

Article

ConAs-GRNs: Sentiment Classification with Construction-Assisted Multi-Scale Graph Reasoning Networks

Bo Chen, Weiming Peng and Jihua Song *

School of Artificial Intelligence, Beijing Normal University, No. 19 Xijiekouwai St., Haidian District, Beijing 100875, China; bochen@mail.bnu.edu.cn (B.C.); pengweiming@bnu.edu.cn (W.P.)

* Correspondence: songjh@bnu.edu.cn

Abstract: Traditional neural networks have limited capabilities in modeling the refined global and contextual semantics of emotional texts and usually ignore the dependencies between different emotional words. To address this limitation, this paper proposes a construction-assisted multi-scale graph reasoning network (ConAs-GRNs), which explores the details of the contextual semantics as well as the emotional dependencies between emotional texts from multiple aspects by focusing on the salient emotional information. In this network, an emotional construction-based multi-scale topological graph is used to describe multiple aspects of emotional dependency, and a sentence dependency tree is utilized to construct a relationship graph based on emotional words and texts. Then, the transfer learning and pooling learning on the topology map is performed. In our case, a weighted edge reduction strategy is used to aggregate the adjacency information which enables the internal transfer of semantic information in a single graph. Moreover, to implement the inter-graph transfer of semantic information, we rely on the construction structure to coordinate the heterogeneous graph information. The extensive experiments conducted on two baseline datasets, SemEval 2014 and ACL-14, demonstrate that the proposed ConAs-GRNs can effectively coordinate and integrate the heterogeneous information from within constructions.

Keywords: sentiment classification; multi-scale topological map; heterogeneous information; construction structure; sentence dependency tree



Citation: Chen, B.; Peng, W.; Song, J. ConAs-GRNs: Sentiment Classification with Construction-Assisted Multi-Scale Graph Reasoning Networks. *Electronics* **2022**, *11*, 1825. <https://doi.org/10.3390/electronics11121825>

Academic Editors: Andrea Prati, Luis Javier García Villalba and Vincent A. Cicirello

Received: 13 April 2022

Accepted: 6 June 2022

Published: 8 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sentiment classification [1,2] aims to interpret the explicit sentiment polarity of a given sentence in a complex context, which is a fundamental natural language processing task that has received much attention in recent years. For instance, in a sentence such as “These MacBooks are encased in a soft rubber enclosure—so you will never know about the razor edge until you buy it”, the sentiment polarities of “rubber enclosure” and “edge” are positive and negative, respectively.

Many existing sentiment classification systems [3–8] focus on statistical methods to develop a set of handcrafted features for sentiment classification. However, these handcrafted feature-based methods usually require human involvement. In such a case, the accuracy is questionable. Moreover, it is difficult for these methods to meet the growing application demands. In recent years, deep learning methods [9,10] have received increasing attention because such methods are able to automatically learn emotional features and generate useful low-dimensional representations from context; they can also achieve high accuracy in the sentiment classification tasks without requiring complex feature engineering. For example, some studies use the attention mechanism [11] to extract important sentiment words in sentences in order to improve classification accuracy. Other studies use the LSTM model [12] to establish long-term dependencies. Recently, graph convolutional networks (GCNs) [13] have been widely used in sentiment classification tasks, in which an information transfer mechanism is used to extract the node neighborhood information and the

local context information. Chen et al. [14] proposed a multiple attention-based LSTM to capture the relevance between sentiment words and their contexts. Lin et al. [15] developed a sentiment semantic coding network based on a multi-head self-attention mechanism, which can model the contextual semantics of sentiment words.

Although the research methods above have made great progress in sentiment classification, most of the existing methods only model the semantics of sentiment words, ignoring the dependencies between emotional words and the semantic relationship between sentiment words. For instance, as shown in Figure 1, we can see that “battery” has a negative sentiment polarity in the first position. We then guess that the second “battery” may be negative, while the emotional polarity of “battery” after the conjunction “but” and the qualifier “upgrade” may also be positive. Overall, the same sentiment text has multiple constructions that can help us to judge the sentiment polarity of sentiment words.

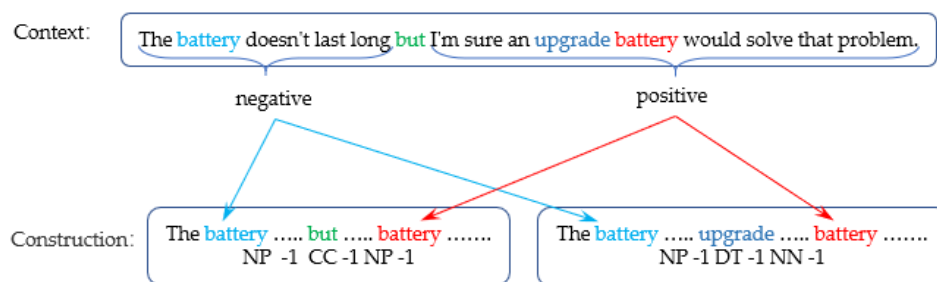


Figure 1. The construction of sentiment context. “Context” represents the emotional text (or sentence), and “Construction” represents the construction of a sentence, e.g., “The battery won’t last long but I’m sure this will be fixed by upgrading the battery.” This sentence might be constructed as “NP -1 CC -1 NP -1” and “NP -1 DT -1 NN -1”, where NP is a noun phrase, CC is a conjunction, and DT is a determiner. The blue and red arrows represent negative and positive emotions, respectively.

