

## Article

# GCAT-GTCU: Graph-Connected Attention Network and Gate Than Change Unit for Aspect-Level Sentiment Analysis

Chunming Ma , Xiuhong Li \*, Huiru Wang and Ying Zheng 

School of Information Science and Engineering, Xinjiang University, Urumqi 830046, China

\* Correspondence: xjulxh@xju.edu.cn

**Abstract:** Currently, attention mechanisms are widely used in aspect-level sentiment analysis tasks. Previous studies have only used attention mechanisms combined with neural networks for aspect-level sentiment classification, and the feature extraction of the model is insufficient. When the same aspect and sentiment polarity appear in multiple sentences, the semantic information sharing of the same domain is also ignored, resulting in low model performance. To address these problems, the paper proposes an aspect-level sentiment analysis model, GCAT-GTCU, which combines a Graph-connected Attention Network containing symmetry with Gate Than Change Unit. Three nodes of words, sentences, and aspects are constructed, and local and deep-level features of sentences are extracted using CNN splicing BiGRU; node connection information is added to GAT to form a GCAT containing symmetry to realize the information interaction of three nodes, pay attention to the contextual information, and update the shared information of three nodes at any time; a new gating mechanism GTCU is constructed to filter noisy information and control the flow of sentiment information; finally, the three nodes are extracted information to predict the final sentiment polarity. The experimental results on four publicly available datasets show that the model outperforms the baseline model against which it is compared in some very controlled situations.

**Keywords:** aspect-LEVEL sentiment analysis; GCAT; Graph-connected Attention Network; GTCU; Gate Than Change Unit



**Citation:** Ma, C.; Li, X.; Wang, H.; Zheng, Y. GCAT-GTCU: Graph-Connected Attention Network and Gate Than Change Unit for Aspect-Level Sentiment Analysis. *Symmetry* **2023**, *15*, 309. <https://doi.org/10.3390/sym15020309>

Academic Editors: Michel Planat and Christos Volos

Received: 22 December 2022

Revised: 11 January 2023

Accepted: 18 January 2023

Published: 22 January 2023



**Copyright:** © 2023 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

With the development of technology and the continuous progress of society, social media is becoming more and more active, and social comments have penetrated into every aspect of our lives. Sentiment analysis for comments can provide users with more comprehensive sentiment information. Text sentiment analysis can be divided into three categories: sentence-level sentiment analysis, chapter-level sentiment analysis, and aspect-level sentiment analysis according to the granularity of the analysis. Unlike the other two types of sentiment analysis, aspect-level sentiment analysis can predict the sentiment tendency expressed in different aspects of the text, and the context of the text and the sentiment information of different aspects are the focus of accurate sentiment prediction. The final predicted sentiment polarity is classified as positive, negative, or neutral [1]. For example, in the sentence “while the food was very delicious, the service was terrible”, the affective tendency of “food” is positive, but the affective tendency of “service” is negative. However, in sentence-level and chapter-level sentiment analysis, it is not possible to analyze the sentiment expressed by different aspects [2].

Aspect-level sentiment analysis is a fine-grained sentiment classification task [3]. Aspect-level sentiment analysis tasks are divided into two categories according to existing fine-grained sentiment analysis models: ACSA and ATSA. In ACSA, aspect terms abstractly represent entity categories in the text and are used to predict the sentiment polarity of a given aspect term in the text, while in ATSA, nouns, and entities directly present in the text are used as aspect terms. Therefore, ATSA refers to analyzing the sentiment polarity

associated with the target entity in the text. In previous aspect-level sentiment analysis tasks, both sentiment lexicon-based and machine learning-based sentiment analysis methods can have good results in sentiment, but for good features, the selection is the superiority and inferiority of these methods. Deep learning methods [4,5] solve the problems of the above methods, which have good results in both image and speech recognition and are widely used up in natural language processing tasks. Due to the rapid development of deep learning, the ACSA task has been significantly advanced. Based on the attention mechanism or gating mechanism is the approach currently employed in most ACSA tasks, the category of aspects to be analyzed guided by deep learning models focus on.

There are currently three main types of sentiment analysis predictions using deep learning neural networks: RNN-based, CNN-based, and GNN-based. For RNN-based sentiment analysis models, Ref. [6] uses LSTM and GRU to incorporate a gating mechanism on RNN as a way to capture the semantic relationships between aspect words and their contexts; Ref. [7] uses attention mechanism-based LSTM to generate aspect-oriented embeddings and connect sentence embeddings and aspect-oriented embeddings to obtain the final features, and Ref. [8] combines recurrent and recursive neural models to handle aspect-level sentiment analysis tasks. For CNN-based sentiment analysis models, Ref. [9] uses CNN as feature extractors and gating units to control the flow of information, and these components are easily parallelized for training, improving the efficiency of the model. Ref. [10] chose CNN as the base encoder to combine user and product information in a neural network model for sentiment classification for the first time, which greatly improved the model performance. For GNN-based sentiment analysis models, GNNs are divided into several categories such as GCNs and GATs. In GCN-based sentiment analysis, a GCN based on sentence dependency trees was built in Ref. [11], and the model outperformed other baseline models. Ref. [12] defined sentence and aspect nodes and used a heterogeneous graph convolutional network for learning sentence and aspect features with good results. In GAT-based sentiment analysis, Ref. [13] proposed a relational GAT to encode the dependency parse tree structure for sentiment prediction, which greatly improved the efficiency over the baseline model. Ref. [14] proposed an aspect-level sentiment analysis model based on GAT, which encodes grammatical structures into aspect representations and refines them using contrast loss, with good experimental results. In addition, the use of cross-domain network structures for classification tasks is a research approach worth exploring, Ref. [15] constructs a new class of hierarchical fractal networks by iteration in certain cases; Ref. [16] investigated the coherence of networks with recursive features and proposed a class of nested weighted n-polygon networks; Ref. [17–19] has carried out a lot of research on cross-domain network structures based on contrastive learning, and applying these methods in the field of aspect-level sentiment analysis is also a great breakthrough; In Ref. [20–22], cross-domain network structures are used to conduct a large number of studies on text classification tasks when there are few corpora. As one of the classification tasks, whether aspect-level sentiment analysis can draw lessons from these methods is also worth exploring for subsequent studies.

