Information Processing and Management 60 (2023) 103223

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Knowledge-guided multi-granularity GCN for ABSA

Zhenfang Zhu [a], Dianyuan Zhang [a,*], Lin Li [b], Kefeng Li [a], Jiangtao Qi [a], Wenling Wang [c], Guangyuan Zhang [a], Peiyu Liu [d]

[a] School of Information Science and Electrical Engineering, Shandong Jiao Tong University, Jinan 250357, China
[b] School of Computer and Artificial Intelligence, Wuhan University of Technology, Wuhan 430070, China
[c] Chinese Lexicography Research Center, Lu Dong University, Yantai 264025, China
[d] School of Information Science and Engineering, Shandong Normal University, Jinan 250358, China

ARTICLE INFO

Keywords:
Sentiment analysis
Graph neural network
Conceptual knowledge
Robustness analysis
Attention mechanism

ABSTRACT

Aspect-based sentiment analysis aims to determine sentiment polarities toward specific aspect terms within the same sentence or document. Most recent studies adopted attention-based neural network models to implicitly connect aspect terms with context words. However, these studies were limited by insufficient interaction between aspect terms and opinion words, leading to poor performance on robustness test sets. In addition, we have found that robustness test sets create new sentences that interfere with the original information of a sentence, which often makes the text too long and leads to the problem of long-distance dependence. Simultaneously, these new sentences produce more non-target aspect terms, misleading the model because of the lack of relevant knowledge guidance. This study proposes a knowledge guided multi-granularity graph convolutional neural network (KMGCN) to solve these problems. The multi-granularity attention mechanism is designed to enhance the interaction between aspect terms and opinion words. To address the long-distance dependence, KMGCN uses a graph convolutional network that relies on a semantic map based on fine-tuning pre-trained models. In particular, KMGCN uses a mask mechanism guided by conceptual knowledge to encounter more aspect terms (including target and non-target aspect terms). Experiments are conducted on 12 SemEval-2014 variant benchmarking datasets, and the results demonstrated the effectiveness of the proposed framework.

1. Introduction

Aspect-based sentiment analysis (ABSA) (Pontiki et al., 2016; Schouten & Frasincar, 2015) is a subtask of text sentiment analysis that, differs from the traditional sentiment analysis of documents or sentences. ABSA aims to summarize the sentiment polarities of users toward specific aspect terms in a sentence. For example, in the sentence “Great food but the service was dreadful!”, the sentiment polarities for the aspect terms “food” and “service” are “positive” and “negative” respectively. Because the two aspect terms in this example express completely opposite sentiment polarities, assigning a sentence-level sentiment polarity to them is unreasonable. In this regard, compared with sentence-level or document-level sentiment analysis, ABSA can provide better insight into user reviews.

In addition, using held-out datasets that are often not comprehensive tends to result in trained models that contain the same biases as the training data, which causes ABSA to encounter more severe challenges. The text robustness assessment method focuses

* Corresponding author.
E-mail addresses: zhuzf@sdjtu.edu.cn (Z. Zhu), csdzdy@163.com (D. Zhang), cathylilin@whut.edu.cn (L. Li), 205073@sdjtu.edu.cn (K. Li), qijiangtaohappy@hotmail.com (J. Qi), 362607793@qq.com (W. Wang), xdzhanggy@163.com (G. Zhang), liupy@sdnu.edu.cn (P. Liu).

https://doi.org/10.1016/j.ipm.2022.103223
Received 12 June 2022; Received in revised form 28 November 2022; Accepted 3 December 2022
Available online 19 December 2022
0306-4573/

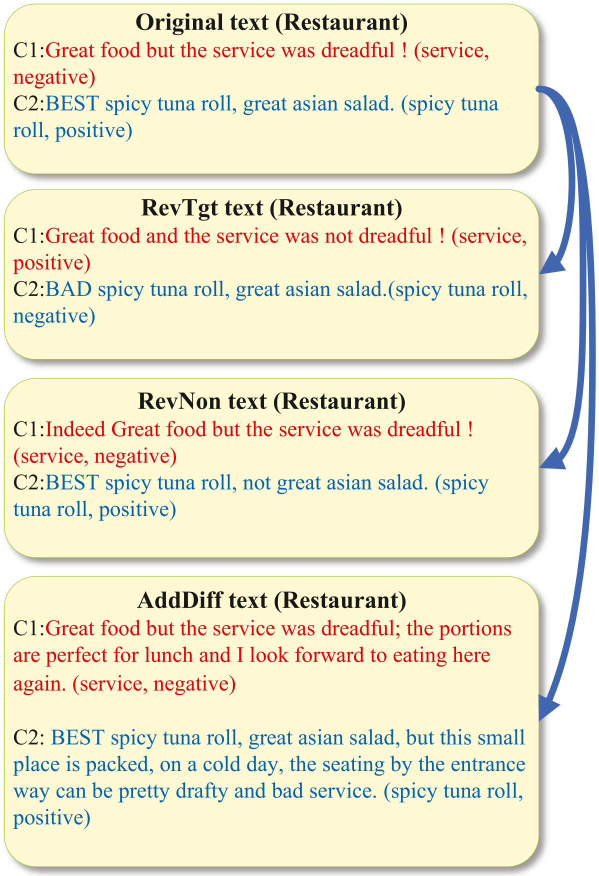

Fig. 1. Examples from robustness test datasets.

on slightly modifying the input and results in different predictions. Studies on robustness (Xing et al., 2020) have detected whether a neural network model has learned real semantic information from training text. The robustness test can be divided into three categories. As shown in Fig. 1, we provide two robustness test examples from the SemEval-2014 benchmarking dataset. In Fig. 1, "Original text" represents the test sample from the original restaurant review dataset, and the three boxes below represent the conversion of the three types of data. "RevTgt text" represents the sentiment of reversing the aspect term. In C1, the sentiment polarity of the service is positive and after RevTgt, the sentiment polarity of the service becomes negative. "RevNon" is the reverse of the sentiment of the non-target aspect terms that are originally the same sentiment as the target. "AddDiff" is the addition of aspect terms with the opposite sentiment from the target aspect. These robustness test datasets are derived by Gui et al. (2021).

We identify three problems that remain unsolved in current state-of-the-art methods for improving the robustness of ABSA. One problem is that these methods do not actually learn the interactive information of aspect terms and context words in the semantic space, they perform extremely poorly on the test set. Poria, Cambria, and Gelbukh (2016) used convolutional neural networks (CNN) to capture the connection between aspect terms and context words. Based on these studies. Tang, Qin, and Liu (2016) adopt multiple layers of long short-term memory (LSTM) with memory cells to capture the hidden state of context words. Xue and Li (2018) proposed an easily parallelized model based on CNNs and gating mechanisms, that selectively captures the connection between aspect terms and context words. In addition, recurrent neural network (RNN) augmented by attention mechanism (He, Lee, Ng, & Dahlmeier, 2018; Ma, Li, Zhang, & Wang, 2017) have been widely used to capture the association information between context words and aspect terms. However, these methods only fused aspect information and context information, and did not fully use their link information in the semantic space. The lack of this semantic interaction significantly limits the performance of these methods on the robustness test dataset. This problem exists not only in ABSA, but also in other natural language processing tasks such as question answer task and summary task. This is an urgent problem in natural language processing. Therefore, we designed a multi-granularity attention mechanism to enhance the interaction between different words. The multi-granularity attention mechanism enhances the interaction between different words by combining coarse and fine grained attention, particularly on the robustness test dataset, because we artificially add some noise elements to the robustness test dataset. The multi-granularity attention mechanism dynamically simulates the relationship between aspect terms (target and non-target aspect terms) and opinion words by increasing the interactive information between different words. This helps the model pay more attention to the target aspect terms and corresponding opinion words, ignoring the sentiment information of secondary non-target aspect terms, to maintain the stability of the subject sentiment orientation of the semantic space, thereby improving the robustness of the model.

