

Aspect-Based Sentiment Analysis Based on Knowledge Enhancement and Pre-trained Language Model

Lizhu Ye*

School of Information Technology and Engineering, Guangzhou College of Commerce, Guangzhou 511363, P. R. China
School of Graduate Studies, Management and Science University, 40100 Shah Alam, Selangor, Malaysia
lizhuye2018@gcc.edu.cn

Md Gapar Md Johar

Software Engineering and Digital Innovation Center, Management and Science University
40100 Shah Alam, Selangor, Malaysia
mdgapar@msu.edu.my

Mohammed Hazim Alkawaz

Department of Computer Science, College of Education for Pure Science, University of Mosul, Mosul 41002, Iraq
mohammed.ameen@uomosul.edu.iq

*Corresponding author: Lizhu Ye

Received June 14, 2024; revised September 30, 2024; accepted January 13, 2025.

ABSTRACT. *Aspect-Based Sentiment Analysis (ABSA) is a significant task in the field of natural language processing, and the existing ABSA models not only lack the mining of sentiment knowledge, but also ignore the semantic feature representation of text. To address the above issues, this article designs an ABSA method relied on knowledge enhancement and Pre-trained Language Model (PLM). Firstly, a syntactic dependency graph is constructed based on the sentiment knowledge of the text to integrate the syntactic and sentiment information in the sentence. On this basis, PLM is utilized to encode words and sentences in the text context to enrich the semantic feature representation of the text. Secondly, the dependency graph of the sentence is enhanced by using emotional knowledge and the dependency types between words, and the node features are modeled by graph convolutional network to get the specific aspectual representation enhanced by emotional knowledge; meanwhile, the conceptual graph is enhanced by using conceptual graph for the aspectual word concepts and then fused with the specific aspectual representation to get the aspectual representation enhanced by multi-source knowledge. Then designing a dual interaction attention mechanism to realize the fusion of context and aspect features. Finally, the fused characteristic information is adopted as the input of the fully connected layer, and the softmax function is used to output the final sentiment polarity classification. The experimental outcome indicates that the suggested method outperforms the comparison model on the Restaurant and Twitter datasets, and better experimental outcome are obtained.*

Keywords: Aspect-based sentiment analysis; Knowledge enhancement; PLM; Graph convolutional network; Interactive attention mechanism

1. **Introduction.** As social networks and Internet applications rapidly growing, more and more user reviews appear on major Internet platforms. Usually, the sentiment tendency of these comments is valuable for mining users' preferences and understanding their

behaviors. However, traditional sentiment analysis only judges the overall sentiment polarity of the comment text, which cannot satisfy the users' demand for fine-grained analysis of different aspects of sentiment tendency, therefore, researchers proposed Aspect Based Sentiment Analysis (ABSA) [1, 2]. Unlike traditional coarse-grained emotional analysis, ABSA is a fine-grained emotional analysis task. Since this technique can directly extract the sentiment corresponding to the aspect words that the user cares about, thus helping the user to understand the point of view expressed in the sentence [3, 4], how to discriminate the sentiments of different aspect words present in a sentence has rapidly become a hot research topic in the field of Natural Language Processing (NLP) since the ABSA task was first proposed.

1.1. Related work. Early ABSA methods usually base their polarity judgments on sentiment lexicons and linguistic rules, or build supervised sentiment classifiers based on machine learning. Hardeniya and Borikar [5] proposed a lexicon-based approach to extract the corresponding sentiment polarity for both explicit and implicit aspects. Haselmayer and Jenny [6] optimized the lexicon-based approach by using a tree kernel to identify the association between aspect and viewpoint, and achieved better performance. Rao et al. [7] proposed a combination of sentiment lexicon and SVM to extract aspect-based sentiment polarity, which achieved excellent ABSA performance at that time. Jain et al. [8] joined the enhanced PSO method with SVM categorization method, which effectively avoided the defects of traditional sentiment lexicon that is affected by word order and different contexts.

As deep learning technology developing, neural networks have been applied to the field of NLP in a large amount of applications. Zeng et al. [9] used LSTM [10] to model the context before and after the target aspect respectively. Xu et al. [11] modeled the syntactic and emotional relation among the aspect items in a sentence and the surrounding context words by BiLSTM. Onan [12] used recurrent neural networks to learn emotional intensity words, negation words, and context-sensitive words in sentences, which enhanced the sentiment representation of the text from the perspective of contextual syntax. Li and Chai [13] used statistical methods to extract attribute words of the comment text to construct attribute-level text feature vectors, and used CNN networks for fine-grained sentiment analysis. Rammo and Al-Hamdani [14] proposed using CNN to classify languages from short audio samples, and their methods, such as feature extraction and deep learning frameworks, have potential implications for multi-emotion classification in sentiment analysis tasks.

Recently, considering the importance of syntactic knowledge in ABSA tasks, Graphical convolutional network (GCN) models based on dependency trees have gradually been a research hotspot of ABSA. Wei et al. [15] built an undirected graph based on dependency trees and used GCN to learn contextual representations that contain syntactic knowledge. Huang et al. [16] input syntactic dependency graphs into a GCN to perform syntactic augmented embedding of sentence representations. Liu et al. [17] embed knowledge graph triples into word sequences and correspond multiple targets to text to realize knowledge enhancement in multiple scenarios.

Deep learning-based sentiment analysis approaches for different domains need to be targeted and trained from scratch, which will cost much training time, Pre-trained Language Model (PLM) was suggested in 2018, which deals with exactly the above-mentioned issue. OpenAI suggested GPT [18], which utilizes Transformer [19] as a characteristic extractor. Google suggested the BERT [20] pre-training language model, and the effect was recognized by the industry. The PLM BERT-wwm [21], which is specialized for Chinese, and the model ERNIE [22], which incorporates emotional knowledge, have appeared. Even

Ding [23] and Tan [24] combined PLM with deep learning networks and verified that this method is superior to the traditional deep learning-only or pre-trained language sentiment analysis algorithms.