In the aspect-level sentiment analysis task, when the same aspect and sentiment predictions occur in a sentence, the model performance is degraded due to the influence of the features of the aspect category on the wrong sentiment prediction. Most of the current studies have used neural networks with attention mechanisms for aspect-level sentiment analysis, and these models are able to direct the attention mechanism to focus on the semantic information of sentences and aspect categories to extract semantic relations about aspect categories, by extracting aspect-related information from sentences to generate aspect embeddings, and then introducing the attention mechanism to predict sentiment polarity [7,9,11]. However, when the same aspect and sentiment are predicted in a sentence, due to insufficient deep feature extraction and neglect of semantic information sharing, these models can only generate sentence and aspect embedding information through a sentence with no interaction between sentences, which leads to the loss of inter-sentence relations and thus affects the overall performance of the model.

In addition, semantic information for aspect-level sentiment analysis can be learned through sentence and aspect interaction [23–25]. In sentences with the same aspect and sentiment prediction polarity, different semantic expressions of the same aspect will appear, and learning sentiment knowledge from different semantic expressions can improve the generalization of the model, nowadays there are already many methods to extract semantic information from only one sentence [26–28], how to extract deep-level features and semantic information in multiple sentences and achieve interaction to improve the overall performance of the model is a problem worth exploring.

To alleviate the above challenges, we propose a new aspect-level sentiment analysis model, GCAT-GTCU. First construct sentence nodes, word nodes, and aspect nodes, splice CNN and BiGRU to extract deep local and deep features of sentences; then, add node connection information to GAT to form GCAT containing symmetry, update the three nodes and realize information interaction to make semantic information flow in the three nodes; then, introduce a new gating mechanism GTCU to control the path of sentiment flow to the pooling layer; and finally outputting sentiment classification by extracting information from the three nodes.

The main contributions can be summarized as follows:

- The method of constructing word nodes, sentence nodes and aspect nodes and using CNN to splice BiGRU to extract local and deep-level features of sentences is proposed. Which CNN is used to extract local features of sentences; BiGRU focuses on the semantic information of sentence context and extracts relevant features, which solves the problem that CNN cannot extract long sequence features and makes the overall performance of the model improve greatly;
- An aspect-level sentiment analysis model, GCAT-GTCU, which combines a Graph-connected Attention Network containing symmetry with Gate Than Change Unit is proposed. The model adds connection information between nodes in GAT to form GCAT containing symmetry, updates the embedding information of three different nodes of words, sentences and aspects at any time, transmits semantic information between related sentences and aspect words through the interaction of the three nodes, and realizes the sharing of semantic information for the same aspect and sentiment polarity prediction in multiple sentences;
- A new gating mechanism GTCU is proposed to effectively remove the updated shared noise information and control the flow of sentiment information after updating. Finally, the model's overall performance is improved by extracting the aspect sentiment polarity prediction through three kinds of node information.

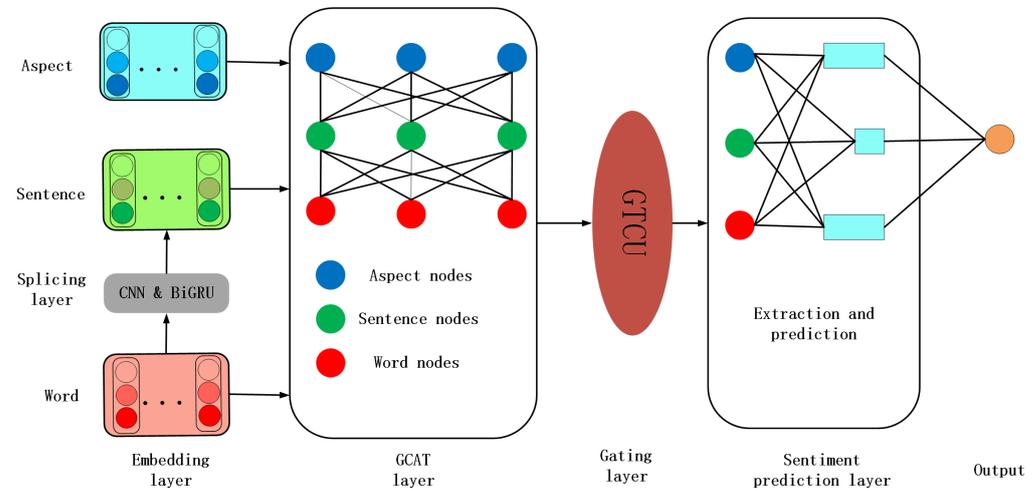
## 2. Method

This paper deals with the ACSA task, which customizes the given sentence  $C = \{s_1, s_2, \dots, s_n\}$ , where the  $i$ th word in the sentence is denoted as  $s_i$ , the aspect category mentioned in the sentence is defined as  $A^C = \{A_1^C, A_2^C, \dots, A_M^C\}$ , a predefined aspect category is denoted as  $A_m^C$ , and the definition of  $O = \{O_1^C, O_2^C, \dots, O_M^C\}$ , where  $O_m^C$  denotes the three emotional polarities of positive, neutral and negative. The main task of this paper is to predict the sentiment tendency of specific aspects of a given text, and the main goal is to enhance feature extraction between multiple sentences and improve information sharing between multiple sentences in a text as a way to improve the accuracy of sentiment prediction.

