Cross and Self Attention Based Graph Convolutional Network for Aspect-Based Sentiment Analysis

Mao Zhang, Sijie Teng, and Linli Xu[✉]

School of Computer Science and Technology, University of Science and Technology of China, Hefei, China
{zmyyy,yunmo}@mail.ustc.edu.cn,
linlixu@ustc.edu.cn

Abstract. Aspect-based sentiment analysis aims to recognize the sentiment polarity of an aspect in reviews. In general, to analyze the sentiment of an aspect in a sentence, it is essential to capture the dependencies between aspects and the corresponding contexts. Recently, graph neural networks over global dependency structures like dependency trees or self-attention score matrices have been explored for this task. However, these models rely heavily on the quality of information extracted from the global dependency structures. In the meantime, the pairwise correlations between aspects and contexts provide an equally important perspective for sentiment analysis, which is usually ignored in previous works. Motivated by this, we propose a novel approach for aspect-based sentiment analysis by integrating the information extracted from global dependency structures as well as pairwise correlations. To capture the aspect-to-text correlations, we design a CAGCN module based on the cross-attention mechanism. Meanwhile, to effectively exploit the syntactic graph, we design an SAGCN module with the self-attention mechanism to build the overall text-to-text connections. Experimental results on five benchmarks show the effectiveness of our proposed model, producing significantly better results than the baselines.

Keywords: Aspect-based sentiment analysis · Attention mechanism · Dependency graph

1 Introduction

Aspect-based sentiment analysis (ABSA) is a fine-grained sentiment analysis task, which aims to determine the sentiment polarities of given aspects in a sentence. For example, in the comment "the food is so delicious, but the service is horrible", the sentiments of the two aspects "food" and "service" are positive and negative, respectively.

To tackle the ABSA task which is essentially a classification problem, early works [17,19] have leveraged recurrent neural networks (RNNs) or convolutional neural networks (CNNs) to build the sentiment classifier. Nevertheless, these methods treat the sentence as a word sequence, and it is hard for them to model the dependency relationship between an aspect and its corresponding opinion expressions which can be far away from the aspect term.

To address that, many attention-based models [4,19] have been proposed with appealing results. They use the attention mechanism to model the dependency relationship between an aspect and its corresponding opinion expressions. However, they can be susceptible to noises in the dependency information. For example, there may exist several aspects with different opinion expressions, which may lead the attention mechanism to mistakenly connect an aspect with syntactically unrelated context words.
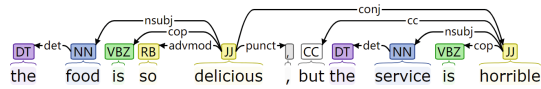

Fig. 1. An example of a dependency tree where opinion expressions (yellow) and the aspect expressions (blue) are connected based on their syntactic dependencies. (Color figure online)

To better exploit the dependency information, more recent efforts [15,22] have been devoted to incorporate the dependency tree into the graph based models. For example, the dependency tree depicted in Fig. 1 connects the aspect term "service" with the opinion word "horrible" with a single path, which allows the graph based models to better capture the syntactic dependencies. Although these models have shown better performance than those without considering syntactic relations, they still suffer from two potential limitations. First, they are vulnerable to parsing errors. Second, informal expressions in tweets, blogs and review comments also keep these models from working as well as expected. To address these issues, several recent works [1,6,23] explore the idea of combining different types of graph. However, these models mainly focus on extracting information from the global dependency structures, without considering the aspect-to-text correlations, which also provide valuable information.

To tackle the challenges mentioned before, we integrate the information extracted from global dependency structures as well as pairwise aspect-to-text correlations by constructing two separate graphs, based on which we propose a cross and self attention based graph convolutional network (CASAGCN) in this paper. On the one hand, we capture the aspect-to-text correlations using a sparse cross-attention mechanism. Motivated by [1], we replace the softmax function with α-entmax function [10] to project the resulting matrix into a sparse probability simplex to connect aspect words with highly relevant items. On the other hand, we build the global text-to-text connections based on the self-attention mechanism. Moreover, motivated by [14], which employs the saliency map as a priori knowledge to holistically refine the attention distribution, we propose to impose a syntactical guidance on the attention weights using dependency probability matrix of the sentence.