1.2. Contribution. In the task of ABSA, existing research focuses on mining semantic information and syntactic dependency constraints of comment statements, and fails to comprehensively consider the effects of affective knowledge, conceptual knowledge, and the type of syntactic dependency between words on the accuracy of aspectual sentiment tendency discrimination. To address this issue, an ABSA method based on knowledge enhancement and PLM is proposed. Firstly, a syntactic dependency graph is constructed based on the sentiment knowledge of the text to better extract aspect-oriented sentiment information. On this basis, PLM is utilized to encode words and sentences in the text context to improve the semantic representation of the text. Next, external sentiment knowledge is utilized to weight the adjacency matrix of the GCN to effectively incorporate the sentiment information. Meanwhile, conceptual knowledge is utilized to enrich the aspect representation. Finally, the dual interaction attention method is adopted to realize the feature fusion of context and aspect. The fused features are used as inputs to the fully connected level, and the softmax operation is adopted to output the final sentiment polarity classification.

2. Related theoretical analysis.

2.1. Pre-trained language model. BERT is a PLM based on the Transformer structure, and its key innovation lies in its bi-directional context encoding method [25]. While traditional unidirectional language models only consider the context information on the left or right side of the target vocabulary, BERT is able to consider the context information on both sides of the target vocabulary. Compared to previous techniques such as GPT and ELMo, which employ unidirectional or shallow bi-directional strategies respectively, BERT's bi-directional context encoding approach provides better semantic representation, and BERT captures deep bi-directional contextual representations through unsupervised learning in the pre-trained phase, and then adapts the model for a specific task in the supervised fine-tuning phase, thus achieving excellent performance on a variety of NLP tasks.

2.2. Graphical convolutional network. Although neural network models such as RNN, LSTM, etc. have shown their powerful ability to extract text features in NLP tasks, there are still a large number of unstructured features such as referential and syntactic relations in natural language, which can only be accurately represented by graph structures. The fundamental idea of GCN [26] is to extract the graph information by aggregating the features of each node in the graph structure with the features of its neighboring nodes. GCN extends the convolutional operation from the traditional data to the data of the graph structure. The structure of the GCN is indicated in Figure 1.

In the GCN learning process, not only the feature vector representation of each node is input to the model, but also the connection relationship between nodes, i.e., the adjacency matrix. For a stacked multi-level GCN model, the mapping relationship of each level is as bellow.

$$H^{l+1} = f(H^l, A) \quad (1)$$

GCN in the learning process, the neighbor node information of a node can be obtained and aggregated simply by multiplying the adjacency matrix with the node features, as implied below.

$$f(H^l, A) = \delta(AH^lV^l) \quad (2)$$

where V^l is the weight matrix of the l -th level of the GCN and δ is a nonlinear activated operation.

Further, the adjacency matrix A in the above model is normalized by adding the diagonal matrix of node degrees as implied below.

$$f(H^l, A) = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^lV^l\right) \quad (3)$$

where $\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}$ is also known as the Laplace matrix of the graph.

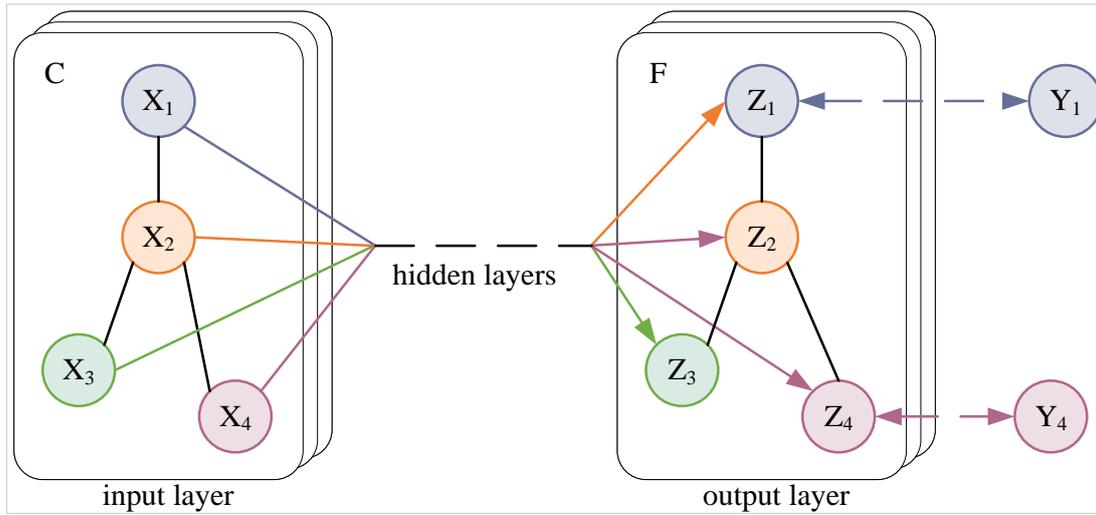


Figure 1. The framework of the GCN model

3. Syntactic dependency graph construction based on affective knowledge. For better ABSA, this chapter constructs a syntactic dependency graph based on the sentiment knowledge of the text. The dependency graph integrates syntactic and sentiment information in sentences to better extract aspect-oriented sentiment information. The syntactic dependencies are enhanced by integrating the syntactic knowledge from the SenticNet lexicon [27] and incorporating the attention mechanism. The steps in detail are as follows:

(1) Construct syntactic dependency graph. Firstly, we use spacy syntactic parser to achieve the syntactic dependency tree of the text, take "The food is very good" as an example, and improve the directed syntactic dependency to an undirected syntactic dependency, and consider the self-cycling of each word, as implied in Figure 2. Relied on the syntactic analysis, the syntactic dependency $D_{ij}^{\text{Syn}} \in R^{n \times n}$ of the text is derived, and the mathematical expression is as follows.

$$D_{ij} = \begin{cases} 1 & w_i, w_j \text{ contains dependency} \\ 0 & \text{others} \end{cases} \quad (4)$$

$$D_{ij}^{\text{Syn}} = D_{ij} + (D_{ij})^T + I \quad (5)$$

where I refers to the unit matrix; and $(D_{ij})^T$ is the transpose matrix of D_{ij} .