### 2.1. General Structure of the Model

The GCAT-GTCU model in this paper is divided into five main layers: embedding layer, splicing layer, GCAT layer, Gating layer, and Sentiment prediction layer. The embedding layer initializes the embedding of nodes and edges; the splicing layer mines the local and deep features of sentences; the GCAT layer updates the embedding information of nodes and enables the sharing of semantic information for sentiment prediction; the Gating layer uses the proposed GTCU gating mechanism to control the sentiment flow of shared

embedding and to denoise them; the Sentiment prediction layer uses sentence and aspect nodes to predict sentiment polarity. The specific information of each layer will be described in detail in the following. The overall structure of the model is shown in Figure 1 below.



**Figure 1.** Overall structure of GCAT-GTCU model. The model consists of five layers: 1. embedding layer. 2. splicing layer. 3. GCAT layer. 4. gating layer. 5. sentiment prediction layer. The model uses CNN to splice BiGRU in splicing layer; GCAT layer realizes the information interaction of word nodes, sentence nodes and aspect nodes; GTCU layer filters the noisy information and controls the flow of sentiment information; sentiment prediction layer extracts the information of the three nodes and predicts the final sentiment polarity.

## 2.2. Embedding Layer

In the embedding layer, the dimension of the word embedding is denoted as  $d_s$ . The embedding of nodes and edges in the word is initialized using the pre-trained word embedding  $X_s \in R^{n \times d_s}$ , where  $n \times d_s$  denotes the product of multiple word embeddings, and the sentence is converted into a stitch of word embedding and fed to the CNN and BiGRU layers with different kernel sizes to extract features; the dimension of the sentence embedding is denoted as  $d_c$ , and  $X_c \in R^{h \times d_c}$  denotes the sentence node's final features, where  $h \times d_c$  denotes the product of multiple sentence embeddings; the dimension of aspect embedding is denoted as  $d_a$ , and the aspect nodes are initialized with a one-hot vector and then sent to the linear layer to obtain  $X_a \in R^{m \times d_a}$ , where  $m \times d_a$  denotes the product of multiple aspect embeddings. The positional embedding of the edges of word nodes and sentence nodes, using transformer positional coding [29], can solve the problem that the same word appearing in different positions of a sentence may have different sentiment predictions. In performing feature extraction, a combination of CNN and BiGRU is used to exploit both the ability of CNN to mine local features of sentences and the ability of BiGRU to effectively capture long sequences and extract contextual information from word vectors.

## 2.3. Splicing Layer

The splicing layer mainly uses CNN to splice BiGRU. Due to the capture of continuous word features by the convolutional kernel, CNN is able to extract local features of sentences. the CNN is divided into five parts: input layer, convolutional layer, pooling layer, fully connected layer, and output layer. The input layer takes the output of the word embedding talked about above as the input. The convolution layer sets up the filter sizes as 2, 3 and 4, to extract the local features of the input sentences, as in expression (1):

$$c_s = f\{\theta \cdot X_{s:s+o-1} + b\} \quad (1)$$

where  $\theta$  denotes the convolution kernel,  $o$  denotes the dimension, and the sentence vector  $X_{s:s+o-1}$  consists of  $s$  to  $s+o-1$  words with bias  $b$ , and the feature matrix  $c = [c_p]$ , with  $p = 1, 2, 3, \dots, n - o + 1$  is obtained after the convolution layer.

The pooling layer downsampling the obtained local feature matrix to obtain the optimal solution for the local values, as in expression (2):

$$D_s = \max(c_k), k = 1, 2, 3, \dots, n - l + 1 \quad (2)$$

where  $c$  denotes the obtained local feature matrix and  $D_s$  denotes the optimal solution.

The fully connected layer solves the problem of sequence breakage after the pooling layer, as in expression (3):

$$U = [D_u], u = 1, 2, 3, \dots, n \quad (3)$$

where  $U$  denotes the feature matrix concatenated by the pooled vector  $D_s$ ; it is fully concatenated in order to be stitched into the BiGRU where the serialized structure information must be input.

The output layer outputs the final serialized structure information.

BiGRU is composed of bidirectional GRU, which can combine words in order, each word can obtain the information of the previous words, and the final word vector contains the context information of all sentences, which can realize the extraction of deep-level features of sentences. Therefore, BiGRU is used to obtain contextual information simultaneously. The output of CNN is used as the input of BiGRU, which consists of forwarding GRU, reverse GRU and forward-backward output state connection layer.

If the hidden state of the forward GRU output at the moment  $t$  is  $h_t^f$  and the hidden state of the reverse GRU output is  $h_t^r$ , then the hidden state of the BiGRU output is  $h_t$ , and the calculation process is as in expressions (4) to (6):

$$h_t^f = GRU(h_{t-1}^f, D_t) \quad (4)$$

$$h_t^r = GRU(h_{t-1}^r, D_t) \quad (5)$$

$$h_t = w_t h_t^f + v_t h_t^r - 1 + b_t \quad (6)$$

where  $w_t, v_t$  denote the weight matrix, the GRU input at time  $t$  is  $D_t$ , the bias vector is  $b_t$ .

#### 2.4. GCAT Layer

In the GCAT layer, the initialized aspect graph  $G = \{N_w, N_s, N_a, L_{ws}, L_{sa}\}$  is firstly given,  $N_w, N_s, N_a$  represent word nodes, sentence nodes and aspect nodes respectively, and  $L_{ws}, L_{sa}$  represent the connection information of word nodes and sentence nodes, sentence nodes and aspect nodes respectively. As shown in the GCAT layer in Figure 1 above.