In this paper, we develop a construction-aided multi-scale graph reasoning network (ConAs-GRNs) to capture the contextual and global semantic information of sentiment texts and to model their sentiment dependencies. More specifically, the graph reasoning networks can obtain interdependent information from rich relational data and enhance interdependent interactivity. For each node in the topological and relational graph, the graph inference network encodes its neighborhood information into a low-dimensional feature vector. As shown in Figure 1, “battery” is regarded as a node, which can be used to construct a corresponding dependency graph with the adjacent sentiment words. After that, on the basis of the structure, a large-scale topological graph containing all emotional sentences can be constructed by exploiting the relationship between the emotional words and texts. Our model learns the emotional dependencies of these heterogeneous graphs at different scales and aggregates information between heterogeneous nodes to obtain global and contextual semantics. To capture the emotion-specific representations and the important salient information, before using the graph inference network, we adopt a triple attention mechanism with positional encoding, i.e., we focus on key features in three directions: horizontal, vertical, and depth. It is worth noting that in the graph inference network, we adopt a weighted edge reduction strategy to reduce the number of nodes in order to ease the computational complexity. The main contributions of this paper are summarized as follows:

- To the best of our knowledge, this work is the first attempt at building a heterogeneous relation graph based on sentiment construction for sentiment words and texts. Moreover, we utilize Graph Reasoning Networks (GRNs) to capture the details of the contextual and global semantics of sentences.
- We design an emotional dependency graph and a constructional relationship graph, where the constructive details are used to assist the learning model for sentiment classification.

- In the process of constructing heterogeneous graphs, a weighted edge reduction strategy is used to refine the node information and reduce the computational complexity of the model. The extensive experiments conducted on two baseline datasets, SemEval 2014 and ACL-14, demonstrate that our method outperforms other sentiment classification methods.

The rest of this article is organized as follows. We review the related work in Section 2 and present the sentiment classification framework ConAs-GRNs in Section 3. The ablation research and experimental results are provided in Section 4. Finally, Section 5 presents the conclusion of our paper.

2. Related Work

Sentiment classification has received extensive attention in recent years [10,15]. In order to improve the accuracy and efficiency of emotion classification, researchers have developed many classification methods based on traditional handcrafted features [7] and deep learning [9]. This section will introduce the related research in detail.

Many existing sentiment classification methods extract the handcrafted features by feature engineering. For instance, Pathik N et al. [16] discussed the relationship between the Latent Dirichlet Allocation (LDA) and probabilistic modeling. Although LDA can model different topics, it relies too much on statistical models and does not exploit high-dimensional features. Machine learning methods increasingly use visualized text and high-dimensional data as input, which provide more dimensional statistical features. Thakur et al. [6] developed a Kernel Optimized-Support Vector Machine (KO-SVM) model for sentiment classification. In their scheme, the sentiment features were fed into the classifier. More specifically, they also improved the SVM classifier by replacing the exponential kernel with an optimized kernel and used a self-adaptive lion algorithm to improve the optimizer in order to make it more sensitive to dominant features. Nafis et al. [7] proposed an improved hybrid feature selection method using Term Frequency-Inverse Document Frequency (TF-IDF) and Support Vector Machine (SVM-RFE) for sentiment classification. Fauzi et al. [8] used the random forests for Indonesian sentiment classification and explored some variations of the term weights in the classification results. However, this method did not significantly improve the performance of the random forest.

As the network layer of deep learning deepens, features that are more conducive to sentiment analysis can be extracted. The recurrent neural network (RNN) [17] has been proved to be suitable for language sentiment classification tasks in many methods. Zhang et al. [10] proposed a Gated Neural Network, which used a gating mechanism to control the importance of the context to the target. In this case, only one target with the highest probability is concerned in a sentence. However, the context relationship of the sentence is difficult to determine according to the target. To address this limitation, Ma et al. [12] leveraged the object-level and sentence-level attention with commonsense knowledge to enhance the long short-term memory (LSTM) network, thus improving the network's accuracy.

Due to the control effect of the attention mechanism on the target, many studies combine attention mechanism and RNN. Huddar et al. [17] proposed a pair-wise attention mechanism to analyze and understand multimodal context and semantics. Chen et al. analyzed a corpus of 3200 English tweets using an attention-based LSTM. Zeng et al. [18] introduced position-aware vectors to solve the common problem of the attention mechanism being unable to focus on the contextual position. However, the RNN structures still have many limitations. The introduction of graph convolutional networks (GCNs) may give a promising solution for a better sentiment classification. Zhou et al. [13] combined grammar and knowledge to improve the graph convolution, which could improve the performance of emotion classification. Zhu et al. [19] used the local and global structure dependency to guide the graph convolutional network and then adaptively fused information with the use of a gate mechanism. Zhang et al. [20] extracted the sentence structure of the dependency tree and solved the long-term multi-word dependency problem, thus enabling the graph

convolutional network to achieve better results in aspect sentiment classification. However, this method does not take full advantage of the label information of edges, and there is still room for further improvement.

3. ConAs-GRNs Frameworks

In this section, we elaborate on the proposed ConAs-GRNs framework for sentiment classification. The word feature vectors of emotional sentences are obtained by the BERT method, which are then fed into the fully connected layer to extract the emotional words and related construction information to obtain rich underlying semantic features, i.e., prior knowledge. Secondly, a triple attention mechanism is designed to encode emotional text in order to obtain the position information of emotional words in the directions of horizontal, vertical, and depth and then model the context and global semantics. More specifically, it assigns the corresponding weights to emotional words while paying attention to salient features, which reduces the use of redundant information. Finally, a multi-scale graph reasoning network is used to learn heterogeneous graphs constructed by emotional words and constructions, which can transfer and aggregate the heterogeneous node features from its neighbors. Note that for the emotional dependency graph, the dependencies between the sentiment words in a sentence and the edges between nodes are measured by the cosine similarity. The relationship graph contains emotions. The overall flow of ConAs-GRNs is shown in Figure 2.