Our contributions are as follows:

– We integrate the information extracted from global dependency structures as well as pairwise aspect-to-text correlations by constructing two separate graphs, based on which we propose a cross and self attention based graph convolutional network (CASAGCN).

– We introduce the sparse cross-attention that enables the CAGCN module to capture the highly relevant context words for the given aspect. And we guide the self-attention score matrix with the dependency probability matrix by shortening their distance in the SAGCN module.
– Extensive experiment results on five public standard datasets verify the importance of integrating the overall dependency structures with the pairwise aspect-to-text correlations, and demonstrate the effectiveness of our model.

2   Related Work

ABSA is a fine-grained branch of sentiment classification, the goal of which is to identify the sentiment polarity of the given aspect in one sentence. Recent studies focus on developing various deep learning models. Among them, pioneering LSTM-based models [17] have been proposed to capture the contextual information which is highly related to the given aspect. These models use relatively simple methods to retrieve important information which is semantically related to the given aspect in the sentence context. Despite the effectiveness of these methods, it is still challenging for them to discriminate different sentiment polarities when facing complicated sentences with long-distance dependency. In addition to LSTM-based neural networks, attention mechanisms have been widely employed to model the relation between an aspect and its corresponding context [7, 16, 19, 21]. For instance, an attention-over-attention neural network is proposed in [19] to explicitly capture the interactions between aspects and contexts. [16] focuses on learning extra aspect embeddings and identifying the conflict opinions using positive and negative attention.

More recently, a line of works leverage the syntactic knowledge to help build the connections between aspects and opinions. Specifically, [15] uses a GCN to model the sentence's structure through its dependency tree. [1] proposes three methods to induce a latent graph, and combine it with the dependency graph to learn aspect-specific features. Instead of using a static tree obtained from off-the-shelf dependency parsers,

Further, several studies integrate different sources of information for the ABSA task. For example, [23] utilizes the word co-occurrence matrix and dependency tree to incorporate the statistic and syntactic information, followed by constructing a Bi-level GCN to distinguish different edges in a graph. [6] combines a semantic graph and a syntactic graph to alleviate the issues of parsing errors, informal expressions, and the complexity of online reviews.

3   Preliminaries

We start with a brief introduction of the GCN, which is a crucial part in our model.

3.1   Graph Convolutional Network (GCN)

GCN [5] can be considered as a CNN variant that encodes information for structured data. Specifically, GCN aggregates information from directly connected nodes. Further,

with multi-layer GCNs, each node in a graph can get more information from distant nodes. Formally, given a graph with n nodes and its corresponding adjacency matrix $A \in \mathbb{R}^{n \times n}$, the hidden representation of the i-th node at the l-th layer, denoted as h_i^l, is updated as follow:

$$h_i^l = \sigma(\sum_{j=1}^{n} A_{ij} W^l h_j^{l-1} + b^l) \tag{1}$$

where W^l is the parameter matrix, b^l is a bias term and σ is an activation function.

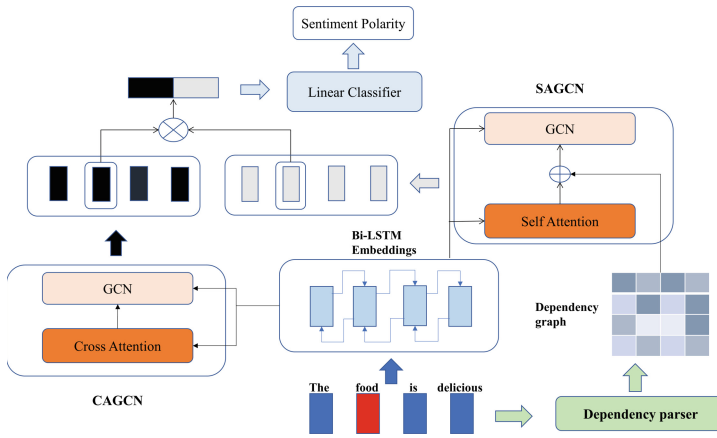## 4 Cross and Self Attention Based Graph Convolutional Network (CASAGCN)

