ is the sentence node, $w_j$ is the word node, the connection is established between words and sentences. The formula of edge weight $e_{ij}$ between node $v_i$ and node $v_j$ is shown as follows.

$$e_{ij} = \begin{cases} 
\text{PMI}(i,j) & \text{node } i \text{ and node } j \text{ are word nodes, } \text{PMI}(i,j) > 0; \\
\text{TF} - \text{IDF}(i,j) & \text{node } i \text{ is sentence, and node } j \text{ is word;}
\end{cases}$$

$$1 \quad i = j;$$

$$0 \quad \text{others}. \quad (6)$$

The PMI value is used to measure the semantic relationship between two words. When the value of $\text{PMI}(i,j)$ is greater than zero, it indicates a close semantic relationship between words $i$ and $j$. On the other hand, when the value of $\text{PMI}(i,j)$ is less than zero, the semantic relationship between words $i$ and $j$ is not close, or there is no semantic relationship. Therefore, the connection is established between two words only when the PMI value is greater than zero. The formula of PMI is as follows, in which $p(i,j)$ is the probability that the words $i$ and $j$ appear simultaneously in the statistical window, $p(i)$ and $p(j)$ are the probability of the words $i$ and $j$ appearing in the statistical window, respectively.

$$\text{PMI}(i,j) = \log \frac{p(i,j)}{p(i)p(j)} \quad (7)$$

Measure the importance of words to the sentence using The TF-IDF value. The importance of words increases in proportion to the frequency of words appearing in the sentence and decreases inversely with the frequency of words appearing in the corpus. The formula of TF-IDF is shown as follows, where $n_{ij}$ indicates the frequency of word $i$ appearing in sentence $j$, $|S|$ denotes the number of sentences in corpus, $\{j : t_i \in d_j\}$ represents the number of sentences containing the word $i$.

$$\text{TF-IDF}(i,j) = \frac{n_{ij}}{\sum_k n_{kj}} \cdot \log \frac{|S|}{|\{j : t_i \in d_j\}| + 1} \quad (8)$$

3.3.3 Architecture of GACNN model

The architecture of the GACNN model including three parts is shown as Figure 3:
1. The constructed semantic graph $G$ is input into graph convolutional neural network GCN. A convolution operation captures the semantic relations between words and sentences.

2. The attention mechanism is adopted to calculate the importance weight of words to sentence expression and update the feature vector of the sentence node.

3. The full connection layer and soft-max layer are used to calculate the probability distribution of sentiment categories.

In the part of attention mechanism, employ orthogonal attention constraint and score attention constraint to improve the performance of the proposed model. The orthogonal attention constraint ensures the diversity of the information stored by the multi-head attention mechanism. The score attention constraint ensures that the attention mechanism assigns heavy attention to a certain number of words.

The feature vectors of word nodes $w_i$ and sentence nodes $s_i$ are saved in semantic graph $G$, denoting as $X(x_i), x_i \in \mathbb{R}^m$, in which $m$ is the dimension of the node feature vector. $A(a_{ij})$ is the adjacency matrix of semantic graph $G$, where $a_{ij}$ denotes the edge weight between node $i$ and node $j$.

Graph convolution neural network can propagate semantic information between nodes in semantic graph $G$. The feature vector of nodes contains semantic information and relationship information between nodes after the graph convolution process. In our proposed model, we utilize a graph convolution neural network to update the feature vector of nodes. The formula of graph convolution process is shown as follows, in which $\tilde{A}$ is the regularized Laplacian matrix of adjacency matrix. The element of adjacency matrix is the weight $e_{ij}$ between nodes. $x$ is the input feature vector of all nodes. $W' \in \mathbb{R}^{m \times o}$ is the weight of convolution processing, in which $o$ is the dimension of node feature vector after convolution process.