(2) Integrate the sentiment information from the SenticNet lexicon to obtain sentiment knowledge and generate a text sentiment dependency graph $A_{ij}^{\text{Sco}} \in R^{n \times n}$, as follows.

$$S_{ij} = \text{Score}(w_i) + \text{Score}(w_j) \quad (6)$$

In addition, to highlight aspect terms in syntactic dependencies, an aspect term adjacency matrix $T_{ij} \in R^{n \times n}$ is constructed by setting the value to 1 when an aspect term exists in the sentiment dependency tree.

$$T_{ij} = \begin{cases} 1 & w_i, w_j \text{ is aspect} \\ 0 & \text{others} \end{cases} \quad (7)$$

The final textual sentiment dependency graph $S_{ij}^{\text{Sco}} \in R^{n \times n}$ is implied below.

$$S_{ij}^{\text{Sco}} = S_{ij} + T_{ij} + 1 \quad (8)$$

(3) Generating textual contextual relevance dependency graphs using textual word embedding and attention mechanisms $A_{ij}^{\text{Att}} \in R^{n \times n}$.

$$A_{ij}^{\text{Att}} = \text{softmax} \left(\frac{h_i w_q \times (h_j w_k)^T}{\sqrt{d}} \right) \quad (9)$$

where h_i and h_j are the embedding representations of the i -th and j -th words of the text S , individually, and w_q and w_k are two parameter matrices that can be trained by the neural network.

(4) Enhance the syntactic dependency of the text using the emotional knowledge and contextual relevance of the text, and integrate the contextual semantic relevance and emotional knowledge of the text into the syntactic dependency of the text to obtain the syntactically enhanced dependency graph $A_{ij}^{\text{SAG}} \in R^{n \times n}$.

$$A_{ij}^{\text{SAG}} = A_{ij}^{\text{Att}} \times S_{ij}^{\text{Sco}} \times D_{ij}^{\text{Syn}} \quad (10)$$

Syntax-enhanced graph convolutional networks are constructed based on the above syntax-enhanced dependency graph $A_{ij}^{\text{SAG}} \in R^{n \times n}$ to output text syntactic features for subsequent aspect-based sentiment analysis tasks.

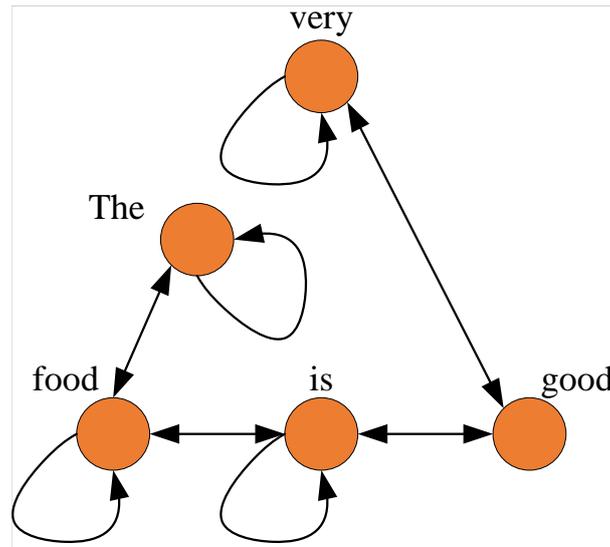


Figure 2. Example diagram of syntactic dependencies

4. Aspect-based sentiment analysis based on knowledge enhancement and PLM.

4.1. PLM-based encoding of contextual information. On the basis of constructing syntactic dependency graphs, this paper proposes an ABSA method based on knowledge augmentation and PLM, which firstly uses PLM to encode words and sentences in the textual context, and secondly, utilizes external sentiment knowledge to weight the adjacency matrix of the graph convolutional network to effectively incorporate the sentiment information. Meanwhile, conceptual knowledge is utilized to enrich the aspect representation. Then, the dual interaction attention mechanism is suggested to realize the fusion of characteristic information of context and aspect. Finally, the softmax operation is used to output the final emotional polarity classification. The whole model of the suggested method is implied in Figure 3.