Fig. 2. Model architecture

This section introduces the proposed framework of Cross and Self Attention based Graph Convolutional Network (CASAGCN) in details. The full architecture is shown in Fig. 2. In the ABSA task, the input is a sentence-aspect pair (s, a), where $s = \{w_1, w_2, ..., w_n\}$, and $a = \{a_1, a_2, ..., a_m\}$ is a sub-sequence of s. Given (s, a), we utilize BiLSTM or BERT(base) [2] to get the hidden representations. For the BiLSTM encoder, we first map each word in s into a real value vector and get the sequence of word embeddings $x = \{x_1, x_2, ..., x_n\}$ using the embedding lookup table $E \in \mathbb{R}^{|V| \times d_e}$, where $|V|$ is the size of the vocabulary and d_e is the dimensionality of word embeddings. Then, we feed the word embeddings of a sentence into a BiLSTM encoder to obtain hidden state vectors $H = \{h_1, h_2, ..., h_n\}$, where $h_t \in \mathbb{R}^{2d_l}$ is the output at time t and d_l denotes the output dimensionality of a unidirectional LSTM. For the BERT encoder, we construct the input as "[CLS] sentence [SEP] aspect [SEP]" to get the aspect-aware hidden representations. Next, the hidden state vectors are sent into the CAGCN and SAGCN modules, which will be described below. Finally, we aggregate representations of all aspect words from CAGCN and SAGCN to obtain the sentiment representation.

4.1   Cross-Attention Based GCN (CAGCN)

To fully exploit the pairwise aspect-to-text correlations, we construct a cross atten-tion based graph convolutional network by building the aspect-to-text connections with sparse cross-attention mechanism.

Sparse Cross-Attention. Cross-attention is adopted here to measure the similarity scores between aspect words and context words. Formally, given a sentence represen-tation H, we can mask the non-aspect words to get H^a, the adjacency matrix A^{cag} is then defined as

$$A^{cag} = \text{softmax}(\frac{QW^Q \times (KW^K)^\top}{\sqrt{d}}) \tag{2}$$

where Q and K are copies of H^a and H, respectively. $W^Q \in \mathbb{R}^{d \times d}$ and $W^K \in \mathbb{R}^{d \times d}$ are model parameters. Besides, $\sqrt{d}$ is a scale constant used to prevent dot products from growing large in magnitude.

However, cross-attention connects aspect words with every word in the sentence, which can introduce noise from irrelevant contexts. To address that, we consider sparse cross-attention to connect aspect words with highly relevant items. Specifically, we replace the softmax function in Eq. (2) with the α-entmax[1] function [10] which is more likely to assign a low-scoring choice with a zero probability to make the constructed adjacency matrix more sparse, that is:

$$A^{cag} = \alpha\text{-entmax}(\frac{QW^Q \times (KW^K)^\top}{\sqrt{d}}) \tag{3}$$

where $\alpha\text{-entmax}(z) = \arg\max_{p \in \Delta^d} p^\top x + H_\alpha^\top(p)$ and $H_\alpha^\top(p)$ is an entropy function.

After initializing the node representations with the hidden state vectors H, we apply GCN on the cross-attention graph A^{cag} constructed above. Using Eq. (1), we obtain the final graph representations $H^{cag} = \{h_1^{cag}, h_2^{cag}, ..., h_n^{cag}\}$ from the CAGCN module, where $h_i^{cag} \in \mathbb{R}^d$ is the hidden representation of the i-th node. Specially, for the aspect nodes, we denote their hidden representations as H_a^{cag}.

4.2   Self-Attention Based GCN (SAGCN)

In addition to the pairwise aspect-to-text correlations, we also capture the text-to-text relations at both the semantic and syntactic levels. However, instead of directly incor-porating the syntactic graph, we take it as a soft guide to mitigate the effects of parsing errors and informal expressions. Specifically, in the SAGCN module, we first capture the global semantic correlations via the self-attention mechanism. Meanwhile, consid-ering that the self-attention mechanism may take the wrong words as descriptors, we shorten the distance between the attention matrix and the dependency probability matrix from the dependency parser to guide the self-attention score distribution.

---

[1] We use the implementation from https://github.com/deep-spin/entmax.