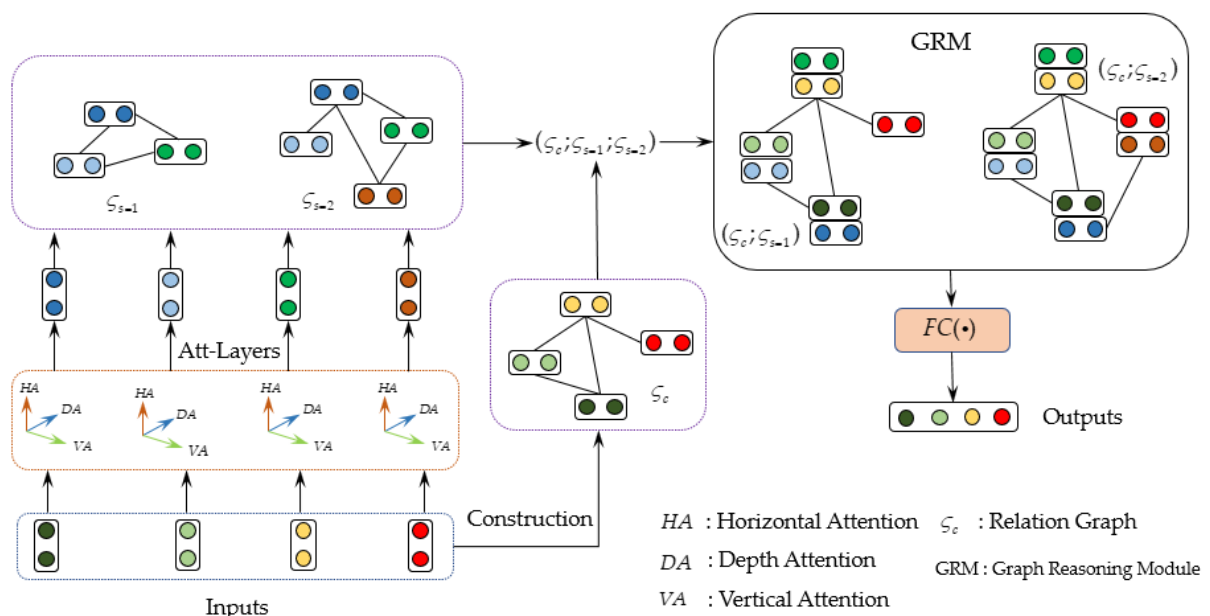


Figure 2. The overall network structure of ConAs-GRNs. $s = 1, 2$ indicates the size of the dependency graph with sentiment words in sentences. ζ_c indicates the construction relation graph with sentiment sentences. $FC(\cdot)$ indicates the operation of fully connected layers. VA, DA, HA indicate the attention mechanism acting in the vertical, deep, and horizontal directions, respectively. $(\zeta_c; \zeta_{s=1}; \zeta_{s=2})$ indicates a weighted fusion of different scale dependency graphs and construction relation graphs. “GRM” indicates the graph reasoning module.

Next, we will introduce the ConAs-GRNs framework, which includes semantic encoding, the construction of heterogeneous graphs, and the reasoning of multi-scale graphs.

3.1. Semantic Encoding Module

The semantic encoding module consists of a Bert [21] layer, a fully connected layer, and a three-directional attention mechanism. First, the mapping vectors extracted by the Bert layer are fed into the fully connected layer (FC) in order to extract the rich low-level

semantic information from emotional sentences. Second, the positions and interrelationships of sentiment words in sentiment sentences are encoded by the three-directional attention [22,23] mechanism. Then, we can obtain the fine-grained semantic information of different sentiment words.

We define the input sentiment sentence as X , where each $X \in \{W_1, \dots, W_N\}$ is composed of N words. More specifically, the word vector dimension is represented by D , and the word vector feature of a sentence is represented by $X^{N \times D}$. The encoded features $f_e^{x_i} \in R^{N \times D}$ are obtained by the FC layer. The three-directional attention mechanism is expressed as follows:

$$\begin{cases} f_{fc}^{x_i} = FC(x_i), x_i \in X \in R^{N \times D} \\ f_e^{x_i} = Conv_{1 \times 1}(HA(f_{fc}^{x_i}), VA(f_{fc}^{x_i}), DA(f_{fc}^{x_i})) \end{cases} \quad (1)$$

where $f_{fc}^{x_i}$ indicates the output features of fully connected layers; $FC(\cdot)$ indicates the fully connected operation; $Conv_{1 \times 1}$ indicates a convolution operation with a kernel size of 1×1 ; $HA(\cdot)$, $VA(\cdot)$, $DA(\cdot)$ indicates the horizontal, vertical, and depth attention operation; x_i indicates the i -th emotional sentence in the corpus.

3.2. Establishment of Heterogeneous Graph

In order to obtain the details of the global and contextual semantics, we construct two topological graphs, i.e., the multi-scale affective dependency graphs $\zeta_{s=k}$ and the constructional relational graphs ζ_c , which can be used to explore effective dependencies and constructional relationships that facilitate effective interaction.

According to Equation (1), we encode the features of sentiment sentences as $f_e^{x_i} \in R^{N \times D}$, and we calculate the correlation between emotional words in each sentence by their feature similarity, where each sentiment word is regarded as a node v^j , and the edge ϵ^j between any two nodes is considered as a dependency. Then, we can construct a dependency topology graph $\zeta_{s=k}(v^j, \epsilon^j)$ that contains multiple scales. The corresponding adjacency matrix A^s of $\zeta_{s=k}$ can be represented by the following equation:

$$A_{j,u}^s = \begin{cases} \cos(v^j, v^u) & , v^j \neq v^u \\ 1 & , v^j = v^u \\ 0 & , otherwise \end{cases} \quad (2)$$

where v^j and v^u can be any two nodes in $\zeta_{s=k}$, $A_{j,u}^s = 1$ indicates a self-looping, and $A_{j,u}^s = 0$ indicates that there is no edge between v_j and v^u .