$$x' = \rho(\tilde{A}xW') \quad (9)$$

After the graph convolution process, we adopt an attention mechanism to calculate the importance of neighbor nodes to central nodes. The importance indicates the contribution of words to the sentiment expression of a sentence. Attention mechanism calculation of node $v_i$ is shown in Figure 4, in which $v_i$ is the central node, $v_j(j = 1, 2, \ldots)$ is the neighbor node, $a_{ij}(j = 1, 2, \ldots)$ is the weight between nodes. In other words, $a_{ij}$ indicates the importance of a neighbor node to a central node.
The attention mechanism is a single-layer feedforward neural network. The formula of attention is shown as follows: $x_i$ and $x_j$ are the feature vector of node $i$ and node $j$. $\mathcal{N}_i$ is all neighbor nodes of the node $i$. $Q \in \mathbb{R}^{m \times t}$ is the query weight, $K \in \mathbb{R}^{m \times t}$ is the keyword weight. The symbol $a^T \in \mathbb{R}^{2m}$ is the weight vector of attention mechanism. $\cdot^T$ represents transposition. $\| \|$ is the concatenation operation. Adopt ReLU as a nonlinear activation function in the formula.

$$
\alpha_{ij} = \text{softmax}_j(a^T[x_iQ \| x_jK]) = \frac{\exp(\text{ReLU}(a^T[x_iQ \| x_jK]))}{\sum_{k \in \mathcal{N}_i}(\text{ReLU}(a^T[x_iQ \| x_kK]))}
$$

(10)

Once obtained normalized attention coefficients, it is used to update the feature vector of nodes in semantic graph $G$. The formula of update for nodes is shown as follows, in which $\sigma$ is the nonlinear activation function. $W'' \in \mathbb{R}^{l \times m}$ is the weight of the attention mechanism for updating feature vector, and the symbol $l$ is the dimension of the output feature vector. According to the theory of attention mechanism, the greater the attention weight of neighbor nodes, the more contribution they make to the feature vector of central nodes.

$$
x''_i = \sigma \left( \sum_{j \in \mathcal{N}_i} \alpha_{ij}x'_i W'' \right)
$$

(11)

In the process of attention mechanism calculation, multi-head attention can be used to calculate the weight. The formula of updating the feature vector of nodes is shown as follows, in which $K$ is the number of attention heads.

$$
x''_i = \sigma \left( \frac{1}{K} \sum_{k=1}^{K} \sum_{j \in \mathcal{N}_i} \alpha_{ik}x'_i W^k \right)
$$

(12)

In order to improve the rationality of attention weight and the effect of attention mechanism, orthogonal attention constraint and score attention constraint are adopted in the attention mechanism. The detail of attention constraint will be introduced below.

After updating the feature vector of nodes, the probability distribution of sentiment categories is calculated by the feature vector of sentence nodes. The full connection layer and soft-max function are utilized to calculate the probability of the sentiment category. The formula of full connection is shown as follows, in which $v_i \in S$ indicates that input data are the feature vector of sentence nodes. $W_1 \in \mathbb{R}^{l \times c}$ is the weight of the full connection layer, and the symbol $c$ is the number of sentiment categories.

$$
y = \sigma(x''W_1)_{h_i \in D}
$$

(13)

Soft-max is the normalization function for calculating the probability distribution of sentiment categories. The formula of soft-max is shown as follows, in which $z_i$ indicates the probability that a sentence belongs to the sentiment category $i$.

$$
z_i = \frac{\exp(y_i)}{\sum_{k=1}^{c} \exp(y_k)}
$$

(14)
The cross-entropy loss function is used as the loss function of the proposed model. The formula of the loss function is shown as follows, in which $S$ is the dataset of sentences. The symbol $t_{sk}$ is the correct tag of a sentence. The symbol $z_{sk}$ is the predicted tag of a sentence.

$$L_1 = - \sum_{s \in S} \sum_{k=1}^c t_{sk} \log z_{sk}$$  \hspace{1cm} (15)

Attention constraint is implemented by the loss function, where $L_2$ and $L_3$ are produced from orthogonal attention constraint and score attention constraint, respectively. As a result, the loss function of the model can be described as follows. Then, the stochastic gradient descent algorithm is used to optimize the model parameters so that the loss function gradually decreases and finally obtain the optimal implicit sentiment analysis model.

$$L = L_1 + L_2 + L_3$$ \hspace{1cm} (16)