Stacking network layers using GAT allows obtaining the neighborhood characteristics of each node and assigning different weights to different nodes in the neighborhood. Since GAT does not depend completely on the graph structure, but only on the edges, they do not need to focus first on the graph structure information and eliminate the need for costly matrix operations [30].

Adding connection information between nodes in GAT forms GCAT containing symmetry. In a sentence, there are many aspect categories. Still, not all aspect information is beneficial to other sentences, so the GCAT is needed to control the information flow and ensure that all useful information is passed between sentences. The semantic model can be learned by forming interactions between sentences and aspects; by updating sentence nodes, word nodes and aspect nodes can learn knowledge from other similar sentence nodes, which enhances sentence features and improves the representation ability of the model; the generalization ability of the model is improved by aspect nodes learning semantic diversity knowledge from sentence nodes. Add connection information to the traditional attention network and connect word nodes and sentence nodes with connection position

information. The hidden states of nodes can be updated by the multi-headed attention mechanism of the GCAT layer, which is expressed in the following expressions (7) to (8):

$$H_c^{t+1} = GCAT(H_c^t, H_c^g) \quad (7)$$

$$\alpha_k^{i,j} = \frac{\exp(f(W_k^a [W_k^n h_i; W_k^v h_j; L_{ij}]))}{\sum_{l \in K_i} \exp(f(W_k^a [W_k^n h_i; W_k^v h_l; L_{il}]))} \quad (8)$$

In the expression (7),  $H_c^t$  is the representation of the node at the  $t$ th iteration and  $H_c^g$  is the representation of the proximity node. In the expression (8),  $h_i$  and  $h_j$  denote the hidden states of node  $i$  and  $j$ , respectively, and  $K_i$  denote the adjacent hidden states of node  $i$ . The learnable weights are denoted as  $W_k^a$ ,  $W_k^n$  and  $W_k^v$ , and  $f(\cdot)$  is the *LeakyReLU* function [31],  $L_{ij}$  denotes the added connection location information, belonging to the connection embedding of nodes  $i$  and  $j$ . After certain iterations, sentence nodes will aggregate information not only from word nodes and aspect nodes, but also from other similar sentence nodes through word nodes and aspect nodes, thus allowing a better representation of the three nodes with structural and semantic information, and for predicting sentiment polarity.

### 2.5. Gating Layer

The gating mechanisms based on GTU and GLU have been widely used in the field of natural language processing with good results. Among them, the GTU is represented as  $\tanh(Wx + b) \odot \sigma(Vx + b)$  and the GLU is represented as  $(Wx + b) \odot \sigma(Vx + b)$  can convey more useful information [32].

In this paper, after sharing information among nodes in the graph attention layer, a new Gate Than Change Unit named GTCU is constructed to control the path of sentiment information flow, prevent information loss, filter out useful information for aspect-level sentiment analysis, and the experimental results prove the effectiveness of this method, which proceeds as follows (9) to (11):

$$q = \tanh(W_q q + b_q) \quad (9)$$

$$\lambda = \sigma(W_\lambda q + b_\lambda) \quad (10)$$

$$g = ReLU(\lambda \odot q + (1 - \lambda) \odot q) \quad (11)$$

The final result obtained is  $g = GTCU(q)$ , where  $W_q$  and  $W_\lambda$  both denote training weights and  $b_q$  and  $b_\lambda$  denote biases; ReLU is a linear function that controls the path of the flow of emotional information and filters out useful information through rectification.

### 2.6. Sentiment Prediction Layer

In the sentiment prediction layer, the polarity probability of the predicted sentiment is represented by the three nodes extracted together. The word node  $r$  is represented as  $N_w^r$ , the sentence node  $i$  is represented as  $N_S^i$ , and the sentiment category  $m$  in aspect node  $j$  is represented as  $N_A^{jm}$ . The predicted sentiment has the following expressions (12) to (14):

$$N_A^j = N_A^{j1} \parallel N_A^{j2} \parallel \dots \parallel N_A^{jm} \quad (12)$$

$$\hat{z}_i^j = \text{softmax}(W_a N_w^r N_S^i N_A^j + b_a) \quad (13)$$

$$\mathcal{L} = - \sum_i \sum_j z_i^j \hat{z}_i^j \quad (14)$$

where  $W_a$  denotes the learning weights, and  $b_a$  denotes the bias;  $\mathcal{L}$  denotes the result of model training, which is trained by the loss of the minimum cross-entropy between the classification accuracy  $z_i^j$  and the sentiment prediction output  $\hat{z}_i^j$ .

### 3. Experiments

#### 3.1. Dataset

Experiments evaluate the validity of the model, and for the ACSA task, this paper uses experiments from four publicly available datasets for validation. SemEval 2014 Task 4 has a dataset on restaurant reviews Restaurant2014 [33], which contains five aspects of food, price, service, atmosphere, and mixture, and contains four sentiments of positive, negative, neutral, and conflicting polarity; the data on restaurant reviews from SemEval 2014 Task 4 [33], SemEval 2015 Task 12 [34] and SemEval-2016 Task 5 [35] were mixed together, conflicting labels were removed, and incompatible data were fixed during merging to form the RestaurantLarge dataset [9]; in the Restaurant2014 and RestaurantLarge datasets, a sentence mostly contains data with only one aspect category, and to detect the effect of the model under different aspect with different sentiment polarity, the Restaurant2014\_hard and RestaurantLarge\_hard datasets [9] were used again, again removing conflicting labels and fixing incompatible data during merging compatible data, both of these two datasets a sentence contains at least two aspect categories of sentiment polarity, which better reflects the effect of the model in this paper. The statistics of the above four datasets are shown in Table 1 below.