For a connected relational graph ζ_c that includes all words, the construction of a sentence (e.g., "The battery doesn't last long but I'm sure an upgrade battery would solve that problem.") can be defined as "NP -1 CC -1 NP -1" and "NP -1 DT -1 NN -1", where "but" is "CC" and "upgrade" is "DT". The polarity of the first "battery", represented by "CC", is opposite to the polarity of the second "battery", represented by "DT", i.e., the first "battery" is negative and the second "battery" is positive. This means that the same emotional word may have opposite polarity by some specific modifications. Subsequently, we can infer the polarity of the emotion by using this particular construction. After that, we can build a construction-based relational graph ζ_c , with the corresponding aspect word, such as "CC" and "DT" in the construction, as the root node. Then, the adjacency matrix A^c of graph ζ_c can be represented by the following equation:

$$A_{t,r}^c = \begin{cases} \|v^t - v^r\| & , v^t \neq v^r \\ 1 & , v^t = v^r \end{cases} \quad (3)$$

where, v^t, v^r indicates the node of graph ζ_c , $A_{t,r}^c = 1$ indicates that the node is self-looping, and $\|\cdot\|$ indicates the normal formulas.

Based on the weighted edge reduction strategy, the graphs $\zeta_{s=k}$ and ζ_c can be used to form the final multi-scale heterogeneous graph ζ_{cs}^* . This process can be represented by the equation below:

$$\zeta_{cs}^* = WERS(\zeta_c; \zeta_{s=k}), k = 1, 2, 3 \quad (4)$$

where $WERS(\cdot)$ indicates the operation of weighted edge reduction, and $s = k$ indicates the scale of sentiment dependency graph $\zeta_{s=k}$. The adjacency matrix A^* for the weighted reduction strategy can be described as follows:

$$A^* = \psi A^s + \varphi A^c = \begin{bmatrix} \psi A_{11}^s + \varphi A_{11}^c & \cdots & \psi A_{1u}^s + \varphi A_{1r}^c & \cdots & \psi A_{1n}^s + \varphi A_{1n}^c \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \psi A_{ju}^s + \varphi A_{tr}^c & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \psi A_{n1}^s + \varphi A_{n1}^c & \cdots & \psi A_{nu}^s + \varphi A_{nr}^c & \cdots & \psi A_{nn}^s + \varphi A_{nn}^c \end{bmatrix} \quad (5)$$

where $\psi = 0.6$, and $\varphi = 0.4$. n indicates the number of nodes. Then, the fused adjacency matrix A^* can be represented by the following equation:

$$A^* = \begin{cases} \varphi \|v^t - v^r\| + \psi \cos(v^j, v^u) & , v^t \neq v^r, v^j \neq v^u, v^t = v^j, v^r = v^u \\ 1 & , v^t = v^r, v^j = v^u \\ 0 & , otherwise \end{cases} \quad (6)$$

In summary, the heterogeneous graph can capture the dependencies between sentiment words and strengthen their interactions by the sentiment constructions. It helps the graph reasoning networks sense the emotional details in changing statements.

3.3. Graph Reasoning Module

In order to accurately predict the polarity of emotional sentences, we input the constructed heterogeneous graph into the dynamically aggregated graph convolution [24,25] for interactive learning. The learning process can be described as follows:

$$\begin{cases} o_\zeta = \tilde{D}^{-\frac{1}{2}} \tilde{A}^* \tilde{D}^{-\frac{1}{2}} X_\zeta \Theta \\ o_\zeta^{(l)} = \tilde{D}^{-\frac{1}{2}} \tilde{A}^* \sigma(\sum_{m=1}^{(l-1)} (\tilde{A}^*)^{(m)}) \tilde{D}^{-\frac{1}{2}} X_\zeta^{(l-1)} \Theta^{(l-1)} \end{cases} \quad (7)$$

where $m \in l$ and o^l are the outputs of the l -th layers.

After that, we feed the detailed semantics obtained by the dynamically aggregated graph convolution into a fully connected layer FC, which can be described as follows:

$$O_{OUT} = \text{SoftMax}(FC(o_\zeta^{(l)})) \quad (8)$$

where O_{OUT} indicates the prediction by the softmax classifier, and $FC(\cdot)$ indicates the operation of the full connection layer.

To be able to use weighted loss during training to obtain better performance for our proposed ConAs-GRNs framework during training, we use the weighted loss as follows:

$$\begin{cases} \tau_{Total} = \lambda_{mce} \tau_{mce} + \lambda_{dice} \tau_{dice} + \lambda_{fl} \tau_{focal} \\ \lambda_{dice} + \lambda_{mce} + \lambda_{fl} = 1 \end{cases} \quad (9)$$

where τ_{mce} indicates multiclass cross-entropy loss, τ_{dice} indicates dice loss, τ_{fl} indicates focal loss, and $\lambda_{mce}, \lambda_{dice}, \lambda_{fl}$ indicates the weights factor.

Moreover, we use the “Adam” to optimize the learning model, as shown in Algorithm 1.

Algorithm 1 Sentiment classification processing by the ConAs-GRNs frameworks.

Input: the feature matrix of the ConAs-GRNs input is $X^{N \times D}$, the number of graph layers l , the feature of initial low-level $f_{fc}^{x_i}$, the adjacency matrices A^* of the heterogeneous graph according to Equation (6), where $X^{(0)} = X^{N \times D}$;
for $l=0$ to L **do**
 Calculate the output $O_{\zeta}^{(l)}$ by Equations (6) and (7);
 The output of the multi-scale graph is by calculating as ζ_{cs}^* according to Equation (4);
 where the graph of the adjacent matrix A^* is optimized according to Equations (2)–(5);
 achieve the feature of output O_{OUT} by graph reasoning module (GRM) according Equation (8);
end
output: optimization of the training by reconstruction of τ_{Total} and recognition of the results of sentiment classification according to Equation (9).