Dependency Graph Guided Self-Attention. Similar to Eq. (2), we first compute the attention score matrix A^{sa} using a self-attention layer, where matrices Q and K are both equal to the hidden representations from the previous BiLSTM layer. Second, instead of using the final discrete output of a dependency parser, we get the dependency probability matrix G from the external dependency parser, which could capture rich structural information by providing all latent syntactic structures [6]. We proceed to transform G to an undirected graph by

$$A^{dep} = G + G^{\top}. \tag{4}$$

In addition, a self loop is included for each node in the dependency graph A^{dep} to keep the information of each node itself. The final graph can then be obtained by combining the two graphs A^{sa} and A^{dep} as follows:

$$A_i^{sag} = \frac{A_i^{sa} + \lambda A_i^{dep}}{\sum_{j=1}^{N} A_{ij}^{sa} + \lambda \sum_{j=1}^{N} A_{ij}^{dep}} \tag{5}$$

where A_i represents the i-th row of a graph, A_{ij} is the j-th element in A_i, and λ is a hyperparameter to control the distance A^{sa} moving towards G^{dep}.

Similarly, we obtain the graph representations H^{sag} from the SAGCN module, and the representations for the aspect nodes are denoted as H_a^{sag}.

To obtain the final feature representation for the ABSA task, we apply average pooling on the aspect node representations from the CAGCN and SAGCN modules, followed by concatenating them. Formally,

$$z = \text{Concat}(f(H_a^{cag}), f(H_a^{sag})) \tag{6}$$

where $f(\cdot)$ represents the average pooling function, and z is the final aspect-specific representation. Then, z is used to calculate the sentiment probability distribution with a linear classifier. Formally,

$$p = \text{softmax}(W_l z + b_l) \tag{7}$$

where W_l and b_l are model parameters.

Objective. The model is trained to minimize the following loss function:

$$\ell(\theta) = -\sum_{i=1}^{N} \sum_{c \in C} log(p) + \lambda_1 ||\theta||_2^2 + \lambda_2 R_O \tag{8}$$

where θ represents all trainable model parameters, λ_1 and λ_2 are regularization coefficients, and C denotes all distinct sentiment polarities. The first two terms represent the standard cross-entropy loss and L_2-regularization, respectively. In addition, following [6], we add the third term to encourage orthogonality among the rows of the self-attention score matrix because the related items of each word should be in different regions in a sentence, which can be formulated as:

$$R_O = ||A^{sag} A^{sag\top} - I||_F \tag{9}$$

5 Experiments

5.1 Datasets

We conduct experiments on five benchmark datasets for aspect-based sentiment analysis, including Lap14, Rest14, Rest15, Rest16 and Twitter. The Lap14, Rest14, Rest15, Rest16 datasets are from the SemEval 2014 ABSA challenge [11], SemEval 2015 ABSA challenge [12] and SemEval 2016 ABSA challenge [13], respectively. Twitter consists of tweets from [3]. The statistics for five datasets are summarized in Table 1.

Table 1. Dataset statistics

Dataset	Lap14	Rest14	Twitter	Rest15	Rest16
Train	2282	3608	6051	1204	1748
Test	632	1119	677	542	616

5.2 Implementation and Parameter Settings

The 300-dimensional Glove vectors[2] [9] are used to initialize word embeddings for all our experiments. Moreover, the dimensionality for part-of-speech (POS) embeddings and position embeddings is set to 30 as in [15] to identify the relative position between each word and the aspect. Then, the word, POS and position embeddings are concatenated and sent to a BiLSTM model as input. The hidden size of the BiLSTM model is set to 50. The 1.25-entmax function is applied to each row of the resulted matrix in CAGCN. The dependency parser we use in SAGCN is LAL-Parser[3] [8]. The hyperparameter λ is set to 1.0, 1.4, 1.0, 1.0 and 1.0 for the five datasets, respectively. The number of GCN layers in CAGCN and SAGCN are set to 1 and 2, respectively. And the dropout rate of the CAGCN and SAGCN modules is set to 0.1. All the model weights are initialized from a uniform distribution. We use the Adam optimizer with learning rate 0.002 for all datasets. The CASAGCN model is trained in 40 epochs with a batch size of 16. λ_1 is set to 10^{-4}, and λ_2 is set to 0.2, 0,2, 0.3, 0.2 and 0.2 for the five datasets respectively. We train our framework on one Nvidia 1080Ti GPU, and it takes less than one hour on each dataset to finish training.

5.3 Baseline Methods

We compare the proposed CASAGCN model to a list of baselines, which are briefly summarized below.

(1) **IAN** [7] employs an interactive attention mechanism to learn the representation for the given aspect.

---

[2] https://nlp.stanford.edu/projects/glove/.
[3] https://github.com/KhalilMrini/LAL-Parser.