**Table 1.** ACSA task dataset statistics.

Dataset	Restaurant2014		Restaurant2014_Hard		Restaurant Large		Restaurant Large_Hard	
	Train	Test	Train	Test	Train	Test	Train	Test
Positive	2179	657	125	21	2710	1505	182	92
Negative	839	222	123	20	1198	680	178	81
Neutral	500	94	47	12	757	241	107	61

#### 3.2. Contrast Model

Using CNN as sentiment feature extraction is a common approach, for example, the TextCNN model uses CNN with different kernel sizes for sentence-level sentiment analysis [36], which can be used as a baseline model for most aspect-level sentiment analysis tasks; the LSTM-based aspect-level sentiment analysis model has a wide range of applications in sentiment classification tasks, especially in sentence-level sentiment analysis with good results [37], and the model through LSTM architecture is a major breakthrough in the study of aspect and sentiment information; the BERT-based aspect-level sentiment analysis model makes full use of the good advantages of BERT and has a great improvement in model performance; given that the research in this paper is dedicated to solving the problems of feature extraction and the shareability of aspect and sentiment semantic information, the comparative experimental models are divided into four categories: standard baseline models, LSTM-based aspect-level sentiment analysis models, BERT-based aspect-level sentiment analysis models, and other aspect-level sentiment analysis models.

The standard baseline model uses the TextCNN model, which applies CNN to text classification tasks and makes full use of the CNN's ability to capture local features [36].

In LSTM-based aspect-level sentiment analysis models, the AT-LSTM jointly models LSTM and attention, and introduces aspect information in the calculation of attention weights to address the relationship between a given aspect and contextual information [7]; the ATAE-LSTM combines aspect embedding and word embedding on the basis of the AT-LSTM to make fuller use of aspect-level information [7].

Among the BERT-based aspect-level sentiment analysis models,  $BERT_{large}$  is a BERT bidirectional language model, which greatly improves the model architecture problem by using MLM as pre-training, with significant performance improvement over using  $BERT_{large}$  [38]; BERT-pair-QA-B aspect-level sentiment analysis has a significant effect in aspect detection, and sentiment classification [39]; SCAN-BERT solves the problem of having multiple aspects and multiple sentiment words in a sentence [40].

Among other aspect-level sentiment analysis models, the GCAE model uses CNN and gating units to improve the speed of model training and solve the problem of difficult parallel computation in model training using RNN and Attention in the ACSA task [9]; the CapsNet model uses a capsule layer for aspect-level sentiment analysis, and its model works well [41]; the As-capsule model was designed for the aspect-level sentiment analysis task, and excellent results were achieved by sharing multiple modules to enable different capsules to exchange information [42].

Under the same configuration environment, the experimental results of the above-mentioned models on four publicly available datasets: Restaurant2014, Restaurant2014\_Hard, RestaurantLarge, and RestaurantLarge\_hard, are compared with the experimental results of the model proposed in this paper, whose experimental setup and result analysis are described below.

### 3.3. Experimental Setup

Since the BERT-based aspect-level sentiment analysis model has significant performance advantages, this paper provides two experiments on the GCAT-GTCU model: GCAT-GTCU\_Glove and GCAT-GTCU\_BERT.

For both GCAT-GTCU\_Glove and GCAT-GTCU\_BERT, this paper uses pre-trained 300-D GloVe vectors to initialize the word nodes [43] and one-hot vectors to initialize the aspect nodes. The difference is that in GCAT-GTCU\_Glove, in this paper, the CNN splicing BiGRU with Glove vectors as input initializes the sentence nodes and extracts the relevant sentence features. The CNN filter size is set to 2, 3, and 4, the number of channels is set to 500, and the hidden layer size of BiGRU is set to 128; while in GCAT-GTCU\_BERT, this paper uses  $BERT_{base}$  to extract sentence features [38], and uses a BERT-based model implemented in PyTorch to initialize the sentence nodes, with both  $BERT_{base}$  and  $BERT_{large}$  are both used with their suggested hyperparameter settings. Still, the training process does not update BERT. In addition, both GCAT-GTCU\_Glove and GCAT-GTCU\_BERT initialize the connections between word nodes and sentence nodes with positional encoding [29], both with a learning rate of 0.001 and a dropout rate of 0.5, and both with a training duration of 200. Assuming that the test set was unavailable during the training period, all the methods were run five times in this paper, and their averages on the test set were taken as the final results.

For the other experiments conducted for comparison, the standard baseline model, the LSTM-based aspect-level sentiment analysis model, and the other aspect-level sentiment analysis models were initialized with pre-trained 300-D GloVe vectors. Among them, the filter sizes of the standard baseline model TextCNN were set to 2, 3, and 4, and the number of channels was set to 256; the  $BERT_{base}$  hidden layer size in the LSTM-based aspect-level sentiment analysis model was set to 128, the number of layers was set to 2, the dropout rate was set to 0.5, and the training period was 100 [37]; the other aspect-level sentiment analysis models, following their original paper's parameter settings. The BERT-based aspect-level sentiment analysis models all use the PyTorch implementation of the BERT-based model to initialize the sentence nodes with their suggested hyperparameter settings.

### 3.4. Analysis of Results

The experimental results of comparing the models in this paper with other models on four publicly available datasets are shown in Table 2 below, which shows the results on the average F1 value and accuracy of each model and the models in this paper on four publicly available datasets, respectively, from which we can draw some conclusions.