### 3.3.4 Orthogonal attention constraint

The weights calculated by the multi-head attention mechanism cover each other, which will reduce the ability of model to extract and represent sentiment information. Therefore, it is necessary to ensure the difference between attention heads. As mentioned above, attention weight $[a_1, a_2, \ldots, a_n]$ has been obtained by multi-head attention, in which $a_i$ is the weight of the single attention head. In order to keep the difference between attention weight vectors and reduce overlapping information, orthogonal attention constraint is adopted in the attention mechanism. The formula of orthogonal attention constraint is shown as follows. The diversity of attention heads is made by minimizing $L_2$.

$$L_2 = \sum \frac{a_i \cdot a_j}{|a_i| \cdot |a_j|} \hspace{1cm} (i,j \in [1, 2, 3 \ldots n], i \neq j)$$  \hspace{1cm} (17)

### 3.3.5 Score attention constraint

In general, there are a certain number of essential words in one sentence that play a crucial role in sentiment expression, and the attention mechanism should focus on some important words. Attention can not only focus on one word. Otherwise, it will also lose much semantic function. Attention also cannot disperse on all words. Otherwise, it cannot play the role of paying attention to some critical words. Therefore, score attention constraint is proposed for focusing on essential words, guaranteeing the rationality of single attention weight.

The neighbor nodes of a sentence node are a set of words that make up the sentence. The attention weights of neighbor nodes are presented as $a_i[a_{11}, a_{12}, \ldots, a_{im}]$, in which $m$ is the number of neighbor nodes of sentence $i$. The variance of the weight vector is calculated firstly in score attention constraint. The value of variance is constrained by unimodal function $f(\rho) = \rho + 1/\rho$. When the value of $\rho$ is 1, the minimum value of function is 2, so that we can realize the purpose of weight concentration on some important words. The formula of score attention constraint is shown as follows, in which $\rho_i$ is the variance of weight vector of sentence $i$, $b$ is the hyper parameters of score constraint, $N$ is the number of sentences. Control single attention weight focus on some important words by minimizing $L_3$.

$$L_3 = \frac{1}{N} \sum_i \rho_i + b/\rho_i$$  \hspace{1cm} (18)

### 4 EXPERIMENT

In this section, we experiment on benchmark datasets to verify the performance of the proposed model. And analyze the effect of the attention constraint mechanism for sentiment classification.
4.1 Datasets and evaluation metrics

SMP-ECISA2019 dataset is chosen as the experiment dataset to evaluate the performance of the proposed model. In 2019, the Eighth National Conference on Social Media Processing (SMP) organized the Chinese implicit sentiment assessment SMP-ECISA and released the Chinese implicit sentiment analysis dataset for evaluation. Shanxi University provides the dataset. The data source mainly includes microblog, tourism website, and product forum. The data’s main fields are Spring Festival Gala, haze, Le Television, national examination, tourism, and Dragon Boat Festival. The text that contains explicit sentiment words has been filtered out by using a large-scale sentiment dictionary. The data were labeled as positive implicit sentiment (tag 1), negative implicit sentiment (tag 2), and sentences without sentiment tendency (tag 3). The training dataset, validation dataset, and test dataset are shown in Table 2. The training set and verification set have public tags. Here, the labeled data are used to train and test according to the ratio of 8:2. The total number of labeled sentences is 19,917.

The precision value \( P \), recall rate \( R \), and \( F \) value were adopted as evaluation metrics. The formula of evaluation metrics is shown as follows, in which TP is the number of positive classes predicted as positive classes, FP is the number of negative classes predicted as positive classes, FN is the number of positive classes predicted as negative classes. Finally, the value of \( F \) is a harmonic value considering accuracy and recall rate, which reflects the overall performance of the model.

\[
P = \frac{TP}{TP + FP} \tag{19}
\]

\[
R = \frac{TP}{TP + FN} \tag{20}
\]

\[
F = \frac{2 \times P \times R}{P + R} \tag{21}
\]