4. Experimental Results

In our experiments, we use the SemEval 2014 and ACL-14 datasets to verify the validity of ConAS-GRN. Moreover, we discuss the needs of different components used in the proposed ConAS-GRN. In the following sections, we describe the details of our experimental setup.

4.1. Datasets Preparation

SemEval 2014: This dataset includes two parts, “Restaurant” and “Laptop”, where each part consists of three categories: positive, negative, and neutral.

ACL-14: This dataset includes negative, neutral, and positive comments about celebrities, products, and companies. In order to be fair, the negative, neutral, and positive emotion texts in the dataset were divided into 25%, 50%, and 25%. More specifically, the number of training and test samples were set to 6248 and 692, respectively. The details of the dataset are shown in Table 1.

Table 1. Details of the SemEval2014 and ACL-14 datasets; “Restaurant” and “Laptop” represent the SemEval2014 dataset, and “Construction” indicates construction grammar, namely important situations relevant to the human experience.

Datasets	Restaurant		Laptop		ACL-14	
	Training	Testing	Training	Testing	Training	Testing
Positive	2164	728	994	341	3142	346
Negative	807	196	870	128	1562	173
Neutral	637	196	464	169	1562	173
Construction	100,043	1,105,665	241,546	992,438	819,242	286,552

To verify the reliability and effectiveness of ConAs-GRNs, we use Accuracy (Acc) and F1 as performance evaluation metrics.

4.2. Experiment Settings

Datasets processing. Firstly, the emotional sentence corresponding to each construction is mapped into a low-dimensional space, and a specific keyword is used as the root node. Then, we construct a large topology map of the entire statement. Finally, all topological graphs are fused with the use of an edge weight reduction strategy to form a heterogeneous graph that can be directly input to the graph reasoning network that learns their emotional relationships.

Training parameters. In our experiments, some important training details are as follows: (i) the learning rate is set to $1e-4$, the number of iterations is set to 300; (ii) the batch is set to 32, and (iii) the word vector dimension is set to 300. More specifically, our framework is optimized by the “Adam”.

Environment configuration. We implement our ConAs-GRNs model by using the Pytorch platform, and all codes are developed based on python3.7. To ensure fairness, all experiments were carried out on two RTX3090 GPU cards.

4.3. Comparison with Other Sentiment Classification Methods

Taking SemEval2014 and ACL-14 as evaluation samples, we conducted extensive experiments to demonstrate the effectiveness of the proposed ConAS-GRN sentiment classification framework. The specific evaluation results are shown in Table 2.

Table 2. Experimental information of different sentiment classification methods. (Bert) indicates the use of bert for a general word vector. Other methods use gloves for word vector mapping. ConAs-GRNs (Glove) and ConAs-GRNs (Bert) are the proposed sentiment classification frameworks.

Datasets	Laptop		Restaurant		ACL-14	
	Acc	F1	Acc	F1	Acc	F1
SVM	70.13	62.44	77.22	68.61	78.59	70.98
LSTM	70.48	63.92	78.44	70.02	80.33	72.08
BiLSTM	71.87	64.96	79.06	70.28	80.62	71.17
IAN	72.58	63.08	79.64	71.09	81.92	71.46
PBAN	74.22	64.77	81.99	72.41	82.42	72.04
ASCNN	72.64	63.78	80.94	71.61	83.21	73.06
TSN	73.07	63.96	80.87	71.29	81.55	71.73
GCNs	73.44	67.91	79.99	70.08	81.47	71.13
Text-GCNs	75.35	70.11	80.08	70.12	82.54	71.85
DGCNs [26]	78.51	76.15	81.79	73.88	83.62	73.58
AEN(Bert)	78.66	70.30	82.46	72.69	84.02	74.07
ASGCN(Bert)	75.65	71.60	80.87	70.18	82.87	72.53
SDGCN(Bert)	80.35	78.44	83.34	75.69	84.56	75.03
ConAs-GRNs (Glove)	79.42	77.08	82.41	73.99	83.92	74.86
ConAs-GRNs (Bert)	81.43	79.11	84.46	76.92	86.22	76.66

From the evaluation results shown in Table 2, we can draw the following conclusions:

(1) Compared with other sentiment classification methods, using the graph structure to model sentiment semantics can significantly improve the classification accuracy. For example, in the Laptop dataset, the Acc of Text-GCNs are 3.48% and 1.13% higher than that of BiLSTM and PBAN, respectively. The main reason may be that the graph convolutional network strengthens the dependencies between emotional words (nodes) by aggregating the semantic information from the neighboring nodes. Moreover, the use of three-directional attention reduces the use of redundant information.

(2) The ConAs-GRNs (Glove) method can achieve the best performance on the two baseline datasets, SemEval2014 (Restaurant and Laptop) and ACL-14. For example, the classification performance of the ConAs-GRNs (Glove) method on the ACL-14 dataset is higher than that of the ASGCN (Bert) method. The main reason is that multi-scale information can better represent sentiment words and context nodes. Furthermore, multi-scale structural and heterogeneous information can describe the emotional relationships from multiple levels.

(3) The proposed ConAs-GRNs (Bert) method achieved good performance. As we expected, the ConAs-GRNs (Bert) outperformed the SDGCN (Bert) on the SemEval2014 (Restaurant) datasets. Moreover, ConAs-GRNs (Bert) used the weighted edge reduction strategy to prune the primary graph structure; it also used the triple attention mechanism with a positional encoding module to extract the useful features.

Finally, we can see that ConAs-GRNs (Bert) gives give the best performance and is better than ConAs-GRNs (glove) on the two baseline datasets.