Another problem is that as sentences increase in length and the distance between words increases, the model's attention to some important words weakens. This is a common problem in the field of natural language processing. Traditional neural networks rely on serialized processing methods; therefore, long-distance dependence is a common defect. Graph neural networks (GNNs) eventually broke the deadlock, as they treats each word as an independent node in the graph, and then capture their hidden states through relationships between different words. GNNs (Wu et al., 2020; Zhou et al., 2020) use each context word as a node to obtain the global information of the context, have made breakthroughs in addressing the problem. Zhao, Hou, and Wu (2020) introduced position encoding for graph convolutional networks. Wang, Shen, Yang, Quan, and Wang (2020) proposed a relational graph attention network to encode a data structure for sentiment prediction. GNN has achieved good results with short and medium length text, but performs poorly when processing long sequences of text. Therefore, we propose using a combination of a fine-tuned pre-training model and GCN (Wang et al., 2019) to solve this problem. We first obtain domain applicable pre-trained language model by pre-training on domain-specific datasets. Then we break the traditional sequence structure and rely only on the semantic information of the input sequence to build a semantic graph. Finally we use a GCN to obtain the hidden state of the text. This method completely relies on the semantic information of the text to construct a new graph data structure method, which abandons the limitation of the sequence structure of the text, and provides an idea for solving long-distance dependence.

Finally the existing ABSA model lacks guidance from external knowledge. A knowledge graph can effectively introduce external knowledge into a neural network model. The introduction of external knowledge can guide the model to better understand the semantic text information. Therefore, incorporating commonsense knowledge into deep learning models has become a popular research topic in the field of natural language processing (NLP), such as in question answering (Dong, Wei, Zhou, & Xu, 2015) and machine reading comprehension (Mihaylov & Frank, 2018). However, few studies have been conducted on improving the robustness of the model using external knowledge. We propose a approach for solving this problem. Many disturbing aspect terms and opinion words were artificially added to the robustness test datasets. This may lead to a confusing problem of opinion words misleading the model, thereby interfering with the model's judgment of the sentiment polarity of target aspect terms. This method can also provide a higher decision-making status for target aspect terms. For example, the sentence "great burgers, grilled cheeses and french fries, but worst sushi and service is slow". have seven aspect terms in the 15 words; however, we only require the model to recognize the sentiment polarity of the target aspect terms. The opinion word "slow" of the non-target aspect term "service" that does not belong to the food category obviously has a very low correlation with the aspect terms of the food category. Therefore, it has less interference with the affective polarity of the target aspect term. However, if the model ignores the non-target aspect term "sushi", and "french fries" is near the word "worst" in both semantic space and sequence structure, the model's judgment of the sentiment polarity of "french fries" may be inaccurate. We import conceptual knowledge as a hypernym of "french fries", which can be combined with the food-related words "burgers, grilled cheeses, sushi". At the same time, the non-target aspect terms and their corresponding opinion words are given lower priority, thereby providing a basis for forming a semantic space dominated by the target aspect terms concept.

In this study, we first fine-tuned a pre-trained model on the restaurant and laptop datasets to obtain pre-trained features with higher domain-related features from these datasets. Then we built domain-related semantic maps based on the fine-tuned pre-training model, and used the semantic maps as the input of the GCN. For aspect terms, we first introduced external concept knowledge, and then used a multi-granularity attention mechanism to enhance the interaction between aspect terms and context words. The main contributions of this paper are summarized as follows:

• We propose a multi-granularity attention mechanism, that enhances the interaction between aspect terms and context words, and helps the model focus on the core target aspect terms and their opinion words.

• To solve the problem of long-distance dependence, we propose a GCN based on a fine-tuned pre-training model.

• To avoid disturbing the model by ignoring artificially added non-target aspect terms and their opinion words, we introduce external conceptual knowledge into the model.

• We conducted experiments using 12 robustness test datasets. The results demonstrate that our network outperforms state-of-the-art methods on these datasets.

2. Related studies

2.1. Aspect-based sentiment analysis

Compared with traditional sentiment analysis task, ABSA is a more fine-grained sentiment analysis task. ABSA can be used to conduct a more standardized and rationalized analysis of users' online comments. In this section, we will review studies related to ABSA.

Deep neural networks can generate dense word vectors without manual features. Hence, they have attracted increasing attention. Tang et al. (2016) proposed multiple layers in an LSTM, and each layer is an attention model that adds additional memory units to each layer to capture connections. Huang and Carley (2018) proposed using parameterized filters and gates and incorporated context information into a convolutional neural network. Then the attention mechanism, which has been successfully applied to machine translation tasks (Bahdanau, Cho, & Bengio, 2015), was used to filter important context information according to certain rules. Gu, Zhang, Hou, and Song (2018) used a bidirectional attention mechanism to model context words and aspect terms, while incorporating position attention. Zhang, Li, and Song (2019b) used a weighted convolutional network to provide aspect-specific syntactically aware context representations. Song, Wang, Jiang, Liu, and Rao (2019) proposed an attentional encoder network, that uses attention-based encoders to model context words and aspect terms. Zhang et al. (2020) proposed using a bidirectional long short-term memory network (BI-LSTM) and capsule attention network to parse sentence structures.

2.2. Graph convolutional networks

Although many studies have been conducted on neural networks based on context sequences, their long-distance dependence problem is still not well solved. Until the emergence of graph neural networks broke this deadlock. A GNN treats each word as an independent node in the graph, and then captures their hidden state through the relationship between different words. Zhang, Li, and Song (2019a) first proposed the use of GCN for aspect-based sentiment analysis. This model first introduced a multi-grained attention mechanism with position encoding to model aspect-specific representations. A GCN is then used to capture the sentiment dependencies between different aspect terms in a sentence. Zhao et al. (2020) first introduced a bidirectional attention mechanism with position encoding to model aspect-specific representations, and then used a GCN to capture the sentiment dependencies between different aspect terms in a sentence. Wang et al. (2020) first built a dependency tree, and then proposed a relational graph attention network (R-GAT) to encode a new tree structure for sentiment prediction. Tian, Chen, and Song (2021) proposed explicitly using dependency types for aspect-based sentiment analysis with type-aware GCN. However, few studies have applied to apply the hybrid architecture model to aspect-based sentiment analysis task. In this study, we propose a combination of an internal-attention graph network and multi-granularity attention for joint learning, which provides a new idea for solving the long-distance problem.

2.3. Pre-trained language models

In recent years, pre-trained language models have been widely used in many NLP tasks. Typically, language models are first trained using large unlabeled corpora, and then applied to downstream tasks. Devlin, Chang, Lee, and Toutanova (2019) successfully achieved state-of-the-art results for 11 NLP tasks. Song et al. (2019) applied pre-trained bidirectional encoder representation from transformers (BERT) to aspect-based sentiment analysis and achieved good results. And since then more researchers have used BERT for aspect-based sentiment analysis.

2.4. Knowledge graph

Knowledge graph describing common sense and facts have become a widely used knowledge representation method in academia and industry, they (Lin, Liu, Sun, Liu, & Zhu, 2015; Wang, Mao, Wang, & Guo, 2017) have received much attention because they can introduce external knowledge to help model inference. Knowledge graphs have achieved good results in the field of question answering (Ravichandran & Hovy, 2002) and recommendation systems (Davidson et al., 2010). Wang, Zhang, Xie, and Guo (2018) introduced external knowledge into a recommendation system and discovered potential knowledge-level connections between news. Sheu, Chu, Qi, and Li (2021) employed external knowledge graphs to improve the semantic-level representations of news articles in recommendation systems. Shen et al. (2018) leveraged external knowledge from knowledge graphs to enrich the representational learning of question answering sentences. Lv et al. (2020) proposed the automatic extraction of evidence from heterogeneous knowledge sources, and answered questions based on the extracted evidence.

Furthermore, the robustness of neural networks, as an emerging problem, has a significant impact on the application of sentiment analysis technology. Gui et al. (2021) proposed datasets for the robustness testing for aspect-based sentiment analysis task. Ma, Zhang, and Song (2021) improved the robustness of the model by introducing position bias. However, few studies have been conducted on aspect-based sentiment analysis task. In this study, we proposes a knowledge guided multi-granularity graph convolutional neural network (KMGCN), which provides new ideas for improving the robustness of the model from different perspectives.

3. Proposed methodology

The aspect-based sentiment analysis task aims to predict the sentiment polarity of aspect terms. Suppose given a context sequence $T^c = \{w_1^c, w_2^c, \dots, w_n^c\}$ and a aspect terms sequence $T^t = \{w_1^t, w_2^t, \dots, w_m^t\}$, where T^t is a sub-sequence of T^c. T^c and T^t are respectively composed of n context words and m aspect term words. The goal of this task is to predict the sentiment polarity of all terms contained in T^t through the elements in T^c. For this task we propose a knowledge-guided multi-granularity graph convolutional neural network based on fine-tuning pre-training model, and its overall structure is illustrated in Fig. 2. KMGCN comprises five major components. The first is the embedding layer, used to obtain a more domain-relevant feature representation by the fine-tuning pre-trained model, and is used to obtain context and aspect feature sequences that incorporate conceptual knowledge. The second component is the graph-hidden layer, which solves the problem of long-distance dependence using semantic graphs based on fine-tuning pre-trained models and GCN, which are based on DGL (Wang et al., 2019). The third component is multi-granularity layer, which is used to capture the hidden state between context words and aspect terms. Finally, the fourth component is the sentiment prediction layer, which is used to predict the sentiment polarities of aspect terms. Next, we elaborate on these five major parts.