The models work well on two datasets, Restaurant2014 and RestaurantLarge, but work poorly on two datasets, Restaurant2014\_hard and RestaurantLarge\_hard. This is because a sentence in the Restaurant2014 and RestaurantLarge datasets mostly contains data with only one aspect category, and its task type becomes simple, so it works well. However, in order to detect the effect of the model under different sentiment polarity of different aspects, to make its experiments closer to real scenarios, and to better illustrate

the significance of the model improvement in this paper, the Restaurant2014\_hard, and RestaurantLarge\_hard datasets must be used. The results of this paper's model outperform the baseline model against which it is compared in some very controlled situations on all four publicly available datasets, proving the effectiveness of this paper's method under different aspects of different sentiment polarities and better demonstrating the applicability of this paper's model to various situations.

Further analysis of the reasons why the model in this paper appears to work well are mainly in three aspects below:

- Most previous models achieve feature extraction by CNN or LSTM only, lacking the capture of contextual semantic information, while this paper splices CNN and BiGRU to extract both contextual and local features of sentences;
- The information sharing of previous models in the same domain is insufficient, while this paper introduces a GAT, on which three nodes are built and connection information between them is added to form a graph connected attention network, which enables the sharing of semantic information between nodes through connection interaction, considering both the case of different aspect categories in one sentence and the case of the same aspect and emotional polarity in multiple sentences The situation of;
- Previous models did not filter the information and control the flow direction before output, while this paper filters the noisy information and controls the flow direction of the shared embedded sentiment information by forming a new gating mechanism, GTCU, and finally predicts the final sentiment polarity after extracting useful information from the three nodes.

**Table 2.** Comparison of experimental results.

Dataset	Restaurant 2014		Restaurant 2014_Hard		Restaurant Large		Restaurant Large_Hard	
	F1	Acc	F1	Acc	F1	Acc	F1	Acc
TextCNN [36]	66.14	78.96	31.93	42.73	67.66	80.01	41.06	50.55
LSTM <sub>base</sub> [37]	64.12	76.45	20.03	36.76	66.14	77.6	20.11	36.45
AT-LSTM [7]	71.25	82.17	33.2	43.32	59.66	80.36	40.33	49.81
ATAE-LSTM [7]	66.05	77.67	37.88	47.82	65.4	78.78	21.83	39.96
GCAE [9]	64.05	77.15	21.96	39.89	66.86	79.74	33.98	48.85
CapsNet [41]	69.51	80.76	52.33	57.86	65.89	76.12	49.74	56.98
As-capsule [42]	70.54	82.08	51.63	55.06	66.43	81.69	49.57	55.72
GCAE-GTCU_Glove	<b>71.29</b>	<b>82.22</b>	<b>53.82</b>	<b>62.72</b>	<b>70.11</b>	<b>84.03</b>	<b>57.63</b>	<b>61.45</b>
BERT <sub>base</sub> [38]	75.33	87.16	39.66	49.67	67.64	82.71	36.18	47.16
BERT <sub>large</sub> [38]	75.61	87.74	39.71	50.64	68.21	83.24	39.01	48.27
BERT-pair-QA-B [39]	76.1	87.65	55.56	66.81	72.81	86.67	63.52	68.11
SCAN-BERT [40]	76.82	<b>88.45</b>	55.67	65.83	72.61	86.81	57.81	64.79
GCAE-GTCU_BERT	<b>76.95</b>	86.83	<b>56.53</b>	<b>67.68</b>	<b>74.18</b>	<b>87.22</b>	<b>64.63</b>	<b>69.65</b>

Attachment: The bold shows the best results presented in this experiment under certain conditions.

In addition, this experiment takes advantage of the BERT to make the overall effect of the model greatly improved, and does not improve on the creativity of the model itself, which is also the shortcoming of this experiment, and subsequent research needs to improve the performance from the perspective of the model itself.

### 3.5. Ablation Experiments

In order to further demonstrate the effects of all parties in this paper's model GCAE-GTCU on its overall effectiveness, ablation experiments are set up. Since the two datasets, Restaurant2014\_hard and RestaurantLarge\_hard are more suitable for the application scenarios required to solve the problem in this paper and can better prove the effect of the model improvement in this paper, only the ablation experiments under these two datasets are set up.

In this paper, the experiments are set up in seven cases: without BiGRU, without word nodes, without sentence nodes, without aspect nodes, without nodes connection, without GTCU, and without attention mechanism. The accuracy results in Restaurant2014\_hard and RestaurantLarge\_hard datasets are shown in Table 3 below.

From the table, we know that in feature extraction, removing the BiGRU added in this paper decreases the accuracy on both datasets, which indicates that the feature extraction with the addition of contextual information has some improvement on the overall effect of the model; removing any of the three nodes or removing the connection information between them deteriorates the effect on both datasets, which also proves that the method of constructing three nodes and adding the connection information between them in this paper. This also proves that the method of constructing three nodes and adding connection information between them to achieve information sharing is effective; removing the GTCU gating mechanism proposed in this paper before the sentiment prediction output has a certain effect on the model performance in both datasets, which indicates that the GTCU gating mechanism also plays a role in filtering useful information and controlling the flow of sentiment.

It is worth noting that the model with the attention mechanism removed has the largest decrease in accuracy on both datasets, which is due to the fact that the attention mechanism can control the flow of information between nodes and can focus on different aspects of the sentiment polarity of different nodes. Therefore, using the attention mechanism is also a prerequisite for constructing graph nodes and adding connectivity information between nodes in this paper.