4.4. Ablation Experiments

To demonstrate the effectiveness of different components used in the proposed ConAs-GRNs, we conducted a set of ablation experiments on the two baseline datasets, SemEval2014 and ACL-14. The experiment results are shown in Table 3, where “BERT” indicates the BERT method that generates the word vectors. Similarly, “Glove” indicates the glove method. For example, ConAs-GRNs (NoAtt,Glove) indicate that ConAs-GRNs do not use the attention mechanism, and “ss” indicates that there is a single-scale heterogeneous graph.

Table 3. The experimental results of ConAS-GRNs internal components. “Bert” indicates the use of BERT for a general word vector. Other methods use glove for word vector mapping. ConAs-GRNs (NoAtt,Glove) indicates the methods where no attention mechanism is involved in the operation. “ss” indicates the use of a single scale heterogeneous graph.

Datasets	Laptop		Restaurant		ACL-14	
	Acc	F1	Acc	F1	Acc	F1
GRNs (NoAtt,Glove)	76.19	71.98	78.54	73.57	76.98	71.06
GRNs (Glove)	77.57	72.11	79.51	74.88	77.71	71.93
GRNs (NoAtt,Bert)	77.37	72.24	79.24	74.19	77.85	72.29
GRNs (Bert)	78.31	72.40	79.75	74.97	78.46	72.06
ConAs-GRNs (NoAtt,Glove)	77.42	73.08	80.11	74.09	78.02	73.86
ConAs-GRNs (NoAtt,Bert)	78.84	74.11	81.46	75.02	82.12	74.16
ConAs-GRNs (ss,Glove)	77.02	71.18	78.49	70.99	77.62	70.77
ConAs-GRNs (ss,Bert)	77.99	71.44	79.85	71.84	79.22	72.06

From Table 3, we can draw the following conclusions:

(1) Different word vectors have different impacts on the classification performance. For example, the Acc and F1 values of GRNs (Bert) are 0.74% and 0.29% higher than those of GRNs (Glove) on the Laptop dataset. Furthermore, the attention mechanism has a great impact on classification performance. For example, the Acc and F values of GRNs (Glove) are 0.97% and 1.31% higher than those of GRNs (NoAtt,Glove) on the Restaurant baseline dataset. This is because the attention mechanism can effectively reduce redundant information by focusing on salient features.

(2) From Table 3, we can clearly see that the sentiment classification performance is poor when using a single-scale heterogeneous graph. For example, the Acc and F1 values of ConAs-GRNs (ss, Glove) on the ACL-14 dataset are 77.62% and 70.86%, respectively, which are 0.40% and 2.09% lower than those of ConAs-GRNs (NoAtt,Glove). The main reason may be that the heterogeneous graph with a single structure cannot capture the deep multi-scale information in the emotional text, thus resulting in insufficient feature representation. This leads to a lack of emotional dependencies in different aspects within the neighborhood scale. When context semantics and global semantic information are passed across graphs, a lot of detailed information is lost, therefore reducing the interaction between them.

In conclusion, all the components used in the proposed ConAs-GRNs greatly improved the classification performance.

4.5. The Effectiveness of Different Scales and Dimensionality

To verify the effectiveness of heterogeneous graphs and embedding dimensions (Bert), we evaluated the proposed ConAs-GRNs on the ACL-14 dataset. The results are shown in Figure 3, where $d = 100, 200, 300, 400$ indicates the embedding dimensions of a word vector.

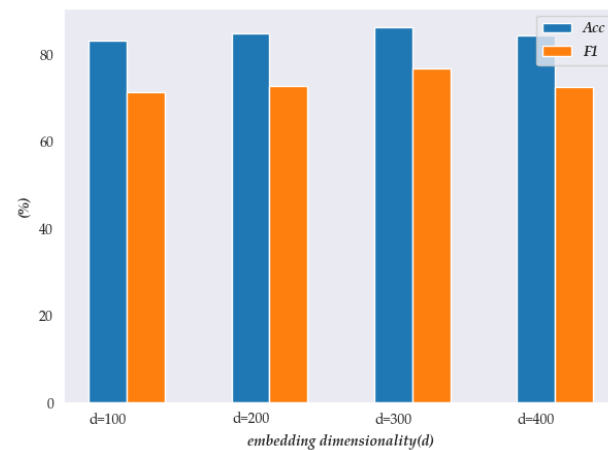


Figure 3. The experimental results of different embedding dimensionalities on our proposed ConAs-GRNs. $d = 100, 200, 300, 400$ indicates the embedding dimensionality of a word vector.

In Figure 3, we can clearly see that as the number of embedding dimensionalities increases, and the classification performance of ConAs-GRNs has a rising trend. For instance, $d = 200$ outperforms $d = 100$ by 1.54% on the two datasets. However, the accuracy is reduced when $d = 400$. The main reason is that when the dimension of the word vector is large, the features are sparse, and the correlation between emotional words is reduced, which makes it difficult to describe the emotional semantics.

Moreover, we selected the values $s = \{1, 2, 3\}$ to investigate the influence on the scale of the heterogeneous graphs.

In Figure 4, we can clearly see that as the number of heterogeneous graph scales increases, the classification accuracy is significantly improved. For example, the Acc of $s = 1, 2, 3$ significantly outperforms that of $s = 1, 2$, $s = 1, 3$, and $s = 2, 3$ on the Laptop dataset, which improved by 1.12%, 1.24%, and 2.49%, respectively. When the scale of the heterogeneous graph is small, it is difficult to obtain rich contextual semantics. Moreover, when the scale is larger, more redundant information is utilized, thus reducing the correlation between different sentiment words.