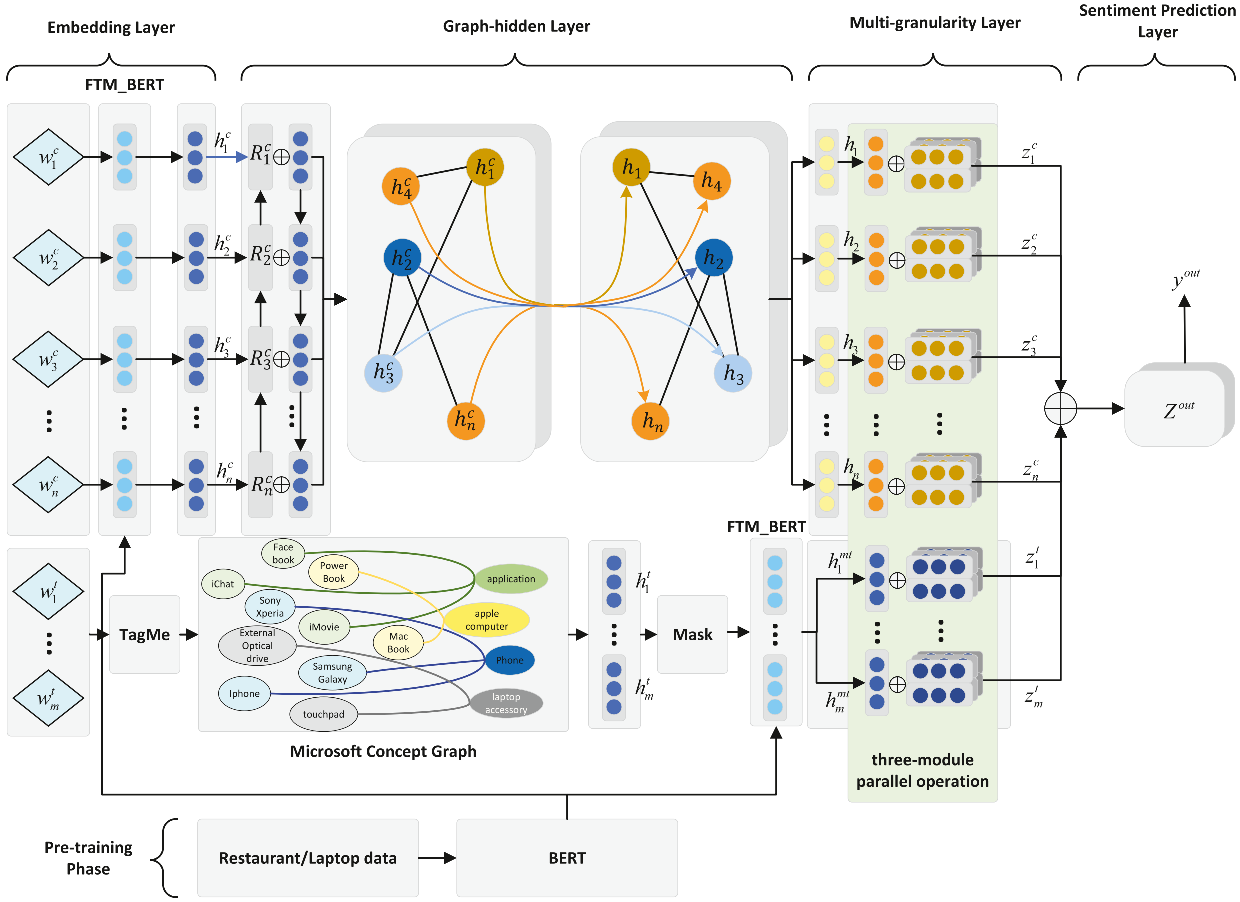

Fig. 2. Overall architecture of KMGCN.

3.1. Embedding layer

3.1.1. Pre-training phase

BERT take transformers as backbone architecture. They can be replaced by the following equation:

$$h^{\hat{(l)}} = \mathrm{LN}(h^{l-1} + \mathrm{MHAtt}(h^{l-1})) \tag{1}$$

$$h^{l-1} = \mathrm{LN}(h^{\hat{(l)}} + \mathrm{FNN}(h^{\hat{(l)}})) \tag{2}$$

where h^0 is the input representation, formed by the sum of token embeddings, position embeddings, and segment embeddings. LN is the layer normalization layer, MHAtt is the multi-head self-attention, FNN contains two linear projection layer and an activation layer. l represents the number of layers of transformer. There are two pre-training methods for BERT, one is masked language modeling (MLM) and the other is next sentence prediction (NSP). MLM randomly masks 15% of the words in each sentence and uses their context to make predictions. Because the ratio of mask 15% is already very high, this will cause certain words to never appear in the fine-tuning stage, in order to solve this problem. 80% of the data uses [mask], “my dog is hairy” → “my dog is [MASK]”; 10% of the data randomly select a word to replace the [mask], “my dog is hairy” → “my dog is apple”; The other 10% remain unchanged, “my dog is hairy” → “my dog is hairy”. NSP select some sentence pairs and predict a sentence from the previous sentence. Specially, 50% of sentence pairs are logically related and the other 50% are randomly selected. This way can effectively learn the relevance of sentences. However, in a specific field, the models trained on these general corpus cannot fully extract the intrinsic meaning of the token, so it needs to be fine-tuned (Devlin et al., 2019; Gao, Yao, & Chen, 2021) on the restaurant data corpus and the laptop corpus. So, we use restaurant data and notebook data as input of fine-tuning corpus for pre-training model (Gao et al., 2021) respectively. In addition, each layer of BERT captures the different features of the input text. We choose the mean value of the output of the last two layers as the final output result. Finally we obtain pre-training weights that are more suitable for these specific fields.

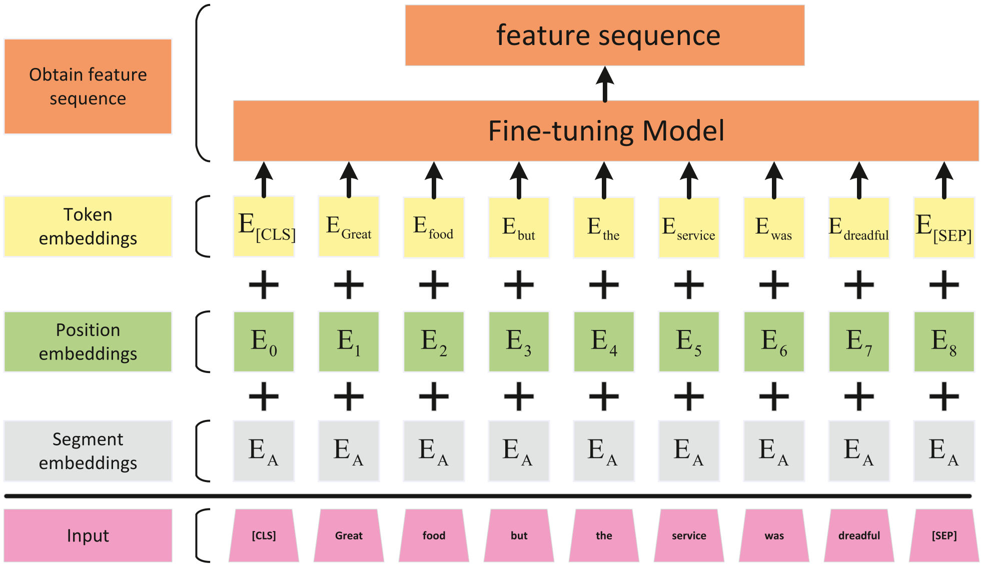

Fig. 3. The process of obtaining context feature sequence from "[CLS] Great food but the service was dreadful [SEP]".