4.2 Implementation details

The hyper parameters of the proposed model are shown in Table 3. In order to make the model get optimal solution more convenient, utilize the normal distribution to initializing the feature vector of semantic graph nodes. The dimension of the feature vector in the input layer is 100. The dimension of the output feature vector is 100 after the graph convolution process. The dimension of the feature vector is 100 after the calculation of the attention mechanism. The proposed model adopts three heads in attention mechanism, in which the weight vector is represented as \([a_1, a_2, a_3]\). We adopt the stochastic gradient descent optimizer, where the learning rate is 0.05, the maximum number of training iterations is 200.

| Dataset | Chapters | Positive sentences | Negative sentences | Neutral sentences | Total sentences |
|---------|----------|--------------------|-------------------|------------------|----------------|
| Train   | 12,664   | 3828               | 3957              | 6989             | 14,774         |
| Validation | 4391   | 1232               | 1358              | 2553             | 5143           |
| Test    | 6380     | 919                | 979               | 1902             | 3800           |

| Parameter                              | Value          |
|----------------------------------------|----------------|
| Initialization method of the feature vector | Normal distribution |
| \( m \) (Dimension of the feature vector in input layer) | 100 |
| \( o \) (Dimension of the output feature vector in graph convolution layer) | 100 |
| \( l \) (Dimension of the output feature vector in attention layer) | 100 |
| Weight vector of the attention layer. | \([a_1, a_2, a_3]\) |
### 4.3 Results and analysis

#### 4.3.1 Overall performance

The experimental results of the model are shown in Table 4. In the table, orthogonal constraint indicates adding orthogonal constraints into the attention model. Score constraint indicates adding score constraint to the attention model. The experimental results show that the $F$ value of the basic model GACNN is 86.98%. The model’s accuracy is improved by adding attention orthogonal constraint and score constraint, respectively, and the $F$ value reached more than 87%, which indicates that both attention constraints affect sentiment classification. While both attention constraints are integrated into the model simultaneously, the $F$ value reaches 88.16%.

#### 4.3.2 Comparative analysis

To comprehensively measure the performance of our proposed model, we compare the GACNN model with several benchmark models in the literature, including Naïve Bayes, CNN, HAN, semantic dependency tree-based CNN (SDT-CNN) model proposed in Reference 17, BiLSTM with attention model proposed in Reference 2, context-specific heterogeneous graph convolutional network (CsHGCN) model proposed in Reference 3 and convolution layer attention model proposed in Reference 37.

**Naïve Bayes:** Naïve Bayes is a fundamental machine learning method for classification. We select appropriate words with high document frequency value as a feature and then use feature words as the input of naïve Bayes to classify sentiment polarity.

**CNN:** Convolutional neural network is used to learn the representation of sentences by extracting semantic features. Word embedding vectors are input CNN layer, and vector representation of sentence is obtained by multi-layer calculation. The classifier is utilized for classification in the last layer, such as the SoftMax classifier.

**SDT-CNN**\(^{17}\): SDT-CNN is a multi-level semantic fused method to learn features based on representation learning, which consists of three different features: sentiment target representation at the word level, structure embedded representation at the sentence level, and context semantic background representation at the document level.

**BiLSTM**\(^{2}\): Wei Ji et al. proposed a BiLSTM model with multi-polarity orthogonal attention for implicit sentiment classification, which identifies the difference between the words and the sentiment orientation regarded as a significant feature. The orthogonal restriction mechanism is adopted in this model to ensure discriminatory performance.

**HAN**\(^{38}\): Hierarchical attention network uses a hierarchical attention mechanism at the word level and the sentence level, which gives different attention to words and sentences based on their importance in the text.

**CsHGCN**\(^{3}\): Context-specific heterogeneous GCN can combine all context representations, which considers the whole context at document level as a heterogeneous graph and captures deep level domain features by graph convolution.

**HNN**\(^{37}\): Hybrid neural network model integrates convolutional neural network and BiLSTM to extract more meaningful semantic features, which uses a convolution network to extract from the text, uses the BiLSTM network to extract context information, and adopts attention mechanism in the framework.