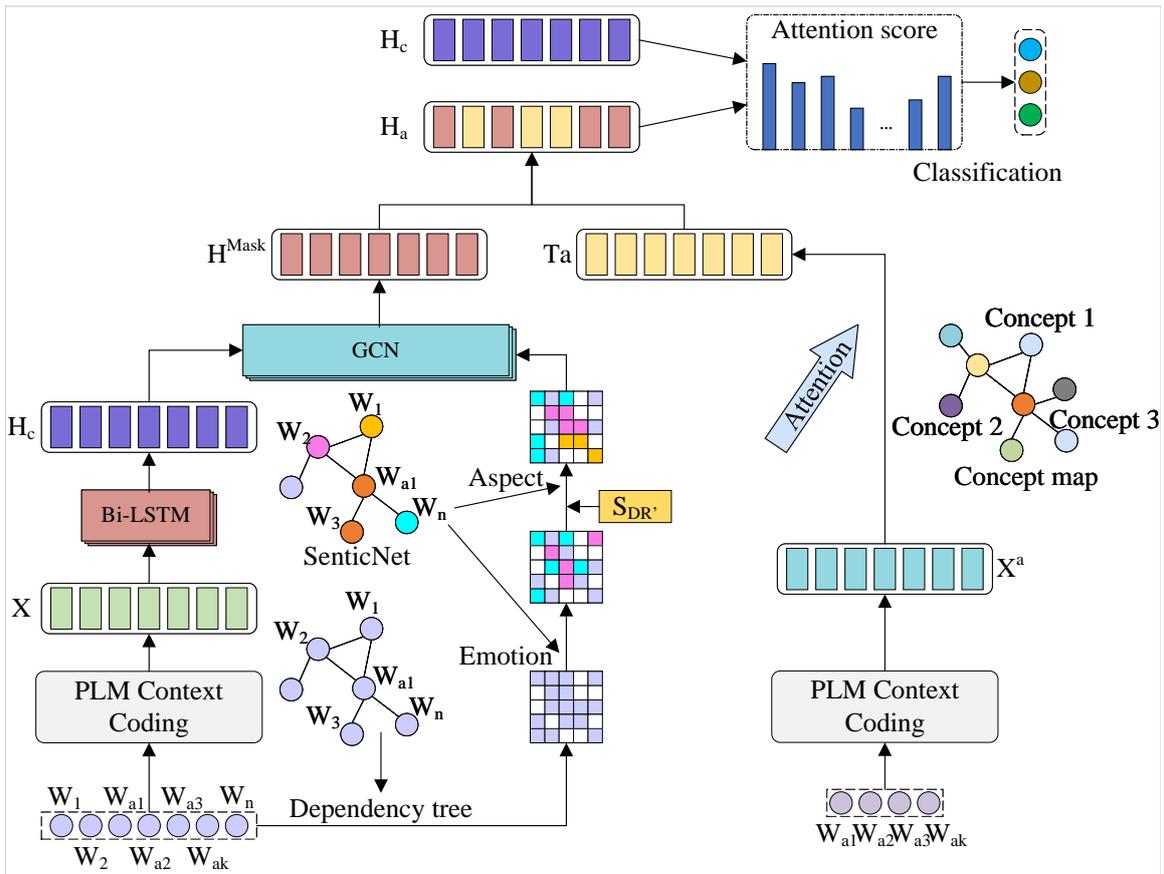


Figure 3. The whole model of the suggested method

Given a word sequence $s = [\varpi_1, \varpi_2, \dots, \varpi_{b_1}, \dots, \varpi_{b_m}, \dots, \varpi_n]$, where $b = [\varpi_{b_1}, \dots, \varpi_{b_m}]$ denotes an aspect containing m words. The text words are embedded into a dimensional feature space c_ϖ using BERT in PLM to obtain the corresponding continuous word vectors $e = [e_1, e_2, \dots, e_{b_1}, \dots, e_{b_m}, \dots, e_n]$, $e_i \in R^{c_\varpi}$. Then the word vectors are input into BiLSTM to obtain the contextualized sentence representation $g^d = [g_1^d, g_2^d, \dots, g_n^d]$.

The obtained vectors and lexical and syntactic dependency graphs are used as inputs to obtain better representations with the help of local hierarchical lexical graphs and hierarchical syntactic graphs, which are further interacted with the context processed by BiLSTM. The local syntactic graph and the lexical graph contain hierarchical structures, so they can be operated by using bi-level GCNs that can aggregate relations.

(1) First, nodes with the same relationship are aggregated to virtual nodes, and then the same sum of normalized hidden features is used as the aggregation function in GCN

to obtain virtual node embeddings, with each relationship type r corresponding to the representation $\tilde{g}_i^{l,r}$. where l is the layer number and i is the target node for aggregation.

(2) Next, all virtual nodes and their relationships are aggregated and the representation of the target node i is updated for different relationship types (virtual nodes) using the average aggregation function.

$$g_i^l = \text{ReLU} \left(V_l \cdot \left(\bigoplus_r \tilde{g}_i^{l,r} \right) \right) \quad (11)$$

where \bigoplus_r denotes the connection corresponding to the representation of different relation types and V_l denotes the weight matrix in the l -th level.

4.2. Aspect information coding based on knowledge enhancement and GCN.

Based on the syntactic dependency graph constructed in Section 3, the SenticNet Sentiment Dictionary is used to analyze the sentiment of the review utterances, and then the extended sensitivity relations are incorporated into the GCN to achieve sentiment enhancement; secondly, the conceptual map is used to enhance the a priori conceptual knowledge of the aspect words.

(1) Emotional knowledge enhancement. First, the neighbor matrix is obtained on the syntactic dependency graph T_{ij} . In this article, the adjacency matrix is constructed by considering the dependencies of the words themselves, i.e., $T_{ii} = T_{jj} = 1$.

On this basis, to fully utilize the emotional information among context words and aspect words, the SenticNet emotional lexicon is used with the help of SenticNet, where $\text{SenticNet}(\varpi_i) \in [-1, 1]$ denotes the sentiment score of word ϖ_i . When $\text{SenticNet}(\varpi_i) = 0$ represents that ϖ_i is neutral or does not exist in SenticNet. The emotional dependency strength of neighboring nodes denoted as S_{ij} , as shown in Equation (6).

In addition, considering whether ϖ_i, ϖ_j are aspect words or not, the aspect-enhanced matrix Z_{ij} is constructed, $Z_{ij} = 1$ when it contains aspect words, and $Z_{ij} = 0$ otherwise, to obtain the sentiment-enhanced matrix B_{ij} .

$$B_{ij} = T_{ij} \times (S_{ij} + Z_{ij} + 1) \quad (12)$$

The processed sentiment adjacency matrix B_{ij} is taken as input and every nodes in the l -th GCN level is updated in terms of the obscured representation of its neighborhood.

$$g_i^l = \text{relu}(\tilde{B}_i^{l-1} V^l + d^l) \quad (13)$$

$$h_i^{l-1} = F(g_i^{l-1}) \quad (14)$$

where h_i^{l-1} denotes the input hidden state of node i at level l in the GCN, \tilde{B} is the standardized symmetrical adjacency matrix: $\tilde{B}_i = B_i / (E_i + 1)$, where $E_i = \sum_{j=1}^n B_{i,j}$, V^l are linear transformation weight matrices, d^l is a predetermine term, and $F(*)$ is a location-aware transformation operation that enhances the importance of the context close to the aspect and reduces the noise generated during the process.