**Table 3.** Results of ablation experiments.

Model	Accuracy on the Restaurant2014_Hard	Accuracy on the RestaurantLarge_Hard
GCAT-GTCU	62.72	61.45
w/o BiGRU	62.28	61.07
w/o word nodes	60.89	60.16
w/o Sentence nodes	57.54	56.88
w/o Aspect nodes	55.38	54.72
w/o Nodes connection	58.43	57.96
w/o GTCU	62.26	61.14
w/o Attention mechanism	50.43	50.04

### 3.6. Model Training Efficiency

Experiments were conducted on the datasets Restaurant2014\_hard and RestaurantLarge\_hard using  $BERT_{base}$  and this paper's model GCAT-GTCU\_BERT, respectively, to further analyze the training efficiency of the models. The training times of  $BERT_{base}$  and GCAT-GTCU\_BERT on one epoch were obtained by averaging 30 experiments on the same hardware platform, and the experimental results are shown in Table 4 below.

**Table 4.** Comparison results of model training efficiency.

Model	Training/Epoch on the Restaurant2014_Hard	Training/Epoch on the RestaurantLarge_Hard
$BERT_{base}$	30.24 s	29.87 s
GCAT-GTCU_BERT	15.79 s	15.34 s

According to the conclusion in the table, it can be obtained that the training efficiency of GCAT-GTCU\_BERT is higher than that of  $BERT_{base}$  both on Restaurant2014\_hard and RestaurantLarge\_hard, which proves that the training efficiency of the model in this paper has some advantages, of course, mainly because there is no update of BERT during the experiment and only BERT is used to initialize the nodes.

#### 4. Conclusions

In this paper, we propose a model GCAT-GTCU for aspect-level sentiment analysis, constructing three nodes of the word, sentence, and aspect and using CNN splicing Bi-GRU to better extract local and deep-level features of sentences, solving the problem of insufficient feature extraction; using GAT and adding connection information between its nodes to achieve the sharing of semantic information between nodes, improving the accuracy of sentiment prediction in two cases: the appearance of different aspect categories in a sentence and the appearance of the same aspect and sentiment polarity in multiple sentences. GCAT-GTCU has been used to construct a new gating mechanism GTCU to filter out useful information and control its flow. The experimental results of GCAT-GTCU in comparison with other models on four public datasets show good results of this paper's mode. However, the improvement of the model effect in this paper is highly dependent on BERT, and future research will shed the advantages of BERT itself to achieve the improvement of the overall performance of the aspect-level sentiment analysis model. In addition, the use of contrastive learning for aspect-level sentiment analysis tasks is also a direction worth exploring, which will be further investigated by the authors of this paper.

**Author Contributions:** Writing—original draft preparation, C.M.; writing—review and editing, X.L.; Data curation, H.W.; methodology, Y.Z.; Funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the open project of key laboratory, Xinjiang Uygur Autonomous Region (No. 2022D04079) and the National Science Foundation of China (No. 11504313, No. U1911401, No. 61433012).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We would like to thank LI Zhe, who is a PhD student at The Hong Kong Polytechnic University, for helping proofread our revised paper after major revisions seriously.

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

#### References

1. Schouten, K.; Frasinca, F. Survey on aspect-level sentiment analysis. *IEEE Trans. Knowl. Data Eng.* **2015**, *28*, 813–830. [\[CrossRef\]](#)
2. Brun, C.; Popa, D.N.; Roux, C. XRCE: Hybrid Classification for Aspect-based Sentiment Analysis. In Proceedings of the COLING, Dublin, Ireland, 23–29 August 2014; pp. 838–842.
3. Luo, J.; Huang, S.; Wang, R. A fine-grained sentiment analysis of online guest reviews of economy hotels in China. *J. Hosp. Mark. Manag.* **2021**, *30*, 71–95. [\[CrossRef\]](#)
4. Do, H.H.; Prasad, P.W.C.; Maag, A.; Alsadoon, A. Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Syst. Appl.* **2019**, *118*, 272–299. [\[CrossRef\]](#)
5. Li, Y.; Xie, M.; Yi, Y. Fine-grained sentiment analysis for social network platform based on deep-learning model. *Appl. Res. Comput.* **2017**, *34*, 743–747.
6. Zhang, Z.; Robinson, D.; Tepper, J. Detecting hate speech on twitter using a convolution-gru based deep neural network. In Proceedings of the European Semantic Web Conference, Heraklion, Crete, Greece, 2 June 2018; pp. 745–760.
7. Wang, Y.; Huang, M.; Zhu, X.; Zhao, L. Attention-based LSTM for aspect-level sentiment classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 606–615.
8. Aydin, C.R.; Güngör, T. Combination of recursive and recurrent neural networks for aspect-based sentiment analysis using inter-aspect relations. *IEEE Access* **2020**, *8*, 77820–77832. [\[CrossRef\]](#)
9. Xue, W.; Li, T. Aspect based sentiment analysis with gated convolutional networks. *arXiv* **2018**, arXiv:1805.07043.
10. Tang, D.; Qin, B.; Liu, T. Learning semantic representations of users and products for document level sentiment classification. In Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP), Beijing, China, 26–31 July 2015; pp. 1014–1023.
11. Zhang, C.; Li, Q.; Song, D. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *arXiv* **2019**, arXiv:1909.03477. 77820–77832.
12. Xu, K.; Zhao, H.; Liu, T. Aspect-specific heterogeneous graph convolutional network for aspect-based sentiment classification. *IEEE Access* **2020**, *8*, 139346–139355. [\[CrossRef\]](#)