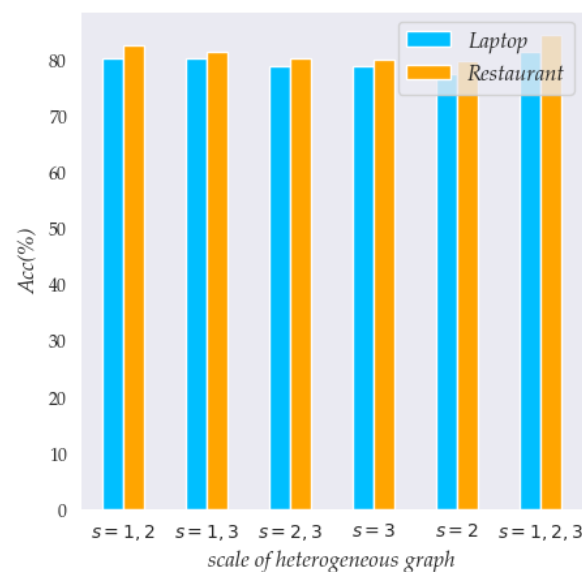


Figure 4. The experimental results of different scales on our proposed ConAs-GRNs, where $s = \{1, 2, 3\}$ indicates the number of heterogeneous graph scales in our proposed classification frameworks.

In addition, we can see that ConAs-GRNs achieved the best performance on the Restaurant datasets when using $s = \{1, 2, 3\}$. This means that multi-scale heterogeneous graphs can strengthen the dependencies between features with different sentiment words and contexts. The experimental results demonstrate the effectiveness of our method.

4.6. The Effectiveness of Different Loss

To demonstrate the effectiveness of our model, we consider the impact of different loss functions on classification performance. The results are shown in Figure 3, where τ_{mce} represents the multi-class cross-entropy, τ_{Focall} aims to suppress the impact of data imbalance, τ_{dice} represents the multi-class dice loss, and τ_{Total} indicates the loss used in our ConAs-GRNs. Then, we can draw the following conclusions:

According to the experimental results of different losses in Table 4, we can see that:

(1) Compared with τ_{dice} and τ_{mce} , τ_{Focall} achieved the best performance on the Laptop and Restaurant datasets. For the Laptop dataset, τ_{Focall} outperforms τ_{mce} by 0.39% and 0.11% respectively. The main reason is that τ_{Focall} effectively suppresses the imbalance of samples. As a result, the proposed ConAs-GRNs can focus on the difficult-to-classify sentiment samples during training.

(2) Compared with τ_{Focall} and τ_{dice} , the proposed ConAs-GRNs uses the special designed loss function τ_{Total} to achieve the best classification performance. For example, the F values are 1.4% and 1.82% higher than τ_{Focall} and τ_{dice} on the Restaurant dataset, respectively. The main reason may be that the weighted τ_{Total} loss function can suppress the data imbalance and keep the loss value in a relatively stable range, thus enabling the network to better learn complex emotional text data.

Table 4. Results of the loss function on the Laptop and Restaurant datasets. τ_{Total} indicates our proposed loss on ConAs-GRNs.

Loss	Laptop		Restaurant	
	Acc	F1	Acc	F1
τ_{mce}	79.82	78.31	82.84	74.75
τ_{Focall}	80.21	78.42	83.47	75.52
τ_{dice}	80.06	78.15	82.91	75.1
τ_{Total}	81.43	79.11	84.46	76.92

4.7. Experimental Results on Different Amounts of Datasets

We also conducted an additional study to investigate the impact of the number of training samples. More specifically, for simplicity, all examples were taken from the ACL-14 dataset. We divided the training sample sizes into 10%, 20%, 30%, and 40%. Figure 5 illustrates the accuracy of ConAs-GRNs with respect to the number of training samples.

As can be seen in Figure 5, with the increase in the number of training samples, the classification accuracy of all models also improves. The main reason may be that when the number of training samples is large, the classification model can learn more useful knowledge from the samples to obtain more powerful discriminative features. Note that our ConAS-GRNs can achieve the best classification performance even on 10% of the training samples, which indicates that ConAS-GRNs has a strong competitive advantage on small-sample datasets.

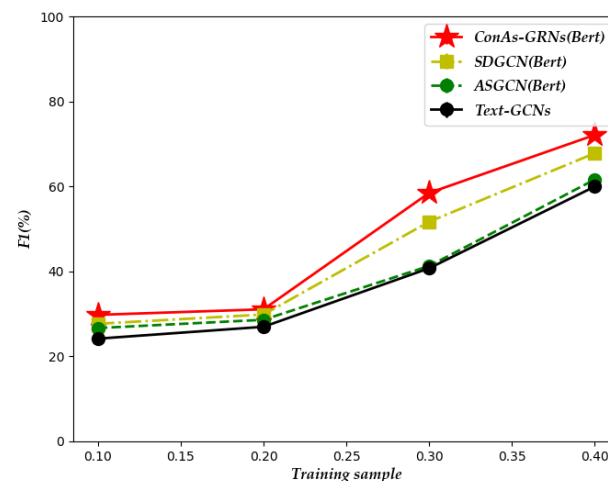


Figure 5. Experimental results of different sizes of datasets on ACL-14.

5. Conclusions and Next Research

In this paper, we developed a sentiment classification algorithm called ConAS-GRN that combines construction information. This framework uses a multi-scale heterogeneous graph and an edge weight reduction strategy to aggregate the context and global semantics of sentiment words in sentiment sentences from multiple levels and scales. Then, we designed a three-directional attention mechanism to emphasize the salient features in sentences from different perspectives, which reduces the use of redundant information and improves the ability of emotional words to express sentences. In order to strengthen the interaction between different topological graphs, a related relational topology was established to avoid the semantic ambiguity of the same sentiment words in different sentences. The extensive experiments validated the effectiveness of our framework. Future research will focus on developing new graph structures to help build a more robust sentiment classification framework.