The output of the l -th layer GCN is $G^l = \{g_1^l, g_2^l, \dots, g_{b_1}^l, \dots, g_{b_m}^l, \dots, g_n^l\}$.

To emphasize the important characteristics of aspect words, aspect-specific masking [28] is utilized to mask the non-aspect words of the output vectors studied by the ultimate GCN level and to keep the aspect representation stiff: $g_t^l = 0$, where $1 \leq t < b_1, b_m < t \leq n$. The sentiment aspect representation is obtained: $S = \{0, \dots, g_{b_1}, \dots, g_{b_m}, \dots, 0\}$.

(2) Conceptual knowledge enhancement. Aspects are conceptualized using Microsoft Knowledge Graph [29] to obtain the set of aspect concepts $C = (c_1, c_2, \dots, c_w)$, w is the

number of concepts taken, and c_i denotes the i -th concept vector in the set of concepts. Self-attention is computed for the concepts and appropriate concepts are filtered out.

$$\sigma_i = \text{softmax}(w^T \tanh(Vc_i) + d) \quad (15)$$

where σ_i is the attentional weight of the i -th concept vector in the concept set, V is the weight matrix, and d is the bias. The attention mechanism assigns weights to concepts according to their importance. Each concept vector is weighted to obtain the final concept representation: $p = \sum_{i=1}^w \sigma_i c_i$.

Then, the aspect vector b is combined with its corresponding concept vector p to obtain the conceptually enhanced aspect part $K = (k_1, k_2, \dots, k_m)$. The conceptually enhanced aspect K is concatenated with the emotionally enhanced aspect S to obtain the final aspect representation b .

4.3. Feature information fusion based on interactive attention mechanism. The two-part representation is interactively modeled using the interactive attention mechanism [30]. First, the initial representation of the context is obtained by averaging the hidden state vectors:

$$c_{avg} = \sum_{i=1}^n \frac{g_i^l}{n}, \quad b_{avg} = \sum_{i=1}^m \frac{b_i}{m}.$$

(1) Aspect-context attention computation. For the context representation, the χ function is utilized to compute the semantic association of the hidden state vector $g_i^l \in G^l$ of the context with the aspect vector b_{avg} , and the associated attention weights α_i are assigned to each context word g_i^l .

$$\alpha_i = \frac{\exp(\chi(g_i^l, b_{avg}))}{\sum_{j=1}^n \exp(\chi(g_j^l, b_{avg}))} \quad (16)$$

The χ function is calculated in the form shown below.

$$\chi(g_i^l, b_{avg}) = \tanh(g_i^l \cdot V_a \cdot b_{avg}^T + a_b) \quad (17)$$

where V_a and a_b are the weight matrix and the bias, individually, \tanh is a nonlinear operation, and b_{avg}^T is the transpose of b_{avg} , \tanh is a nonlinear function.

(2) Context-aspect attention computation. Similarly, a contextual representation c_{avg} is used to compute the attention vector of the aspect representation.

$$\beta_i = \frac{\exp(\chi(b_i, c_{avg}))}{\sum_{j=1}^m \exp(\chi(b_j, c_{avg}))} \quad (18)$$

After the above calculations, the context and aspect representations are obtained.

$$c_r = \sum_{i=1}^n \alpha_i g_i^l \quad (19)$$

$$b_r = \sum_{i=1}^m \beta_i b_i \quad (20)$$

Connect the aspect representation b_r and the context representation c_r , as the final representation $z = c_r \oplus b_r$.

4.4. Aspect level affective classification. Using the output of the interactive attention level, z , as the input to the fully connected level, a softmax function outputs the final sentiment polarity classification.

$$y = \text{softmax}(Vz + d) \quad (21)$$

where z is the interactive attention layer output, V is the fully connected layer weight matrix, and d is the bias vector.

For the sentiment classification task, cross entropy is adopted as the loss function.

$$L_1 = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y^{ij} \ln \hat{y}^{ij} \quad (22)$$

where N and C are the amount of training instances and the amount of sentiment polarity classes, respectively; y and \hat{y} are the true and predicted sentiment polarity, respectively.

To further enhance the influence of affective and conceptual knowledge, an auxiliary task of judging opinion words was introduced. Using the obscured state G_s as the input of the auxiliary task, the output expression is as follows.

$$y_s = \text{Softmax}(V_s G_s + d_s) \quad (23)$$

where V_s and d_s are the trainable weights and biases, individually; y_s denotes the prediction of whether each word is an opinion word.

For the auxiliary task of judging opinion words, a binary cross-entropy loss function is adopted as bellow.

$$L_2 = \frac{1}{N} \sum_{i=1}^N \left\{ -\frac{1}{T} \sum_{j=1}^T (y_s^{ij} \log_a(\hat{y}_s^{ij}) + (1 - y_s^{ij}) \ln(1 - \hat{y}_s^{ij})) \right\} \quad (24)$$

where y_s and \hat{y}_s are the true and predicted labels, individually; N is the amount of training instances; and T is the length of the comment text sentence.

For the emotional categorization task and the auxiliary task of binary categorization of opinion words are trained at the same time, then the entire loss function is the sum of the loss functions of the 2 sub-tasks, and the L_2 regularization is applied to all trainable parameters, and the total loss function is implied as bellow.

$$L = L_1 + L_2 + \mu \|\Theta\|^2 \quad (25)$$

where μ is the regularization coefficient; Θ is the model trainable parameter.