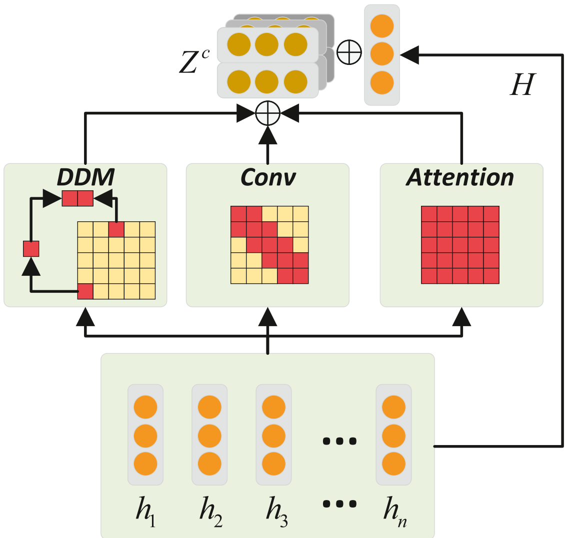

Fig. 4. Overall architecture of multi-granularity layer.

3.1.2. Context embedding

Given a context sequence $w^c = \{w_1^c, w_2^c, \dots, w_n^c\}$ which is composed of n context words, and a aspect terms sequence $w^t = \{w_1^t, w_2^t, \dots, w_m^t\}$ which is composed of m aspect term words, Specially w^t is a subsequence of w^c. In this layer we use the fine-tuned BERT to obtain the context feature sequence $h^c = \{h_1^c, h_2^c, \dots, h_n^c\}$ and aspect terms feature sequence $h^{mt} = \{h_1^{mt}, h_2^{mt}, \dots, h_m^{mt}\}$.

The h^c conversion of input is expressed as follows:

$$h^c = \mathrm{FTM\_BERT}(w^c) \tag{3}$$

FTM_BERT represent fine-tuning model BERT. In addition, to understand this process more intuitively, we show an example with "[CLS] Great food but the service was dreadful [SEP]" as input in Fig. 3.

3.1.3. Aspect terms embedding

For the aspect terms sequence, we first map the different aspect terms to different entities through TagMe (Ferragina & Scaiella, 2010, 2011), which is an entity link tool. We map the retained entities to different entity concepts in Microsoft concept graph, and then we obtain the concepts corresponding to the entities and their mapping probabilities. Microsoft concept map provides 6 algorithms for calculating the correlation between entities and concepts. We use $P(c|e)$ to calculate the degree of association between entities and concepts. $P(c|e)$ represents the probability that the concept corresponding to an entity e is c. The specific calculation process is as follows:

$$P(c|e) = \frac{n(c,e)}{\sum_{e \in c_j} n(c_j,e)} \tag{4}$$

where $n(c,e)$ represents the number of times that entity e and concept c appear at the same time. $\sum_{e \in c_j} n(c_j,e)$ represents the sum of the simultaneous occurrences of all concepts to which entity e and entity e belong. We select the concept that is most closely related to the entities to obtain the entity concept squeeze $w^{con} = \left\{c_1^{con}, c_2^{con}, \dots, c_l^{con}\right\}$, and we use the entities sequence as a bridge to correspond the concept sequence to the aspect terms sequence. Then we randomly mask the sequence of aspect terms by external conceptual knowledge in the training process inspired by bert, and obtain the knowledge-guided aspect terms mask sequence h^m.

$$w^{mt} = \mathrm{MASK}(w^t) \tag{5}$$

Finally we use the aspect terms sequence h^m as the input of FTM-BERT to obtain knowledge-guided aspect terms feature sequence h^{mt}.

$$h^{mt} = \mathrm{FTM\_BERT}(w^{mt}) \tag{6}$$

3.2. Graph-hidden layer

In this layer, we use a variant of the graph convolutional network to capture context features based on global semantic connections. The construction of our graph-hidden layer depends on DGL package (Wang et al., 2019). It further provides a global semantic analysis basis for solving long-distance dependency. At the same time, we consider that due to the large amount of parameters of the downstream model, the knowledge contained in the pre-training features may have a certain loss. Therefore, in order to make full use of the knowledge of the pre-training model, We propose to use the idea of k-nearest neighbor (KNN) to transform the original data into graph structured data which also can be called a semantic graph. Finally, we input the graph structured data to the GCN for obtaining the hidden state of the context feature sequence. Specially, we introduce an internal attention mechanism for GCN. Next, we use the rest of this part to introduce graph-hidden layer in details.

We use the output of embedding layer h^c as the input of graph-hidden layer. We calculate the degree of association between each element in the context feature sequence, and record the degree of association between the ith element and the jth element as r_j^i. Then we group the two nodes with the highest correlation to h_i^c into a class (connect these three points with the ith node as the center).

$$r_{m1}^i, r_{m2}^i = argmax\left\{r_1^i, r_2^i, \dots, r_n^i\right\} \tag{7}$$

We build graph structured data G based on above calculations. In the semantic graph G, $\left\{h_{m1}^c, h_{m2}^c, r_{m1}^i, r_{m2}^i\right\}$ is the semantic information connected to the ith node G_i. h_{m1}^c and h_{m2}^c are the hidden states of the two nodes adjacent to the ith node; r_{m1}^i and r_{m2}^i respectively represent the degree of association between these two adjacent nodes and the i node. Then, we use the graph convolutional neural network with dot-product internal-attention to update the hidden state of each node. The update process of each node is as follows

$$h_i^{(l+1)} = ReLU(\sum_{j \in \mathcal{N}(i)} \alpha_{i,j} h_j^{(l)}) \tag{8}$$

where $\mathcal{N}(i)$ is the set of its one-hop neighbor (which contains two nodes in the context), $h_j^{(l)}$ represents the jth node hidden state in the lth layer of the graph neural network. $\alpha_{i,j}$ is the attention score between node i and node j, and the calculation process of $\alpha_{i,j}$ is as follows:

$$\alpha_{i,j} = \mathrm{softmax}\left\{e_{ij}^{(l)}\right\} \tag{9}$$

$$e_{ij}^{(l)} = r_j^i (W_i^{(l)} h_i^{(l)})^T \cdot W_j^{(l)} h_j^{(l)} \tag{10}$$

where $W_i^{(l)}$ and $W_j^{(l)}$ transform ith node and jth node features into the same dimension, This makes it convenient to introduce dot-product when calculating these features similarity. The lth graph neural network layer is represented in matrix form as follows:

$$H^{(l)} = ReLU(\sum_{j \in \mathcal{N}(i)} A^{(l)} H^{(l)}) \tag{11}$$

where $[A^{(l)}]_{i,j} = \alpha_{i,j}$, l is the number of layers of the graph convolutional network. $H^{(0)}$ is the output of embedding layer, and the output of graph-hidden layer is $H^{(l)} = \left\{h_1, h_2, \dots, h_n\right\}$.

3.3. Multi-granularity layer

In this layer, we adopt a three-module parallel operation method based on Zhao, Sun, Xu, Zhang, and Luo (2019) to further improve the interaction between aspect terms and context words, and its overall structure is illustrated in Fig. 4. Self-attention module capture local features, separable convolution module come from Chollet (2017), Wu, Fan, Baevski, Dauphin, and Auli (2019), which can capture separable convolution features, dynamic decision module capture multi-word aspect term features, In particular, dynamic decision module can also help the model to deal with the problem of multiple words in one aspect term. Then we will introduce the details of these three modules.

3.3.1. Self-attention module

Self-attention module is responsible for learning learn local features from input squeeze. We first project the input sequence into three representations in a way similar to multi-head attention, Key (K), query (Q), and value (V). Then we add a self-attention mechanism based on these three representations. The calculation process is as follow:

$$\mathrm{Attention(H)} = \sigma(W^K K, W^Q Q, W^V V)W \tag{12}$$

$$K, Q, V = \pounds^K(H), \pounds^Q(H), \pounds^V(H) \tag{13}$$

where W, W^K, W^Q and W^V are linear variable weight matrices, $\pounds^K$, $\pounds^Q$ and $\pounds^V$ are three linear transformation functions. σ is the dot-production function, its calculation process is as follows:

$$\sigma(K_i, Q_i, V_i) = softmax(\frac{Q_i K_i^\top}{\sqrt{d_k}})V_i \tag{14}$$

3.3.2. Separable convolution module

Separable convolution module capture separable convolution features for hidden state containing deep semantic information. It includes two independent transformations, namely point projection transformation and information transformation. $Depth\_Conv$ refers to depth convolution in the work of Wu et al. (2019), which can share the same point-wise projecting transformation with self-attention mechanism, and separable convolution module can share the same point-wise projecting transformation with self-attention mechanism. The sub-module of separable convolution module contains multiple cells with different core sizes. They are used to capture features in different ranges. The calculation process of the kernel size k is as follows:

$$\mathrm{Conv(V)} = \mathrm{Dep\_conv}(W^V V)W^{conv} \tag{15}$$

where W^V and W^{conv} are learnable parameters. Its detailed calculation process is as follows:

$$\mathrm{Dep\_conv(V)} = \sum_{j=1}^{k} softmax((\sum_{c=1}^{d} W^{j\_c} V^{j\_c})V_{i+j-\frac{k+1}{2}}) \tag{16}$$