13. Wang, K.; Shen, W.; Yang, Y.; Quan, X.; Wang, R. Relational graph attention network for aspect-based sentiment analysis. *arXiv* **2020**, arXiv:2004.12362.
14. Huang, Z.; Wu, G.; Qian, X.; Zhang, B. Graph Attention Network for Financial Aspect-based Sentiment Classification with Contrastive Learning. In Proceedings of the 2022 IEEE 20th International Conference on Industrial Informatics (INDIN), Perth, Australia, 25–28 July 2022; pp. 668–673.
15. Liu, J.B.; Bao, Y.; Zheng, W.T. Analyses of Some Structural Properties on a Class of Hierarchical Scale-free Networks. *arXiv* **2022**, arXiv:2203.12361.
16. Liu, J.B.; Bao, Y.; Zheng, W.T.; Hayat, S. Network coherence analysis on a family of nested weighted n-polygon networks. *Fractals* **2021**, *29*, 215–260. [[CrossRef](#)]
17. Li, Z.; Mak, M. W.; Meng, H. M. L. Discriminative Speaker Representation via Contrastive Learning with Class-Aware Attention in Angular Space. *arXiv* **2022**, arXiv:2210.16622.
18. Li, Z.; Mak, M. W. Speaker Representation Learning via Contrastive Loss with Maximal Speaker Separability. In Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Chiang Mai, Thailand, 7–10 November 2022; pp. 962–967.
19. Sheng, J.; Zhang, Y.; Cai, J.; Lam, S.K.; Li, Z.; Zhang, J.; Teng, X. Multi-view Contrastive Learning with Additive Margin for Adaptive Nasopharyngeal Carcinoma Radiotherapy Prediction. *arXiv* **2022**, arXiv:2210.15201.
20. Ke, Z.; Sheng, J.; Li, Z.; Silamu, W.; Guo, Q. Knowledge-guided sentiment analysis via learning from natural language explanations. *IEEE Access* **2021**, *9*, 3570–3578. [[CrossRef](#)]
21. Li, Z.; Li, X.; Sheng, J.; Slamun, W. AgglutiFiT: efficient low-resource agglutinative language model fine-tuning. *IEEE Access* **2020**, *8*, 148489–148499. [[CrossRef](#)]
22. Li, X.; Li, Z.; Sheng, J.; Slamun, W. Low-Resource Text Classification via Cross-Lingual Language Model Fine-Tuning. In Proceedings of the China National Conference on Chinese Computational Linguistics, Haikou, Hainan, China, 31 October–1 November 2020; pp. 231–246.
23. Hsu, T.W.; Chen, C.C.; Huang, H.H.; Chen, H.H. Semantics-Preserved Data Augmentation for Aspect-Based Sentiment Analysis. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online, 7–11 November 2021; pp. 4417–4422.
24. Zhang, K.; Zhang, K.; Zhang, M.; Zhao, H.; Liu, Q.; Wu, W.; Chen, E. Incorporating Dynamic Semantics into Pre-Trained Language Model for Aspect-based Sentiment Analysis. *arXiv* **2022**, arXiv:2203.16369.
25. Zhang, Z.; Zhou, Z.; Wang, Y. SSEGCN: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, WA, USA, 10–15 July 2022; pp. 4916–4925.
26. Hosseini-Asl, E.; Liu, W.; Xiong, C. A Generative Language Model for Few-shot Aspect-Based Sentiment Analysis. *arXiv* **2022**, arXiv:2204.05356.
27. Zheng, J.; Friedman, S.; Schmer-Galunder, S.; Magnusson, I.; Wheelock, R.; Gottlieb, J.; Miller, C. Towards a Multi-Entity Aspect-Based Sentiment Analysis for Characterizing Directed Social Regard in Online Messaging. In Proceedings of the Sixth Workshop on Online Abuse and Harms (WOAH), Online, 14 June 2022; pp. 203–208.
28. Tian, Y.; Chen, G.; Song, Y. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Mexico City, Mexico, 6–11 June 2021; pp. 2910–2922.
29. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
30. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, arXiv:1710.10903.
31. Xu, B.; Wang, N.; Chen, T.; Li, M. Empirical evaluation of rectified activations in convolutional network. *arXiv* **2015**, arXiv:1505.00853.
32. Dauphin, Y.N.; Fan, A.; Auli, M.; Grangier, D. Language modeling with gated convolutional networks. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 933–941.
33. Kirange, D.K.; Deshmukh, R.R.; Kirange, M.D.K. Aspect based sentiment analysis semeval-2014 task 4. *Asian J. Comput. Sci. Inf. Technol. (AJCSIT)* **2014**, *4*, 72–75.
34. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; Androutsopoulos, I. Semeval-2015 task 12: Aspect based sentiment analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, CO, USA, 4–5 June 2015; pp. 486–495.
35. Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; L-Smadi, M.A.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; Clercq, O.D.; et al. Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the International Workshop on Semantic Evaluation, San Diego, CA, USA, 16–17 June 2016; pp. 19–30.
36. Kim, Y. Convolutional neural networks for sentence classification. In Proceedings of the Conference Empirical Methods Natural Lang. Process (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1746–1751.
37. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
38. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.

39. Sun, C.; Huang, L.; Qiu, X. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. *arXiv* **2019**, arXiv:1903.09588.
40. Li, Y.; Yin, C.; Zhong, S. Sentence constituent-aware aspect-category sentiment analysis with graph attention networks. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Zhengzhou, Henan, China, 16–18 October 2020; pp. 815–827.
41. Jiang, Q.; Chen, L.; Xu, R.; Ao, X.; Yang, M. A challenge dataset and effective models for aspect-based sentiment analysis. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 5–7 November 2019; pp. 6280–6285.
42. Wang, Y.; Sun, A.; Huang, M.; Zhu, X. Aspect-level sentiment analysis using as-capsules. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 2033–2044.
43. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1532–1543.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.