5. Performance testing and analysis.

5.1. Comparison and analysis of experimental results. To validate the performance of the suggested model KEPLM, this article is trained on 2 publicly available datasets, including Restaurant [31] and Twitter [32] datasets. The datasets are categorized into 3 sentiment polarities, negative, neutral and positive. In the comment text, each sample is further divided into 3 parts, sentences masking aspectual words, aspectual words and corresponding sentiment labels. The specific division of the dataset is implied in Table 1. The model is trained adopting the Adam optimizer with a learning rate of 0.001, regularization factor set to 10^{-5} , batch size set to 64, and the amount of training iterations set to 100.

This article adopts the Accuracy and Macro-F1 to estimate the effectiveness of the method for ABSA tasks. To further evaluate the effectivity of the KEPLM and assess

Table 1. Dataset information

Dataset	Group	Positive	Medium	Negative
Restaurant	Training set	2167	639	812
	Test set	725	192	217
Twitter	Training set	1591	3169	1572
	Test set	174	342	185

the performance in ABSA tasks, the suggested model in this paper is compared with the models in literature [9], literature [11], literature [13] and literature [15] in comparative experiments.

For ease of analysis, the model in literature [9] is denoted as EALSTM, the model in literature [11] is denoted as SALSTM, the model in literature [13] is denoted as FGCNN and the model in literature [15] is denoted as GP-GCN.

This section compares the sentiment categorization performance of the contrast model and the KEPLM on the 2 datasets, the results of which are shown in Table 2, with bold denoting the optimal data.

Table 2. Experimental results of different models

Model	Restaurant		Twitter	
	Accuracy/%	Macro-F1/%	Accuracy/%	Macro-F1/%
EALSTM	73.18	63.07	69.48	66.79
SALSTM	75.37	72.46	72.19	70.88
FGCNN	81.26	84.39	80.28	85.33
GP-GCN	87.58	86.01	84.39	86.35
KEPLM	92.49	91.37	93.23	92.79

As can be seen from Table 2, the overall performance of the KEPLM outperforms all the compared models on the two benchmark datasets. The accuracy of KEPLM on the Restaurant dataset is 19.31%, 17.12%, 11.23%, 4.91% and 3.14% higher than that of the EALSTM, SALSTM, FGCNN and GP-GCN, respectively. The accuracy on Twitter dataset is 23.75%, 21.04%, 12.95%, 8.84% and 5.79% higher than EALSTM, SALSTM, FGCNN and GP-GCN, respectively. Specifically, both EALSTM and SALSTM add attention mechanism to LSTM for sentiment feature enhancement. FGCNN uses static word vectors as a model for text vectorization approach and utilizes CNN for sentiment classification without considering sentiment knowledge enhancement. GP-GCN is an ABSA model based on GCN, which performs better compared to the models using LSTM, RNN and CNN models, it performs better, but GP-GCN does not dynamically generate word vectors based on context, which results in worse classification performance than KEPLM. The KEPLM model uses syntactic dependency trees to establish the syntactic dependency relationship between aspect words and context, effectively incorporating a priori knowledge to enhance the feature representation and solving the problem that the attention mechanism is unable to obtain syntactic information. In addition, KEPLM uses PLM’s powerful linguistic characterization capability to directly obtain the low-dimensional feature representation of the text, which makes the model better than the one that uses static word vectors as the text vectorization method.

Meanwhile, to compare the effect of KEPLM with the other four models more intuitively, Figure 4 and Figure 5 show the change of Macro-F1 value and the convergence of

the loss function of each model with the number of training steps on the dataset Restaurant, respectively. All models fluctuate when the number of training steps is around 10, and when the training reaches 20, the model effect begins to level off, and the KEPLM model is always the model with the fastest increase in Macro-F1 value, the smoothest and the most optimal model, which also proves the effectiveness of the model in this paper. Meanwhile, the decreasing trend of Loss also shows that the model in this paper has the fastest convergence speed and the best convergence effect, which further proves that the designed KEPLM model is very competitive.