Then we introduced a gating mechanism to automatically select the weights of different convolution units, where α_i and α_j represent different attention scores. The calculation process is as follows:

$$\mathrm{Ds\_Conv(V)} = \sum_{i=1}^{n} \frac{exp(\alpha_i)}{\sum_{j=1}^{n} exp(\alpha_j)} Conv(V) \tag{17}$$

3.3.3. Dynamic decision module

Dynamic decision module gives equal status to each word feature in the input sequence. It is called "dynamic decision" because it dynamically changes the subject word of the decision each time. At the same time, dynamic decision module use each word as the subject to select the feature with the deepest semantic relevance. Then we will introduce this module in detail, the process is as follows:

$$r_i^t = \sum_{j=1}^{n} h_j \cdot \mathcal{F}\left(h_i, h_j\right) \tag{18}$$

$$\mathcal{F}\left(h_i, h_j\right) = \frac{exp\left(h_i^\top \cdot h_j\right)}{\sum_{k=1}^{n} exp\left(h_i^\top \cdot h_k\right)} \tag{19}$$

where r_i^t is the tailored word-level feature, and the function $\mathcal{F}$ represents the correlation between h_i and h_j as follows:

$$h_i^{out} = r_i^t \cdot h_i + h_i \tag{20}$$

We input h^{out} into the $ReLU$ function to obtain the output of dynamic decision module. Where w_{out} and b_{out} are learnable parameters. The process is as follows:

$$\mathrm{DDM(H)} = z^{DMM} = ReLU\left(w_{out}^\top h^{out} + b_{out}\right) \tag{21}$$

Table 1
Detailed description of SemEval 2014 Task 4 datasets and its 12 variant datasets.

Dataset	Train	Test
Res	3608	1120
Res_RevTgt	3608	838
Res_RevNon	3608	582
Res_AddDiff	3608	847
Res_RevTgt_RevNon	3608	572
Res_RevTgt_AddDiff	3608	837
Res_RevNon_AddDiff	3608	572
Lap	2328	638
Lap_RevTgt	2328	466
Lap_RevNon	2328	239
Lap_AddDiff	2328	466
Lap_RevTgt_RevNon	2328	238
Lap_RevTgt_AddDiff	2328	464
Lap_RevNon_AddDiff	2328	238

At the end of this section, we give the overall process of multi-granularity layer. we input the output of embedding layer h^{mt} and the output of graph-hidden layer H into multi-granularity layer, and we give the collaborative working process of these three modules:

$$Z^{c} = H + \text{Attention}(H) + \text{Conv}(H) + \text{DDM}(H) \tag{22}$$

$$Z^{t} = h^{mt} + \text{Attention}(h^{mt}) + \text{Conv}(h^{mt}) + \text{DDM}(h^{mt}) \tag{23}$$

where Z^{c} and Z^{t} are based on the output of the context words feature and aspect term words feature respectively. $Attention$ is self-attention module, $Conv$ is separable convolution module and DDM is dynamic decision module. Finally we combine the context-based hidden state and the aspect-based hidden state:

$$Z^{out} = Z^{c} + Z^{t} \tag{24}$$

3.4. Sentiment prediction layer

In this layer, we will predict the sentiment polarity of the hidden state Z^{out} output by multi-granularity layer. The process is as follows:

$$y^{out} = softmax(w_{se}^{\top} z^{out} + b_{se}) \tag{25}$$

where $w_{se}^{\top}$ and b_{se} are learnable parameters, y^{out} stands for sentiment sequence. Via this approach, the sentiment polarity of aspect terms can be accurately predicted.

4. Experiment

4.1. Dataset and experimental settings

To demonstrate the effectiveness of KMGCN, we conduct experiments on 12 variant benchmarking datasets (Gui et al., 2021; Xing et al., 2020) from SemEval 2014 Task 4. The dataset is described in Table 1. Specially, “Res” and “Lap” respectively represent the dataset of restaurant reviews and laptop reviews, the dataset with “Res” are variant datasets of the restaurant dataset. For example, “Res_RevTgt” is reverse the sentiment of the target aspect terms. “RevNon” is reverse the non-target aspect terms sentiment that is the same as the target aspect terms sentiment. “AddDiff” is add aspect terms opposite to the target aspect terms sentiment. In addition, We only use the datasets from SemEval 2014 Task 4 dataset in the training phase. For model performance comparison, we select the accuracy and macro-F1 scores as evaluation metrics. Embedding dimension is 768 for fine-tuning pre-trained BERT (Devlin et al., 2019; Gao et al., 2021). For our method, the learning rate is set to 2e-5, the dropout rate is 0.2, the number of attention heads of self-attention module in multi-granularity layer is 8.

4.2. Comparison models

To evaluate KMGCN along with its components in aspect-based sentiment analysis , we consider multiple baseline models. Most of the experiment results of the robustness test are derived from Gui et al. (2021).

Robustness Test for Baseline Models:

Baselines based on sequence neural network:

Table 2
The main experiment results of 6 variants from the restaurant dataset and laptopdataset.

Restaurant	RevTgt		RevNon		AddDiff		RevTgt_RevNon		RevTgt_AddDiff		RevNon_AddDiff		Average result	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
TNET	21.37	23.21	77.84	52.81	79.46	56.76	20.63	19.41	17.44	19.80	77.20	38.74	48.99	35.12
MGAN	26.33	26.50	73.20	48.96	70.96	52.11	23.08	21.24	14.58	20.47	49.30	34.01	42.91	33.88
BERT-BASE	38.72	31.99	53.09	37.06	55.37	38.27	31.82	24.14	7.17	8.01	16.61	17.10	33.79	26.09
BERT-ASPECT	64.70	48.17	56.19	44.4	81.23	69.97	66.43	44.55	55.56	45.29	48.08	47.89	62.03	50.04
LCF-BERT	54.78	41.81	61.17	44.05	85.83	72.67	54.20	40.25	44.68	36.90	55.07	45.98	59.28	46.94
LCFS-BERT	57.02	41.20	63.23	45.19	79.57	56.21	64.16	35.26	39.65	31.13	71.67	34.73	62.55	40.62
Sentic GCN-BERT	65.17	53.30	69.42	48.88	88.31	77.82	64.69	42.19	58.25	48.69	67.60	48.59	68.91	53.25
KMGCN	69.32	54.26	69.76	53.87	89.85	75.58	66.08	51.20	63.08	53.83	65.41	51.56	70.58	56.72

Laptop	RevTgt		RevNon		AddDiff		RevTgt_RevNon		RevTgt_AddDiff		RevNon_AddDiff		Average result	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
TNET	26.82	26.39	65.27	49.74	75.32	58.35	26.47	24.79	20.22	20.91	67.65	49.45	46.96	38.27
MGAN	28.54	28.59	64.44	48.92	72.53	52.60	23.95	24.39	20.86	22.48	72.27	53.44	47.07	38.41
BERT-BASE	46.14	39.25	43.51	35.87	70.82	43.55	35.71	26.36	22.80	14.75	67.23	34.05	47.70	32.31
BERT-ASPECT	54.08	48.19	51.05	41.01	72.96	49.66	45.38	35.67	27.10	21.72	68.07	40.24	53.11	39.42
LCF-BERT	48.93	44.28	56.90	48.93	77.90	63.23	42.02	37.29	31.83	32.76	60.92	41.89	53.08	44.73
LCFS-BERT	61.37	49.26	61.50	45.91	78.75	61.89	55.46	27.72	43.11	40.48	61.37	49.26	60.26	45.75
Sentic GCN-BERT	58.15	45.28	59.00	46.54	80.04	65.11	56.30	36.60	47.84	41.10	70.46	38.21	61.97	45.47
KMGCN	62.45	45.56	62.27	52.09	80.26	60.70	58.82	39.11	55.39	31.86	71.43	52.26	65.10	46.93

• **TNET** (Li, Bing, Lam, & Shi, 2018) employs a CNN layer to extract salient features, and propose a component to generate target-specific representations of words in the sentence, meanwhile incorporate a mechanism for preserving the original contextual information.