**BERT**\(^{39}\): BERT is an outstanding language representation model, and achieves great performance in a wide range of task. We fine-tuned pre-trained BERT model on sentiment dataset and predicted the sentiment polarity.
TABLE 5  The result of comparative experiment

| Model        | P (%) | R (%) | F (%) |
|--------------|-------|-------|-------|
| Naïve Bayes  | 67.70 | 63.40 | 65.40 |
| CNN          | 73.40 | 73.40 | 73.40 |
| SDT-CNN17    | 73.4  | 75.8  | 74.5  |
| BiLSTM2      | -     | -     | 77.50 |
| HAN38        | 76.67 | 73.69 | 74.72 |
| CsHGCN3      | 79.46 | 76.00 | 76.94 |
| HNN37        | 77.00 | 84.00 | 80.00 |
| BERT39       | 78.20 | 79.21 | 78.70 |
| GACNN        | 88.31 | 88.02 | 88.16 |

**Example 1:**
After reading it, I am full of memories and many elements of that era.

**Example 2:**
I can't stand the haze. I can't see my fingers. I don't want to go out these two days!
I can't stand the haze. I can't see my fingers. I don't want to go out these two days!
I can't stand the haze. I can't see my fingers. I don't want to go out these two days!

**Example 3:**
I went to the museum when it opened, recording the history of Chengdu and even Sichuan.
I went to the museum when it opened, recording the history of Chengdu and even Sichuan.
I went to the museum when it opened, recording the history of Chengdu and even Sichuan.

**FIGURE 5**  The visualization of attention weight for sentiment representation

The results of the comparative experiment are shown in Table 5. The evaluation metrics values of Naïve Bayes and CNN are obtained from Reference 2. The horizontal lines in the table indicate that there is no such index in the literature. The indicators in Reference 3 adopt the average value of three sentiment categories. It is obvious in experimental results that the F value of the model proposed in this paper is 88.16%, which is much higher than that of other benchmark models and 8.16% higher than that of the benchmark model in literature named HNN15.

4.3.3  Effects of the attention mechanism

The attention mechanism is used to extract sentiment features in the proposed model GACNN. The weight of words indicates the importance of sentiment representation. We visualized the heat map of weight with three attention heads. The visualization of three examples is shown in Figure 5. The color depth indicates the importance of words for sentiment representation.

Multi-head attention plays a vital role in neural network models, so we analyze the performance of multi-head attention with the different number of attention heads in the model. The result of attention experiments shows in Figure 6. When the number of heads is equal to 3, the evaluation metrics value reaches maximum. It is not hard to find that the value of F reached a minimum when no attention mechanism was used in the model. However, the value of metrics begins to drop when the number of heads is greater than 4. The result shows that the attention mechanism plays a positive role in our model with an appropriate head number.
4.3.4 Effects of the score attention constraint

In the proposed model, we use a unimodal function $f(\rho) = \rho + b/\rho$ to limit the weight of attention to the finite number of essential words in a sentence. We take the experiment on parameter $b$ to examine the effect of the score attention constraint. The result of the experiment shows in Figure 7. The performance of the proposed model changes with different values of parameter $b$. When the value of parameter $b$ equals 4, the evaluation metrics reach maximum, and the value drops slightly with other values.

5 CONCLUSION AND FUTURE WORK

This paper investigated the implicit sentiment analysis task based on the graph attention neural network and proposed a novel model called GACNN. The model’s motivation, principle, and structure are described, and the effectiveness and advanced nature of the model is verified on the implicit sentiment analysis evaluation dataset. The character of implicit sentiment analysis is that it does not contain explicit emotional words, making it more indistinct to express sentiment and more challenging to extract sentiment features. In this paper, we employ a GACNN to extract sentiment features. Orthogonal attention constraint and score attention constraint are utilized to differentiate multiple attention and pay more attention to certain essential words. The effect of graph attention neural network model and attention mechanism on sentiment analysis is verified by experimental analysis. This study has not considered the influence of external knowledge on implicit sentiment analysis. In the following research, we will explore the application of semantic information such as part of speech, dependency relationship, and other information in the sentiment analysis model based on graph neural network.
ACKNOWLEDGMENTS

The authors would like to thank all anonymous reviewers for their valuable comments and suggestions, which have significantly improved the quality and presentation of this paper. This work was supported by the National Key R&D Program of China (2018YFB1403302); the National Social Science Foundation of China (20CTQ012); the Scientific Research Fund of Shandong University of Technology (419038).