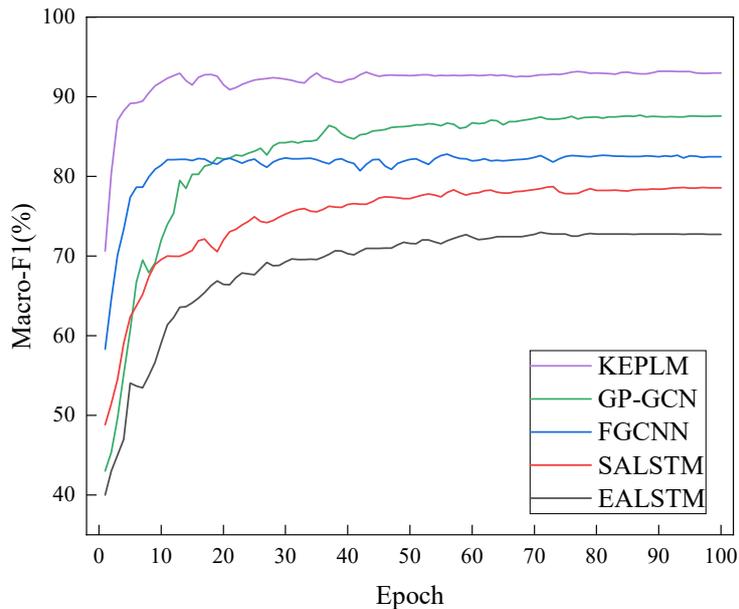


Figure 4. Variation of Macro-F1 values on the dataset Restaurant

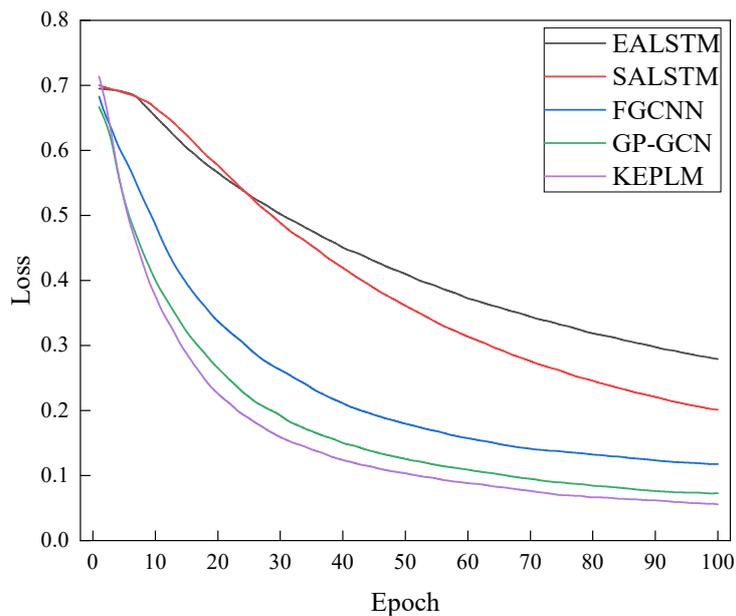


Figure 5. Convergence of the loss function on the dataset Restaurant

5.2. Ablation experiment. To further examine the effect of every component of the KEPLM on the performance, an ablation study of the KEPLM is conducted in this article and the outcome are implied in Table 3. The ablation experiments for each component are as bellow.

(1) KEPLM-PLM: Remove the lexical and syntactic coding part of the PLM coding module for contextual information from the model.

(2) KEPLM-EM: Remove the emotional knowledge enhancement part of the model.

(3) KEPLM-CO: Remove the conceptual knowledge enhancement part of the model.

(4) KEPLM-IN: Remove the feature information fusion module in the model, and splice the two parts of the representation to do the final classification.

Table 3. Ablation results of KEPLM model

Model	Restaurant		Twitter	
	Accuracy/%	Macro-F1/%	Accuracy/%	Macro-F1/%
KEPLM	92.49	91.37	93.23	92.79
KEPLM-PLM	78.16	80.04	81.57	80.37
KEPLM-EM	76.32	78.39	75.61	76.92
KEPLM-CO	78.68	80.95	80.83	78.51
KEPLM-IN	85.29	88.14	87.13	90.66

It can be seen that KEPLM-IN with the interactive attention mechanism removed outperforms the other models and the gap with the full model is relatively small. This suggests that linguistic and external knowledge profoundly affects the whole sentiment analysis task, and it is important to utilize this knowledge if it can be fully exploited. Comparing the KEPLM-EM and KEPLM-CO, which remove affective and conceptual knowledge respectively, the performance on the two datasets is not different, which shows that affective knowledge and conceptual knowledge are equally important and should not be neglected. Meanwhile, KEPLM-PLM with PLM linguistic knowledge removed has a large decrease in the Restaurant, Twitter dataset compared to KEPLM-IN, which indicates that although armed with emotional and conceptual knowledge, context has more important semantic information, and it is indispensable to utilize PLM for linguistic knowledge training. The KEPLM designed in this paper incorporates multilevel knowledge to enhance the model’s classification ability.

6. Conclusion. ABSA task is a fundamental work in natural language processing. This article suggests an ABSA based on knowledge augmentation and PLM to address the issue that current ABSA research fails to comprehensively consider the effects of affective knowledge, conceptual knowledge, and syntactic dependency types between words on the accuracy of aspectual sentiment tendency discrimination. Firstly, a syntactic dependency graph is constructed based on the affective knowledge of the text. Based on this, PLM is utilized to encode words and sentences to enhance the contextual representation. Next, external sentiment knowledge is utilized to weight the adjacency matrix of the GCN to effectively incorporate sentiment information. Meanwhile, conceptual knowledge is utilized to enrich the aspect representation. Finally, the dual interaction attention mechanism is designed to realize the feature fusion of context and aspect. Finally, the fused feature information is adopted as the input of the fully connected layer, and the final sentiment polarity classification is output by softmax function. Validation experiments are conducted on the public datasets Restaurant and Twitter, respectively, and the accuracies reach 92.49% and 93.23%, indicating that the suggested model has strong applicability and competitiveness in the ABSA task.

Acknowledgment. This work is funded by the Teaching Quality and Teaching Reform Project of Guangdong Province in 2022 (No: 2022JXGG01); School-level Quality Engineering Construction Project of Guangzhou College of Commerce in 2024 (No: 2024ZLGC11); Higher Education Research Project of 2024 in the 14th Five-Year Plan of Guangdong Higher Education Association (No: 24GYB75); Teaching Quality Management Professional Committee of Private Colleges and Universities of Guangdong Province Teaching Management Association 2024 Project (No: GDZLGL2428); "Artificial Intelligence+" Pilot Course Project "Data Collection Technology" (No: 2024rgznsdkc06). The authors are grateful to the the Guangzhou College of Commerce in China, Management and Science University in Malaysia, and the University of Mosul in Iraq for their support.