• **MGAN** (Fan, Feng, & Zhao, 2018) combines coarse-grained and fine-grained analysis, location encoding was introduced and the influence of location information on the task was considered. In addition, the aspect alignment loss was used to describe interactions between aspects that have the same context.

Baseline models based on pre-trained model:

• **BERT-BASE** (Devlin et al., 2019) uses the basic version of BERT for aspect-based sentiment analysis.

• **BERT-ASPECT** (Devlin et al., 2019; Li, Zou, Zhang, Zhang, & Wei, 2021) incorporates aspect terms information on the basis of BERT-BASE.

Baselines based on attention network:

• **LCF-BERT** (Zeng, Yang, Xu, Zhou, & Han, 2019) uses a local context focus(LCF) mechanism based on multi-head self-attention for aspect-based sentiment analysis. This mechanism is called LCF design, and utilizes the context features dynamic mask and context features dynamic weighted layers to pay more attention to the local context words.

• **LCFS-BERT** (Phan & Ogunbona, 2020) combines part-of-speech embeddings, dependency based embeddings and contextualized embeddings to enhance the performance of the aspect extractor. LCFS-BERT also propose syntactic relative distance to downplay the adverse effects of irrelevant words.

• **Sentic GCN-BERT** (Liang, Su, Gui, Cambria, & Xu, 2022) constructs the graph neural networks via integrating the affective knowledge from SenticNet to enhance the dependency graphs of sentences.

4.3. Main analysis

Table 2 present the performances of baseline models and KMGCN, and these results demonstrate that KMGCN achieved a new state-of-the-art performance on all datasets.

KMGCN significantly outperformed TNET on almost all datasets. In particular on the RevTgt dataset KMGCN realized Acc values that were 47.95% and 35.63% higher, and a breakthrough in the F1 value was observed. Compared with MGAN, we obtained a huge advantage in all indicators of the 9 datasets. On the RevNon dataset MGAN outperformed KMGCN in Acc values, and outperformed KMGCN in overall performance on the RevNon_AddDiff laptop dataset. We think this is due to MGAN's combination of coarse-grained and fine-grained attention to capture the interactive information between aspect terms and contexts. KMGCN mainly considers word level interaction, putting coarse-grained information at a secondary level. But the overall performance of KMGCN is better than MGAN. Therefore, we believe that obtaining more complete fine-grained semantic information and to a certain extent ignoring coarse-grained information is a reasonable strategy for KMGCN. Although this strategy reduces Acc value on some datasets, it improves the overall performance of the model. In addition, KMGCN is superior to TNET and MGAN for another reason that our pre-training model is superior to TNET and MGAN. Our pre-trained model can learn more accurate representations of different words in the semantic space from a large amount of text.

Compared with other bert-based models, we have observed that BERT-based models generally outperform the models considered above. Next, we compared KMGCN with two bert-based models (BERT-BASE and BERT-ASPECT). KMGCN outperformed BERT-BASE on the 12 datasets. In particular in AddDiff, KMGCM achieved Acc and F1 scores that were 34.48% and 37.31%, respectively, higher

Table 3
The ablation study results of 6 variants from the restaurant dataset and laptopdataset.

Restaurant dataset	RevTgt		RevNon		AddDiff		RevTgt_RevNon		RevTgt_AddDiff		RevNon_AddDiff		Average result	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
KMGCN +transformer	70.72	51.91	70.62	52.44	87.72	74.32	67.83	48.30	62.96	49.13	63.29	48.85	70.52	54.16
KMGAN +BERT	69.78	51.39	70.27	52.76	89.02	76.75	65.03	46.71	67.03	51.96	64.69	48.95	70.97	54.75
KMGCN w/o self attention	67.77	56.06	58.93	47.07	88.78	76.56	67.83	50.16	61.65	53.94	62.41	48.73	67.73	55.42
KMGCN w/o dynamic decision	68.71	55.66	67.18	48.16	88.67	75.66	66.43	47.59	64.76	47.60	63.64	52.64	69.89	54.55
KMGCN w/o knowledge	68.60	54.30	68.90	53.86	88.16	75.41	65.73	51.28	62.13	49.12	64.86	51.47	69.73	55.91
KMGCN	69.32	54.26	69.76	53.87	89.85	75.58	66.08	51.20	63.08	53.83	65.41	51.56	70.58	56.72

Laptop dataset	RevTgt		RevNon		AddDiff		RevTgt_RevNon		RevTgt_AddDiff		RevNon_AddDiff		Average result	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
KMGCN +transformer	57.30	43.84	62.76	46.44	78.76	65.16	57.14	38.76	55.05	30.42	66.81	51.67	62.97	46.05
KMGAN +BERT	58.80	45.03	61.09	46.31	79.61	62.64	52.94	39.07	52.47	40.87	70.17	48.22	62.51	47.02
KMGCN w/o self attention	59.66	45.24	56.07	49.21	77.25	59.51	57.56	39.14	52.04	30.29	69.75	50.09	62.06	45.58
KMGCN w/o dynamic decision	61.80	44.40	58.16	45.27	77.04	58.77	57.98	43.57	57.85	31.70	68.49	49.49	63.55	45.53
KMGCN w/o knowledge	59.44	46.19	62.76	49.02	78.97	60.56	57.56	42.11	54.41	29.44	69.33	50.94	63.74	46.37
KMGCN	62.45	45.56	62.27	52.09	80.26	60.70	58.82	39.11	55.39	31.86	71.43	52.26	65.10	46.93

than BERT-BASE on the restaurant datasets. We believe that this is because of the addition of text information, which results in the context sentence being too long, and that BERT-BASE cannot accurately eliminate interference noise in the sentence, thereby ignoring some important aspect terms of information. BERT-ASPECT compensates for this shortcoming by introducing additional aspect term information in the sentence. This method of directly introducing additional aspect terms of information effectively improved the capabilities of the model. However, it has shortcomings in capturing the deep semantics of the context. For example, RevNon creates noise by reversing the sentiment of non-target aspect terms, and then uses this noise to detect whether the model has constructed an effective semantic space. Both BERT-BASE and BERT-ASPECT achieved poor results on this dataset, and KMGCN surpassed BERT-BASE and BERT-ASPECT for all indicators. This may be because BERT-BASE and BERT-ASPECT over-rely on the semantic spatial distribution in the training dataset, leading to poor model robustness.

KMGCN outperformed LCF-BERT, compared with LCF-BERT and we realized breakthroughs in Acc and F1 on all datasets. LCF-BERT uses a local context focus mechanism (LCF) based on multi-head self-attention to capture the correlation between the aspect term sentiment polarity and local context. Compared with BERT-BASE and BERT-ASPECT, LCF better uses multi-head attention (Vaswani et al., 2017) to capture the connection between the aspect terms in the sentence and context, However a significant gap remains between LCF-BERT and KMGCN. There remains a serious problem in encountering new sequences. The performance of KMGCN is far better than LCF-BERT which may be because of LCF-BERT's inability to accurately assign the correct weight to an additional aspect term in the test text. LCFS-BERT explores the grammatical aspect terms of a sentence and uses a self-attention mechanism for syntactical learning. Our method interrupts the original sentence sequence and constructs graph structured data using the semantic connections between different context words. In addition KMGCN outperformed LCFS-BERT on almost all datasets. This may be because the original syntactic structure of the original test dataset is disrupted during the robustness test, and the syntactic information learned by the model may interfere with the judgment of sentiment polarity. The graph structured data of KMGCN completely relies on the connection in the semantic space of words. Fundamentally this avoid the problem of noise generated by syntactic information. Compared to Sentic GCN-BERT, KMGCN outperforms Sentic GCN-BERT on almost all metrics. Similar to KMGCN, Sentic GCN-BERT also uses external knowledge and GCN, which we believe benefits from the combination of our conceptual knowledge encoding and multiple attention. In addition, we also show the average results for all models on the variant datasets in Table 2. This further illustrates that KMGCN achieves new state-of-the-art performance on these 12 datasets.

4.4. Ablation study

The ablation study is conducted mainly to evaluate the influences of various components of the network on the overall performance of the network. Table 3 presents the results of our ablation experiments on KMGCN.

KMGCN+transformer refers to using transformer (Dai, Yan, Sun, Liu, & Qiu, 2021) to replace GCN, KMGCN significantly outperformed KMGCN+transformer on almost all datasets. We believe this is because GCN is more suitable for our textual semantic graph structured data. KMGCN+BERT refers to using the original BERT. We can see that although it outperforms KMGCN on some datasets, but the overall performance is slightly worse than KMCGN. KMGCN w/o self decision and KMGCN w/o dynamic decision refers to the attention mechanism that removes the multi-granularity layer. KMGCN w/o knowledge refers to remove external knowledge. The experimental results demonstrate that the cooperative work of these three components is critical to the model.