Author Contributions: Conceptualization, B.C. and W.P.; methodology, B.C. and W.P.; writing—original draft preparation, B.C.; writing—review and editing, W.P. and J.S.; supervision, W.P. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grants No. 61877004 and No. 62007004) and the Major Program of National Social Science Foundation of China (Grant No. 18ZDA295).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ConAs-GRNs	Construction-assisted multi-scale graph reasoning networks
HA	Horizontal Attention
VA	Vertical Attention
DA	Deep Attention
GRM	Graph reasoning module

References

1. Pang, B.; Lee, L. Opinion mining and sentiment analysis. *Synth. Lect. Hum. Lang. Technol.* **2008**, *2*, 1–135.

2. Liu, B. Sentiment analysis and opinion mining. *Found. Trends Inf. Retr.* **2012**, *5*, 1–167.
3. Kiritchenko, S.; Zhu, X.; Cherry, C.; Mohammad, S. Nrcanada-2014: Detecting aspects and sentiment in customer reviews. In Proceedings of the 8th International Workshop on Semantic Evaluation, Dublin, Ireland, 23–24 August 2014; pp. 437–442.
4. Wagner, J.; Arora, P.; Cortes, S.; Barman, U.; Bogdanova, D.; Foster, J.; Tounsi, L. Dcu: Aspect-based polarity classification for semeval task 4. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 23–24 August 2014; pp. 223–229.
5. Blei, D.M. Probabilistic topic models. *Commun. ACM* **2012**, *55*, 77–84. [\[CrossRef\]](#)
6. Thakur, R.K.; Deshpande, M.V. Kernel Optimized-Support Vector Machine and Mapreduce framework for sentiment classification of train reviews. *Int. J. Uncertain. Fuzziness-Knowl. Based Syst.* **2019**, *27*, 1025–1050. [\[CrossRef\]](#)
7. Nafis, N.S.M.; Awang, S. An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification. *IEEE Access* **2021**, *9*, 52177–52192. [\[CrossRef\]](#)
8. Fauzi, M.A. Random Forest Approach fo Sentiment Analysis in Indonesian. *Indones. J. Electr. Eng. Comput. Sci.* **2018**, *12*, 46–50. [\[CrossRef\]](#)
9. Zhou, J.; Huang, J.X.; Chen, Q.; Hu, Q.V.; Wang, T.; He, L. Deep learning for aspect-level sentiment classification: Survey, vision, and challenges. *IEEE Access* **2019**, *7*, 78454–78483. [\[CrossRef\]](#)
10. Zhang, M.; Zhang, Y.; Vo, D.T. Gated neural networks for targeted sentiment analysis. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
11. Huang, B.; Ou, Y.; Carley, K.M. Aspect level sentiment classification with attention-over-attention neural networks. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*; Springer: Cham, Germany, 2018; pp. 197–206.
12. Ma, Y.; Peng, H.; Cambria, E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
13. Zhou, J.; Huang, J.X.; Hu, Q.V.; He, L. Sk-gcn: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification. *Knowl. Based Syst.* **2020**, *205*, 106292. [\[CrossRef\]](#)
14. Chen, Y.; Yuan, J.; You, Q.; Luo, J. Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM. In Proceedings of the 26th ACM International Conference on Multimedia, Vancouver, BC, Canada, 27–31 October 2018; pp. 117–125.
15. Lin, Y.; Wang, C.; Song, H.; Li, Y. Multi-head self-attention transformation networks for aspect-based sentiment analysis. *IEEE Access* **2021**, *9*, 8762–8770. [\[CrossRef\]](#)
16. Pathik, N.; Shukla, P. Aspect Based Sentiment Analysis of Unlabeled Reviews Using Linguistic Rule Based LDA. *J. Cases Inf. Technol. (JCIT)* **2022**, *24*, 1–19. [\[CrossRef\]](#)
17. Huddar, M.G.; Sannakki, S.S.; Rajpurohit, V.S. Attention-based Multi-modal Sentiment Analysis and Emotion Detection in Conversation using RNN. *Int. J. Interact. Multimed. Artif. Intell.* **2021**, *6*, 112–121. [\[CrossRef\]](#)
18. Zeng, J.; Ma, X.; Zhou, K. Enhancing attention-based LSTM with position context for aspect-level sentiment classification. *IEEE Access* **2019**, *7*, 20462–20471. [\[CrossRef\]](#)
19. Zhu, X.; Zhu, L.; Guo, J.; Liang, S.; Dietze, S. GL-GCN: Global and Local Dependency Guided Graph Convolutional Networks for aspect-based sentiment classification. *Expert Syst. Appl.* **2021**, *186*, 115712. [\[CrossRef\]](#)
20. Zhang, C.; Li, Q.; Song, D. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *arXiv* **2019**, arXiv:1909.03477.
21. Islam, S.M.; Bhattacharya, S. AR-BERT: Aspect-relation enhanced Aspect-level Sentiment Classification with Multi-modal Explanations. In Proceedings of the ACM Web Conference 2022, Virtual Event, 25–29 April 2022; pp. 987–998.
22. Yu, J.; Marujo, L.; Jiang, J.; Karuturi, P.; Brendel, W. Improving Multi-Label Emotion Classification via Sentiment Classification with Dual Attention Transfer Network. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018.
23. Zhu, Y.; Zheng, W.; Tang, H. Interactive dual attention network for text sentiment classification. *Comput. Intell. Neurosci.* **2020**, *2020*, 8858717. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
25. Zhao, P.; Hou, L.; Wu, O. Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification. *Knowl. Based Syst.* **2020**, *193*, 105443. [\[CrossRef\]](#)
26. Xiao, L.; Xue, Y.; Wang, H.; Hu, X.; Gu, D.; Zhu, Y. Exploring fine-grained syntactic information for aspect-based sentiment classification with dual graph neural networks. *Neurocomputing* **2022**, *471*, 48–59. [\[CrossRef\]](#)