(2) **ASGCN** [22] implements a multi-layered GCN on top of the LSTM output and uses a masking mechanism to obtain high-level aspect-specific features.

(3) **CDT** [15] encodes the dependency tree using GCN to propagate dependency information from opinion words to aspect words.

(4) **DualGCN** [6] designs a dual graph convolutional networks which takes both syntactic and semantic information into consideration.

(5) **DGEDT-BERT** [18] jointly considers the flat representations and graph-based representations learnt from the corresponding dependency graph in an iterative interaction manner.

(6) **BERT4GCN** [20] incorporates the knowledge from the intermediate layers in BERT which can enhance GCN and obtain better features of ABSA task.

(7) **DualGCN-BERT** [6] is DualGCN that uses a BERT as encoder.

(8) **SSEGCN-BERT** [24] proposes a novel syntactic and semantic enhanced graph convolutional network to learn the aspect-related semantic correlations and obtain comprehensive syntactic structure information.

5.4   Comparison Results

In this subsection, we compare the recent methods with CASAGCN using the accuracy and macro-averaged F1-score as the main evaluation metrics. The results are shown in Table 2. From these results, we observe that our CASAGCN model consistently outperforms all the compared models on the Lap14, Twitter and Rest16 datasets, and achieves competitive performance on the Rest14 and Rest15 datasets. These results demonstrate the effectiveness of our CASAGCN for integrating the information extracted from global dependency structures as well as pairwise aspect-to-text correlations. Compared to the attention-based models like IAN, our CASAGCN model utilizes both cross-attention and self-attention guided by syntactic knowledge simultaneously to model the dependencies between the aspect and opinion words. As a consequence, it can reduce the noise caused by the attention mechanism. Besides, the graph based and syntax integrated methods (ASGCN, CDT) achieve better performance than those without considering syntax. However, informal expressions or parsing errors still degrade the performance of these models, while our CASAGCN can perform better when facing the complicated and informal sentences and mitigate the noises from parsing errors. Moreover, utilizing BERT as the encoder, our CASAGCN-BERT also achieves better performance than BERT-based models (DGEDT-BERT, BERT4GCN, DualGCN-BERT, and SSEGCN-BERT).

5.5   Ablation Study

To investigate the influence of each component in our CASAGCN model, we conduct extensive ablation studies on the Lap14 dataset and show the results in Table 3. As expected, all simplified variants have lowered accuracy. Compared with the complete CASAGCN model, the decreased performance of both CAGCN and SAGCN validates that integrating the information extracted from global dependency structures as well as pairwise aspect-to-text correlations is better than focusing only on one of them. In addition, we find that CAGCN and SAGCN have competitive results, indicating that

Table 2. Performance comparison on five benchmark datasets. The best scores are bolded.

Models	Lap14		Rest14		Twitter		Rest15		Rest16	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
IAN [7]	72.05	67.38	79.26	70.09	72.50	70.81	78.54	52.65	84.74	55.21
ASGCN [22]	75.55	71.05	80.77	72.02	72.15	70.40	79.89	61.89	88.99	67.48
CDT [15]	77.19	72.99	82.30	74.02	74.66	73.66	79.42	61.68	85.58	69.93
DualGCN [6]	78.48	74.74	**84.27**	**78.08**	75.92	74.29	81.37	60.09	88.96	67.58
CASAGCN	**79.43**	**75.80**	84.18	77.55	**76.22**	**74.60**	81.73	64.33	89.45	72.98
DGEDT-BERT [18]	79.80	75.60	86.30	80.00	77.90	75.40	84.00	71.00	91.90	79.00
BERT4GCN [20]	77.49	73.01	84.75	77.11	74.73	73.76	-	-	-	-
DualGCN-BERT [6]	**81.80**	78.10	87.13	81.16	77.40	76.02	84.32	65.26	**92.53**	78.66
SSEGCN-BERT [24]	81.01	77.96	**87.31**	81.09	77.40	76.02	-	-	-	-
CASAGCN-BERT	81.65	**78.39**	87.13	**81.27**	**78.73**	**77.44**	**85.61**	**71.91**	91.90	**79.44**

they have their own contributions. CASAGCN w/o dep indicates that we do not use the dependency probability matrix to guide the self-attention score matrix. Therefore, the performance degrades substantially on the Lap14 dataset which justifies that the dependency graph guided self-attention can better model the dependency between aspects and the corresponding contexts.