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AUTHOR CONTRIBUTIONS

Shanliang Yang: Conceptualization, methodology, software, formal analysis, investigation, writing—original draft, writing—review and editing, visualization. Linlin Xing: Conceptualization, formal analysis, investigation, supervision, project administration, funding acquisition. Yongming Li: Investigation, supervision, project administration, funding acquisition. Zheng Chang: Investigation, supervision, project administration, funding acquisition.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://www.biendata.xyz/competition/smpecisa2019.

ORCID

Shanliang Yang https://orcid.org/0000-0002-3764-4622

REFERENCES

1. Liu B. Sentiment analysis and opinion mining. Synth Lect Hum Lang Technol. 2012;5(1):1-167.
2. Wei J, Liao J, Yang Z, Wang S, Zhao Q. BiLSTM with multi-polarity orthogonal attention for implicit sentiment analysis. Neurocomputing. 2020;383:165-173.
3. Zuo E, Zhao H, Chen B, Chen Q. Context-specific heterogeneous graph convolutional network for implicit sentiment analysis. IEEE Access. 2020;8:37967-37975.
4. Zhang T, Huang M, Zhao L. Learning structured representation for text classification via reinforcement learning. Association for the Advancement of Artificial Intelligence. Palo Alto, CA: AAAI Press; 2018:6053-6066.
5. Kayvan B, Hadi Z, Emad K, Hadi V. Leveraging deep graph-based text representation for sentiment polarity applications. Expert Syst Appl. 2020;144:1-10.
6. Peng B, Wang J, Zhang X. Adversarial learning of sentiment word representations for sentiment analysis. Inf Sci. 2020;541:426-441.
7. Chen F, Huang Y. Knowledge-enhanced neural networks for sentiment analysis of Chinese reviews. Neurocomputing. 2019;368:51-58.
8. Vo AD, Nguyen QP, Ock CY. Semantic and syntactic analysis in learning representation based on a sentiment analysis model. Appl Intell. 2020;50:663-680.
9. Li W, Zhu L, Shi Y, Guo K, Erik C. User reviews: sentiment analysis using lexicon integrated two-channel CNN–LSTM family models. Appl Soft Comput J. 2020;94:106411-106435.
10. Abdi A, Shamsuddin SM, Hasan S, Piran J. Deep learning-based sentiment classification of evaluative text based on multi-feature fusion. Inf Process Manag. 2019;56(4):1245-1259.
11. Han K, Chien W, Chui C, Cheng Y. Application of support vector machine (SVM) in the sentiment analysis of twitter dataset. Appl Sci. 2020;10:1116-1125.
12. Chen T, Xu R, He Y, Wang X. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Syst Appl. 2017;72:221-230.
13. Akhtar MS, Ekbal A, Cambria E. How intense are you? predicting intensities of emotions and sentiments using stacked ensemble. IEEE Comput Intell Mag. 2020;15(1):64-75.
14. Cambria E, Li Y, Xing FZ, Poria S, Kwok K. SenticNet 6: ensemble application of symbolic and subsymbolic AI for sentiment analysis. CIKM ’20: the 29th ACM International Conference on Information and Knowledge Management. Virtual Event, Ireland: ACM; 2020:105-114.
15. Fang Z, Zhang Q, Tang X, Wang A, Baron C. An implicit opinion analysis model based on feature-based implicit opinion patterns. Artif Intell Rev. 2020;53:4547-4574.
16. Cho Y, Wiebe J, Mihalcea R. Coarse-grained +/-effect word sense disambiguation for implicit sentiment analysis. IEEE Trans Affect Comput. 2017;8(4):471-479.
17. Liao J, Wang S, Li D. Identification of fact-implied implicit sentiment based on multi-level semantic fused representation. Knowl-Based Syst. 2019;165:197-207.
18. Wang H, Hou M, Li F, Zhang Y. Chinese implicit sentiment analysis based on hierarchical knowledge enhancement and multi-pooling. *IEEE Access*. 2020;8:126051-126065.