4.5. Effectiveness of node connections

To evaluate the impact of data structure on overall network performance, we conduct detailed experiments in this section. As shown in Fig. 5, original sequence structure represents the original input data; syntactic structure represents that the original data

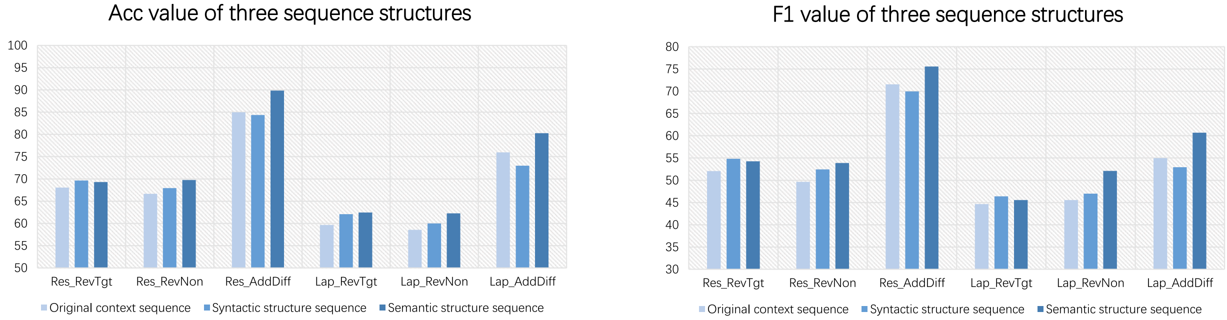

Fig. 5. The impact of different data structures on the overall model.

is constructed as a syntactic dependency tree and the tree structure data is used as input; semantic structure sequence represents the graph structured data constructed based on the semantic connection between words proposed in graph-hidden layer. We can see that when using the same network structure, the sequence-based data structure performance is the worst, the performance of the model based on the syntactic structure outperformed the original sequence structure in most cases. This is because there is no change in the data dependency structure, and syntactic dependency tree built by the model can more effectively extract the hidden state related to the aspect terms. However the performance on the syntactic structure dataset was slightly worse than that on AddDiff, this is because the syntactic analysis will produce more interference information when the sentence structure changes. Simultaneously, semantic structure sequence's performance is significantly outperformed the other two structure sequence. This is because KMGCN is not limited to a specific sequence by introducing external knowledge. When the dataset changes, it will update the hidden state of different words in the semantic space. This way of dynamic linking enhances the robustness of the model. So as to help KMGCN understand the context from the perspective of deep semantics and enhance the sentiment representation based on aspect terms.

4.6. Influence of GCN layers

As shown in Fig. 6, we find that the best number of graph neural network for laptop dataset and restaurant dataset. Although when the number of layers is 1 and 3, the performance outperformed 2 on the laptop dataset. The overall performance of model 2 GCN layers is still the best. An intuitive explanation is that interactive information between different words would not be fully broadcasted when the iteration number is too small. Therefore, with 1 GCN layer the effect of the model was worse than that of 2 GCN layers. The model has over-fitting and redundancy in information transmission, and too many layers of graph neural network will cause performance degradation. Therefore, the overall performance of the model shows a downward trend as the number of layers increases.

4.7. Case study

Fig. 7 visualizes the attention score on the sentence and aspect terms, along with predictions on these examples and the corresponding ground truth labels. MGAN and BERT-ASPECT predict the first example fail, In the second example, we can observe that the above two models mistakenly establish a connection between "french fries" and "worst", and give "worst" a high degree of attention. As a result, their predictions are wrong. LCF-BERT predicted the third example wrong. We can clearly see that LCF-BERT pays more attention to the secondary word "battery" when facing "battery life". Due to this wrong attention, the model may confuse the relationship between different words in aspect terms and sentiment words. It also predicted the first sentence wrong. We can see that it pays attention to "great" and "worst" at the same time, resulting in a conflict in judging the sentiment polarity of the target aspect term, so it predict the wrong sentiment polarity of the target aspect term. KMGCN eliminates these interfering elements by introducing external knowledge and joint learning of different components, and finally successfully predicted the sentiment polarity of these examples.

5. Conclusion and future work

In this study, we propose a knowledge-guided multi-granularity graph convolutional neural (KMGCN) network for aspect-based sentiment analysis, which integrates knowledge via multi-granularity graph convolutional neural simultaneously. In particular, KMGCN outperform all the state-of-the-art approaches across all the 12 datasets and can capture the semantic information more accurately. However, we think that our model still has some limitations. First of all, when the entities in the text are mapped to the conception knowledge graph, some aspect terms may be mapped to the wrong conceptions. Our paper assumes that the concept words of aspect term mapping are correct, without taking into account the problem of wrong concept words. Secondly, due to the scale of the knowledge graph, some words in the aspect term may not be able to find corresponding concept words.

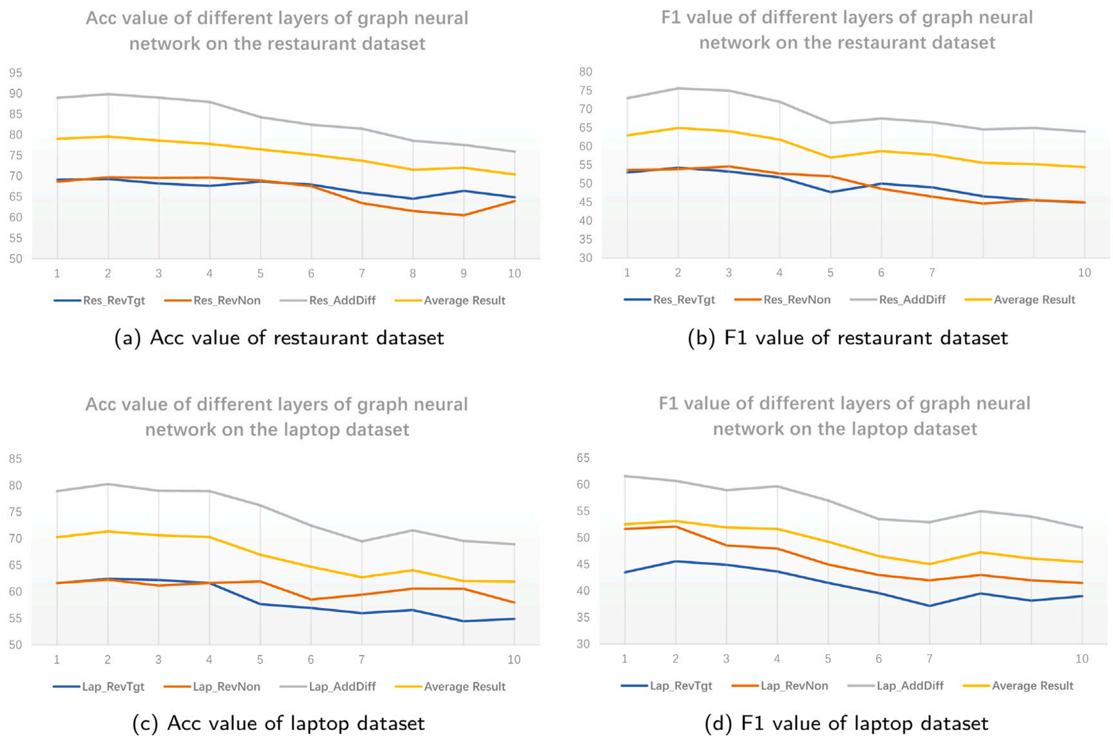

Fig. 6. The effects of GCN with different numbers of layers.

Model	Aspect	Attention visualization	Prediction	Label
MGAN	french fries	Great burgers, grilled cheeses and french fries, but worst sushi and service is slow	Neg ✗	Pos
	battery life	the battery life is not great	Pos ✗	Neg
BERT-ASPECT	french fries	Great burgers, grilled cheeses and french fries, but worst sushi and service is slow	Neg ✗	Pos
	battery life	the battery life is not great	Neg✓	Neg
LCF-BERT	french fries	Great burgers, grilled cheeses and french fries, but worst sushi and service is slow	Neu ✗	Pos
	battery life	the battery life is not great	Pos ✗	Neg
KMGCN	french fries	Great burgers, grilled cheeses and french fries, but worst sushi and service is slow	Pos✓	Pos
	battery life	the battery life is not great	Neg✓	Neg

Fig. 7. Visualization of attention scores for two examples.