Table 3. Ablation study on the Lap14 dataset.

Models	Accuracy	Macro-F1
CAGCN	77.37	73.75
SAGCN	78.32	74.87
CASAGCN w/o dep	78.16	74.55
CASAGCN	**79.43**	**75.80**

5.6 Case Study

To better understand the behaviour of our CASAGCN model, we present the case study on a few sample cases in this subsection. Table 4 shows the results of different models. We denote positive, negative and neutral sentiment as Pos, Neg and Neu, respectively. In the first example, the sentence has a long and complicated structure where the attention-based model IAN fails. For the aspect words "windows 8" in the second example, IAN and CAGCN are unable to make the correct prediction due to the lack of syntax information, while SAGCN and CASAGCN can connect aspect words and the opinion words with the help of dependency graph. Moreover, in the third example, both CAGCN and SAGCN fail to give the right sentiment polarity for the aspect words "touchscreen function". However, by combining the information from these two modules, our CASAGCN successfully capture the feature representations of the key

words "did not". Overall, these three examples demonstrates our CASAGCN, taking both global dependency and pairwise correlations in to consideration, can handle complex and informal sentences in the ABSA task.

Table 4. Case studies of our CASAGCN model compared with baselines. Aspect words are in italic.

#	Reviews	Sentiment	IAN	CAGCN	SAGCN	CASAGCN
1	*Tech support* would not fix the problem unless I bought your plan for $150 plus	Neg	Neu	Neg	Neg	Neg
2	Did not enjoy the new *Windows 8* and touchscreen functions	Neg	Neu	Neu	Neg	Neg
3	Did not enjoy the new Windows 8 and *touchscreen functions*	Neg	Neu	Neu	Pos	Neg

5.7   Attention Visualization

To investigate the effectiveness of the dependency graph guided self-attention in connecting aspect terms and corresponding opinion expressions, we visualize the original self-attention score matrix and dependency graph guided self-attention score matrix. Take the sentence "It's fast, light, and simple to use." with an aspect term "use" as an example. As shown in Fig. 3, the original self-attention score is dense, every word gives other words very close attention scores, which will bring noise in the information propagation stage. In addition, by observing the attention probability distribution of "use" in the 10-th row, we can find that it does not distinguish the corresponding opinion expression "simple" and mistakenly pays too much attention to "to" and "." which are not helpful for judging the sentiment polarity. Further, we can observe that the dependency graph incorrectly connects the word "light" with many other words, which also demonstrates the problems of relying too much on the dependency graph. In contrast, after adding a syntactical guidance, our CASAGCN module produces a more sparse self-attention matrix. The dependency path between words "simple" and "use" allows the attention probability distribution to be adjusted correctly, so that our model can make a right prediction.
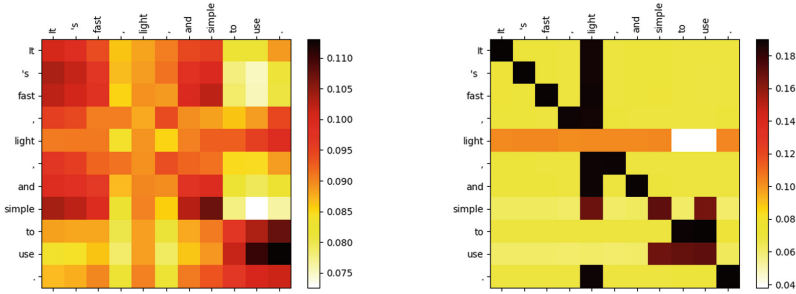
(a) The original self-attention score matrix.   (b) The self-attention score matrix guided by the dependency graph.

Fig. 3. Self-attention score distributions.

6   Conclusion

In this paper, we highlight the importance of combining information from both global dependency structures and pairwise aspect-to-text correlations in the ABSA task and propose a novel framework CASAGCN to exploit these information through the CAGCN and SAGCN modules. We utilize sparse cross-attention to model the aspect-to-text correlations. Moreover, we impose a syntactical guidance for better constructing the text-to-text connections. Extensive experiments on five real-world datasets demonstrate the effectiveness of the proposed CASAGCN model with superior performance.