REFERENCES

- [1] H. H. Do, P. W. Prasad, A. Maag, and A. Alsadoon, "Deep learning for aspect-based sentiment analysis: a comparative review," *Expert Systems with Applications*, vol. 118, pp. 272–299, 2019.
- [2] A. Nazir, Y. Rao, L. Wu, and L. Sun, "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 845–863, 2020.
- [3] K. Schouten and F. Frasinca, "Survey on aspect-level sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 813–830, 2015.
- [4] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731–5780, 2022.
- [5] T. Hardeniya and D. A. Borikar, "Dictionary based approach to sentiment analysis—a review," *International Journal of Advanced Engineering, Management and Science*, vol. 2, no. 5, pp. 239438, 2016.
- [6] M. Haselmayer and M. Jenny, "Sentiment analysis of political communication: Combining a dictionary approach with crowd coding," *Quality & Quantity*, vol. 51, pp. 2623–2646, 2017.
- [7] Y. Rao, J. Lei, L. Wenyin, Q. Li, and M. Chen, "Building emotional dictionary for sentiment analysis of online news," *World Wide Web*, vol. 17, pp. 723–742, 2014.
- [8] A. Jain, B. Pal Nandi, C. Gupta, and D. K. Tayal, "Senti-NSetPSO: large-sized document-level sentiment analysis using Neutrosophic Set and particle swarm optimization," *Soft Computing*, vol. 24, no. 1, pp. 3–15, 2020.
- [9] J. Zeng, X. Ma, and K. Zhou, "Enhancing attention-based LSTM with position context for aspect-level sentiment classification," *IEEE Access*, vol. 7, pp. 20462–20471, 2019.
- [10] K. Smagulova and A. P. James, "A survey on LSTM memristive neural network architectures and applications," *The European Physical Journal Special Topics*, vol. 228, no. 10, pp. 2313–2324, 2019.
- [11] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on BiLSTM," *IEEE Access*, vol. 7, pp. 51522–51532, 2019.
- [12] A. Onan, "Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification," *Journal of King Saud University—Computer and Information Sciences*, vol. 34, no. 5, pp. 2098–2117, 2022.
- [13] H. Li and Y. Chai, "Fine-Grained Sentiment Analysis Based on Convolutional Neural Network," *Data Analysis and Knowledge Discovery*, vol. 3, no. 1, pp. 95–103, 2019.
- [14] F. M. Rammo and M. N. Al-Hamdani, "Detecting the speaker language using CNN deep learning algorithm," *Iraqi Journal for Computer Science and Mathematics*, vol. 3, no. 1, pp. 43–52, 2022.
- [15] S. Wei, G. Zhu, Z. Sun, X. Li, and T. Weng, "GP-GCN: Global features of orthogonal projection and local dependency fused graph convolutional networks for aspect-level sentiment classification," *Connection Science*, vol. 34, no. 1, pp. 1785–1806, 2022.
- [16] B. Huang, J. Zhang, J. Ju, R. Guo, H. Fujita, and J. Liu, "CRF-GCN: An effective syntactic dependency model for aspect-level sentiment analysis," *Knowledge-Based Systems*, vol. 260, pp. 110125, 2023.
- [17] H. Liu, X. Li, W. Lu, K. Cheng, and X. Liu, "Graph Augmentation Networks Based on Dynamic Sentiment Knowledge and Static External Knowledge Graphs for aspect-based sentiment analysis," *Expert Systems with Applications*, vol. 251, pp. 123981, 2024.
- [18] L. Floridi and M. Chiriatti, "GPT-3: Its nature, scope, limits, and consequences," *Minds and Machines*, vol. 30, pp. 681–694, 2020.

- [19] K. Han et al., “A survey on vision transformer,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 87–110, 2022.
- [20] P. Ganesh et al., “Compressing large-scale transformer-based models: A case study on bert,” *Transactions of the Association for Computational Linguistics*, vol. 9, pp. 1061–1080, 2021.
- [21] X. Xu, Y. Chang, J. An, and Y. Du, “Chinese text classification by combining Chinese-BERTology-wm and GCN,” *PeerJ Computer Science*, vol. 9, e1544, 2023.
- [22] X. Zhang and Y. Ma, “An ALBERT-based TextCNN-Hatt hybrid model enhanced with topic knowledge for sentiment analysis of sudden-onset disasters,” *Engineering Applications of Artificial Intelligence*, vol. 123, 106136, 2023.
- [23] N. Ding et al., “Parameter-efficient fine-tuning of large-scale pre-trained language models,” *Nature Machine Intelligence*, vol. 5, no. 3, pp. 220–235, 2023.
- [24] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, “RoBERTa-LSTM: a hybrid model for sentiment analysis with transformer and recurrent neural network,” *IEEE Access*, vol. 10, pp. 21517–21525, 2022.
- [25] X. Dong, A. T. Luu, M. Lin, S. Yan, and H. Zhang, “How should pre-trained language models be fine-tuned towards adversarial robustness?,” in *Advances in Neural Information Processing Systems*, vol. 34, pp. 4356–4369, 2021.
- [26] T.-Y. Wu, X. Guo, Y.-C. Chen, S. Kumari, and C.-M. Chen, “SGXAP: SGX-Based Authentication Protocol in IoV-Enabled Fog Computing,” *Symmetry*, vol. 14, no. 7, p. 1393, 2022.
- [27] T.-Y. Wu, X. Guo, L. Yang, Q. Meng, and C.-M. Chen, “A Lightweight Authenticated Key Agreement Protocol Using Fog Nodes in Social Internet of Vehicles,” *Mobile Information Systems*, vol. 2021, pp. 1–14, 2021.
- [28] A. Feng, X. Zhang, and X. Song, “Unrestricted attention may not be all you need—masked attention mechanism focuses better on relevant parts in aspect-based sentiment analysis,” *IEEE Access*, vol. 10, pp. 8518–8528, 2022.
- [29] L. Ji, Y. Wang, B. Shi, D. Zhang, Z. Wang, and J. Yan, “Microsoft concept graph: Mining semantic concepts for short text understanding,” *Data Intelligence*, vol. 1, no. 3, pp. 238–270, 2019.
- [30] M.-H. Guo et al., “Attention mechanisms in computer vision: A survey,” *Computational Visual Media*, vol. 8, no. 3, pp. 331–368, 2022.
- [31] A. Schmidt, M. W. U. Kabir, and M. T. Hoque, “Machine learning based restaurant sales forecasting,” *Machine Learning and Knowledge Extraction*, vol. 4, no. 1, pp. 105–130, 2022.
- [32] D. Effrosynidis, A. I. Karasakalidis, G. Sylaios, and A. Arampatzis, “The climate change Twitter dataset,” *Expert Systems with Applications*, vol. 204, 117541, 2022.