In the future, we would like to consider how to balance coarse-grained text information with fine-grained text information, and we would like to work on knowledge graph embedding, so as to acquire more advanced conceptual knowledge, and we would like to explore the effectiveness of knowledge-guided GCN for other tasks, such as emotion-cause extraction and aspect sentiment triplet extraction. In addition, exploring common sense knowledge is valuable for effective modeling in the field of sentiment analysis.

CRediT authorship contribution statement

Zhenfang Zhu: Methodology, Software, Data curation. **Dianyuan Zhang:** Conceptualization, Writing – original draft. **Lin Li:** Visualization. **Kefeng Li:** Supervision. **Jiangtao Qi:** Software. **Wenling Wang:** Validation. **Guangyuan Zhang:** Investigation. **Peiyu Liu:** Writing – review & editing.

Data availability

The data that has been used is confidential.

Acknowledgments

This work was supported in part by National Social Science Foundation(19BYY076); of China (ZR2021MF064, ZR2021QG041); Key R & D project of Shandong Province 2019JZZY010129, Shandong Natural Science Foundation, Shandong Social Science Planning Foundation (19BJCJ51).

References

Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *3rd International conference on learning representations*.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251–1258).

Dai, J., Yan, H., Sun, T., Liu, P., & Qiu, X. (2021). Does syntax matter? A strong baseline for aspect-based sentiment analysis with RoBERTa. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1816–1829).

Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., et al. (2010). The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on recommender systems* (pp. 293–296).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers)* (pp. 4171–4186).

Dong, L., Wei, F., Zhou, M., & Xu, K. (2015). Question answering over freebase with multi-column convolutional neural networks. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)* (pp. 260–269).

Fan, F., Feng, Y., & Zhao, D. (2018). Multi-grained attention network for aspect-level sentiment classification. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 3433–3442).

Ferragina, P., & Scaiella, U. (2010). Tagme: On-the-fly annotation of short text fragments (by Wikipedia entities). In *Proceedings of the 19th ACM international conference on information and knowledge management* (pp. 1625–1628).

Ferragina, P., & Scaiella, U. (2011). Fast and accurate annotation of short texts with Wikipedia pages. *IEEE Software*, *29*(1), 70–75.

Gao, T., Yao, X., & Chen, D. (2021). SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 6894–6910).

Gu, S., Zhang, L., Hou, Y., & Song, Y. (2018). A position-aware bidirectional attention network for aspect-level sentiment analysis. In *Proceedings of the 27th international conference on computational linguistics* (pp. 774–784).

Gui, T., Wang, X., Zhang, Q., Liu, Q., Zou, Y., Zhou, X., et al. (2021). TextFlint: Unified multilingual robustness evaluation toolkit for natural language processing. arXiv e-prints, arXiv–2103.

He, R., Lee, W. S., Ng, H. T., & Dahlmeier, D. (2018). Effective attention modeling for aspect-level sentiment classification. In *Proceedings of the 27th international conference on computational linguistics* (pp. 1121–1131).

Huang, B., & Carley, K. M. (2018). Parameterized convolutional neural networks for aspect level sentiment classification. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 1091–1096).

Li, X., Bing, L., Lam, W., & Shi, B. (2018). Transformation networks for target-oriented sentiment classification. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 946–956).

Li, Z., Zou, Y., Zhang, C., Zhang, Q., & Wei, Z. (2021). Learning implicit sentiment in aspect-based sentiment analysis with supervised contrastive pre-training. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 246–256).

Liang, B., Su, H., Gui, L., Cambria, E., & Xu, R. (2022). Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowl.-Based Syst.*, *235*, 107643.

Lin, Y., Liu, Z., Sun, M., Liu, Y., & Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. In *Twenty-ninth AAAI conference on artificial intelligence*.

Lv, S., Guo, D., Xu, J., Tang, D., Duan, N., Gong, M., et al. (2020). Graph-based reasoning over heterogeneous external knowledge for commonsense question answering. In *Proceedings of the AAAI conference on artificial intelligence*: *vol. 34*, (no. 05), (pp. 8449–8456).

Ma, D., Li, S., Zhang, X., & Wang, H. (2017). Interactive attention networks for aspect-level sentiment classification. In *Proceedings of the 26th international joint conference on artificial intelligence* (pp. 4068–4074).

Ma, F., Zhang, C., & Song, D. (2021). Exploiting position bias for robust aspect sentiment classification. arXiv preprint arXiv:2105.14210.

Mihaylov, T., & Frank, A. (2018). Knowledgeable reader: Enhancing cloze-style reading comprehension with external commonsense knowledge. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 821–832).

Phan, M. H., & Ogunbona, P. O. (2020). Modelling context and syntactical features for aspect-based sentiment analysis. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 3211–3220).

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., et al. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *International workshop on semantic evaluation* (pp. 19–30).

Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, *108*, 42–49.

Ravichandran, D., & Hovy, E. (2002). Learning surface text patterns for a question answering system. In *Proceedings of the 40th annual meeting of the association for computational linguistics* (pp. 41–47).

Schouten, K., & Frasincar, F. (2015). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, *28*(3), 813–830.

Shen, Y., Deng, Y., Yang, M., Li, Y., Du, N., Fan, W., et al. (2018). Knowledge-aware attentive neural network for ranking question answer pairs. In *The 41st international ACM SIGIR conference on research & development in information retrieval* (pp. 901–904).

Sheu, H.-S., Chu, Z., Qi, D., & Li, S. (2021). Knowledge-guided article embedding refinement for session-based news recommendation. *IEEE Transactions on Neural Networks and Learning Systems*.

Song, Y., Wang, J., Jiang, T., Liu, Z., & Rao, Y. (2019). Attentional encoder network for targeted sentiment classification. arXiv preprint arXiv:1902.09314.

Tang, D., Qin, B., & Liu, T. (2016). Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 214–224).

Tian, Y., Chen, G., & Song, Y. (2021). 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* (pp. 2910–2922).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, *29*(12), 2724–2743.

Wang, K., Shen, W., Yang, Y., Quan, X., & Wang, R. (2020). Relational graph attention network for aspect-based sentiment analysis. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 3229–3238).

Wang, H., Zhang, F., Xie, X., & Guo, M. (2018). DKN: Deep knowledge-aware network for news recommendation. In *Proceedings of the 2018 world wide web conference* (pp. 1835–1844).

Wang, M., Zheng, D., Ye, Z., Gan, Q., Li, M., Song, X., et al. (2019). Deep graph library: A graph-centric, highly-performant package for graph neural networks. arXiv preprint arXiv:1909.01315.

Wu, F., Fan, A., Baevski, A., Dauphin, Y. N., & Auli, M. (2019). Pay less attention with lightweight and dynamic convolutions. arXiv preprint arXiv:1901.10430.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(1), 4–24.

Xing, X., Jin, Z., Jin, D., Wang, B., Zhang, Q., & Huang, X. J. (2020). Tasty burgers, soggy fries: Probing aspect robustness in aspect-based sentiment analysis. In *Proceedings of the 2020 conference on empirical methods in natural language processing* (pp. 3594–3605).

Xue, W., & Li, T. (2018). Aspect based sentiment analysis with gated convolutional networks. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 2514–2523).

Zeng, B., Yang, H., Xu, R., Zhou, W., & Han, X. (2019). Lcf: A local context focus mechanism for aspect-based sentiment classification. *Applied Sciences*, *9*(16), 3389.

Zhang, C., Li, Q., & Song, D. (2019a). Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing* (pp. 4568–4578).

Zhang, C., Li, Q., & Song, D. (2019b). Syntax-aware aspect-level sentiment classification with proximity-weighted convolution network. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval* (pp. 1145–1148).

Zhang, B., Li, X., Xu, X., Leung, K.-C., Chen, Z., & Ye, Y. (2020). Knowledge guided capsule attention network for aspect-based sentiment analysis. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *28*, 2538–2551.

Zhao, P., Hou, L., & Wu, O. (2020). Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification. *Knowledge-Based Systems*, *193*, Article 105443.

Zhao, G., Sun, X., Xu, J., Zhang, Z., & Luo, L. (2019). Muse: Parallel multi-scale attention for sequence to sequence learning. arXiv preprint arXiv:1911.09483.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., et al. (2020). Graph neural networks: A review of methods and applications. *AI Open*, *1*, 57–81.