References

1. Chen, C., Teng, Z., Zhang, Y.: Inducing target-specific latent structures for aspect sentiment classification. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (2020)
2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: 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. Association for Computational Linguistics, Minneapolis, Minnesota, June 2019. https://doi.org/10.18653/v1/N19-1423. https://aclanthology.org/N19-1423
3. Dong, L., Wei, F., Tan, C., Tang, D., Xu, K.: Adaptive recursive neural network for target-dependent twitter sentiment classification. In: Meeting of the Association for Computational Linguistics (2014)
4. Huang, B., Ou, Y., Carley, K.M.: Aspect Level Sentiment Classification with Attention-Over-Attention Neural Networks. Springer, Cham (2018)

5. Kipf, T., Welling, M.: Semi-supervised classification with graph convolutional networks. ArXiv abs/1609.02907 (2017)

6. Li, R., Chen, H., Feng, F., Ma, Z., Wang, X., Hovy, E.: Dual graph convolutional networks for aspect-based sentiment analysis. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 6319–6329. Association for Computational Linguistics, Online, August 2021. https://doi.org/10.18653/v1/2021.acl-long.494. https://aclanthology.org/2021.acl-long.494

7. Ma, D., Li, S., Zhang, X., Wang, H.: Interactive attention networks for aspect-level sentiment classification. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp. 4068–4074 (2017). https://doi.org/10.24963/ijcai.2017/568. https://doi.org/10.24963/ijcai.2017/568

8. Mrini, K., Dernoncourt, F., Bui, T., Chang, W., Nakashole, N.: Rethinking self-attention: an interpretable self-attentive encoder-decoder parser. arXiv preprint arXiv:1911.03875 (2019)

9. Pennington, J., Socher, R., Manning, C.: Glove: global vectors for word representation. In: Conference on Empirical Methods in Natural Language Processing (2014)

10. Peters, B., Niculae, V., Martins, A.F.T.: Sparse sequence-to-sequence models. ArXiv abs/1905.05702 (2019)

11. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Manandhar, S.: Semeval-2014 task 4: aspect based sentiment analysis. In: Proceedings of International Workshop on Semantic Evaluation at (2014)

12. Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., Androutsopoulos, I.: SemEval-2015 task 12: Aspect based sentiment analysis. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pp. 486–495. Association for Computational Linguistics, Denver, Colorado (Jun 2015). https://doi.org/10.18653/v1/S15-2082. https://aclanthology.org/S15-2082

13. Pontiki, M., et al.: Semeval-2016 task 5: aspect based sentiment analysis. In: International Workshop on Semantic Evaluation (2016). https://api.semanticscholar.org/CorpusID:1021411

14. Song, K., Yao, T., Ling, Q., Mei, T.: Boosting image sentiment analysis with visual attention. Neurocomputing **312**(27), 218–228 (2018)

15. Sun, K., Zhang, R., Mensah, S., Mao, Y., Liu, X.: Aspect-level sentiment analysis via convolution over dependency tree. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (2019)

16. Tan, X., Cai, Y., Zhu, C.: Recognizing conflict opinions in aspect-level sentiment classification with dual attention networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (2019)

17. Tang, D., Qin, B., Feng, X., Liu, T.: Effective lstms for target-dependent sentiment classification. Computer Science (2015)

18. Tang, H., Ji, D., Li, C., Zhou, Q.: Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (2020)

19. Wang, Y., Huang, M., Zhu, X., Zhao, L.: Attention-based lstm for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (2016)

20. Xiao, Z., Wu, J., Chen, Q., Deng, C.: BERT4GCN: using BERT Intermediate Layers to Augment GCN for Aspect-based Sentiment Classification. arXiv e-prints arXiv:2110.00171, September 2021

21. Yang, M., Tu, W., Wang, J., Xu, F., Chen, X.: Attention-based LSTM for target-dependent sentiment classification. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI 2017, pp. 5013–5014. AAAI Press (2017)
22. Zhang, C., Li, Q., Song, D.: Aspect-based Sentiment Classification with Aspect-specific Graph Convolutional Networks. arXiv e-prints arXiv:1909.03477, September 2019
23. Zhang, M., Qian, T.: Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 3540–3549. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.emnlp-main.286. https://aclanthology.org/2020.emnlp-main.286
24. Zhang, Z., Zhou, Z., Wang, Y.: Ssegcn: syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis. In: North American Chapter of the Association for Computational Linguistics (2022). https://api.semanticscholar.org/CorpusID:250391015