19. Liu MZ, Zhou FY, Chen K, Zhao Y. Co-attention networks based on aspect and context for aspect-level sentiment analysis. *Knowl-Based Syst*. 2021;217(2):106810.

20. Li W, Li D, Yin H, Zhang L, Zhu Z, Liu P. Lexicon-enhanced attention network based on text representation for sentiment classification. *Appl Sci*. 2019;3717:9-21.

21. Meb A, Sn A, Ma B, Ec C, Ura D. Acdm: an attention-based bidirectional cnn-rnn deep model for sentiment analysis. *Futur Gener Comput Syst*. 2021;115:279-294.

22. Su J, Tang J, Jiang H, et al. Enhanced aspect-based sentiment analysis models with progressive self-supervised attention learning. *Artif Intell*. 2021;296:103477.

23. Gan C, Wang L, Zhang Z, Wang Z. Sparse attention based separable dilated convolutional neural network for targeted sentiment analysis. *Knowl-Based Syst*. 2020;188:104827.

24. Zhou J, Cui H, Zhang Z, et al. Graph neural networks: a review of methods and applications. *AI Open*. 2020;1:57-81.

25. Zhang S, Tong H, Xu J, Maciejewski R. Graph convolutional networks: a comprehensive review. *Comput Soc Netw*. 2019;6(11):1-23.

26. Velickovi P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. *Proceedings of the 6th International Conference on Learning representations ICLR 2018*; Vancouver Canada: OpenReview.net; 2018:1-12.

27. Zhang J, Shi X., Xie J., Ma H., King L., Yeung D., GaAN: gated attention networks for learning on large and spatiotemporal graphs; 2018. https://arxiv.org/abs/1803.07294.

28. Zhang C, Song D, Huang C. Heterogeneous graph neural network. *Knowledge Discovery and Data Mining*. Anchorage, AK: ACM; 2019:793-803.

29. Wang X, Ji H, Shi C, et al. Heterogeneous graph attention network. *WWW Conf*. 2019;2019:2022-2032.

30. Yao L, Mao C, Luo Y. Graph convolutional networks for text classification. Proceedings of the 33rd AAAI Conference on Artificial Intelligence, AAAI 2019; 2019:7370-7377.

31. Liu X, You X, Zhang X, Wu J, Lv P. Tensor graph convolutional networks for text classification. *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. Vol 2020. Palo Alto, CA: AAAI Press; 2020:8409-8416.

32. Chen J, Hou H, Ji Y, Gao J. Graph convolutional networks with structural attention model for aspect based sentiment analysis. *International Joint Conference on Neural Networks*. Budapest, Hungary: IEEE; 2019:1-7.

33. Xiao L, Hu X, Chen Y, et al. Targeted sentiment classification based on attentional encoding and graph convolutional networks. *Appl Sci*. 2020;10(957):1-17.

34. Zhao P, Hou L, Wu Q. Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification. *Knowl-Based Syst*. 2020;193:105443.

35. Vashishth S., Bhandari M., Yadav P., Rai P., Bhattacharyya P., Talukdar P., Incorporating syntactic and semantic information in word Embeddings using graph convolutional networks. arxiv.org/abs/1809.04283, 2018: 1–11.

36. Knyazev B, Graham WT & Mohamed RA Understanding attention and generalization in graph neural networks. Proceedings of the 33rd Conference on Neural Information Processing Systems, NeurIPS 2019 Vancouver, BC, Canada; 2019:32.

37. Zhao R, Xiong X, Ju S, Li Z, Xie C. Implicit sentiment analysis for Chinese texts based on a hybrid neural network. *JSichuan Univ Nat Sci Ed*. 2020;57:264-270.

38. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference; 2016:1480-1489.

39. Jacob D, Ming-Wei C, Kenton L, Kristina T. BERT: pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Stroudsburg, PA: Association for Computational Linguistics (ACL); 2019:4171-4186.

---

**How to cite this article:** Yang S, Xing L, Li Y, Chang Z. Implicit sentiment analysis based on graph attention neural network. *Engineering Reports*. 2022;4(1):e12452. doi: 10.1002/eng2.12452