Affective Knowledge Enhanced Multiple-Graph Fusion Networks for Aspect-based Sentiment Analysis

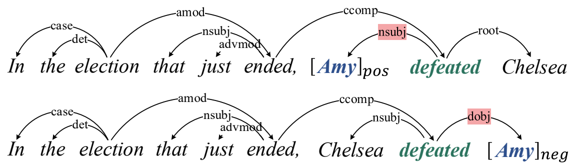
Siyu Tang[1*], **Heyan Chai**[1*], **Ziyi Yao**[1], **Ye Ding**[2†], **Cuiyun Gao**[1],
Binxing Fang[1,3], **Qing Liao**[1,3†]
[1] Harbin Institute of Technology, Shenzhen, China
[2] Dongguan University of Technology, China
[3] Peng Cheng Laboratory, Shenzhen, China
{tangsiyu999, chaiheyan, yaoziyi}@stu.hit.edu.cn,
dingye@dgut.edu.cn, gaocuiyun@hit.edu.cn, fangbx@cae.cn, liaoqing@hit.edu.cn

Abstract

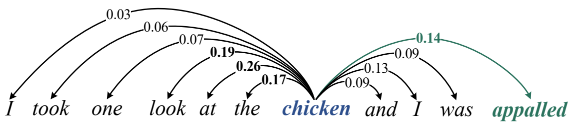
Aspect-based sentiment analysis aims to identify sentiment polarity of social media users toward different aspects. Most recent methods adopt the aspect-centric latent tree to connect aspects and their corresponding opinion words, thinking that would facilitate establishing the relationship between aspects and opinion words. However, these methods ignore the roles of syntax dependency relation labels and affective semantic information in determining the sentiment polarity, resulting in the wrong prediction. In this paper, we propose a novel multi-graph fusion network (MGFN) based on latent graph to leverage the richer syntax dependency relation label information and affective semantic information of words. Specifically, we construct a novel syntax-aware latent graph (SaLG) to fully leverage the syntax dependency relation label information to facilitate the learning of sentiment representations. Subsequently, a multi-graph fusion module is proposed to fuse semantic information of surrounding contexts of aspects adaptively. Furthermore, we design an affective refinement strategy to guide the MGFN to capture significant affective clues. Extensive experiments on three datasets demonstrate that our MGFN model outperforms all state-of-the-art methods and verify the effectiveness of our model.

1 Introduction

Sentiment analysis has been a popular research subject in natural language processing. Aspect-based sentiment analysis (ABSA) (Birjali et al., 2021) is a fine-grained sentiment analysis task. For example, given a sentence "*The **menu** is limited but the **dishes** are excellent.*", there are two aspects mentioned in the sentence and the sentiment polarity of aspects "*menu*" and "*dishes*" are *negative* and *positive*, respectively. Generally, ABSA task is formulated as predicting the polarity of a given sentence-aspect pair. The main challenge of ABSA is to precisely capture the relationship between the aspect and its corresponding opinion expressions.



(a) The dependency parse tree from spaCy.



(b) The dependency tree derived by ACLT.

Figure 1: (a) Two similar sentences with aspect "Amy", each with its own dependency tree. (b) An example,the numbers in arcs denote the weight of edge between aspect word and its contextual words, derived from ACLT (Zhou et al., 2021).

Many existing graph-based methods(Sun et al., 2019a; Zhao et al., 2020; Wang et al., 2020; Li et al., 2021b) have been devoted to obtaining promising performance of ABSA task by constructing graph neural networks (GNNs) over dependency trees. They generally rely on the off-the-shelf dependency parsers to generate the static syntactic relationship between words in a sentence, which is insufficient to adaptively search for the affective clues of aspects from the contexts. Recent efforts (Chen et al., 2020; Zhou et al., 2021) show that latent graph derived from dynamic latent trees can adaptively capture the relationship between words in a sentence, leading to better performance in ABSA.

Despite promising progress made by latent graph based methods, they still suffer from two potential limitations: (1) They ignore the richer syntactic information contained in syntax dependency relation labels[1] (e.g., *nsubj* and *dobj* in Figure 1), leading

[*] The first two authors contributed equally to this work.
[†] Corresponding Authors

[1]The grammatical relation between the head and the de-

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 5352 - 5362
December 7-11, 2022

models to make wrong predictions. We show examples in Figure 1 (a) where these two sentences are very similar and have the same aspect "*Amy*". Noting that aspect "*Amy*" presents the opposite sentiment polarities in these two sentences. The main reason of wrong prediction is that the same aspects may signal different sentiment polarities when they have different syntax dependency relation labels (*nsubj* and *dobj* with red color) with opinion words. Therefore, it is important to model the syntax dependency relations between words and fuse them into the latent graph to improve the performance of ABSA task. (2) They pay more attention to neighbor words of aspects, bringing extra difficulty in capturing the interaction between aspects and their corresponding long-distance opinion words. To illustrate this limitation, we give an example in Figure 1 (b) where attention scores of every word are derived from existing state-of-the-art latent graph method, ACLT (Zhou et al., 2021). Noting that the attention value between aspect "*chicken*" and its corresponding opinion word "*appalled*" is 0.14 which is much lower than that between the aspect and its neighbor words (*e.g.* 0.26 *for "at"*, 0.17 *for "the", etc.*). This implies that the existing latent graph overly focuses excessively on the neighbor words of aspects, while ignoring affective semantic information of words. Such a limitation may prevent the model from accurately capturing the interaction between aspects and their corresponding opinion words, thus degrading performance.

To address the aforementioned two limitations, in this paper, we propose a novel multi-graph fusion network (MGFN) based on latent graph to leverage the richer syntax dependency relation label information and affective semantic information of words. Specifically, we construct a novel syntax-aware latent graph (SaLG) by integrating syntax dependency relation label information to facilitate the learning of sentiment representations in ABSA task. Subsequently, we design a multi-graph fusion module to fuse the information of the syntax-aware latent graph and the semantic graph (SeG), so that the SaLG can leverage the semantic information to capture significant sentiment features. In addition, we design a novel affective refinement strategy to guide the model to determine the significant affective clues from surrounding contexts, which can effectively enable the model to capture the interaction between aspect words and long-distance opinion words.

Our contributions are highlighted as follows:

- We have come up with a kind of syntax-aware latent graph (SaLG) by leveraging the syntax dependency relation label information to facilitate the learning of sentiment representation.
- A novel multi-graph fusion network (MGFN) is proposed by integrating the semantic information learned from semantic graph (SeG) into SaLG to capture more accurate sentiment representations.
- We also propose an affective refinement strategy to guide MGFN model to pay more attention to opinion expressions of aspect words.
- Experimental results illustrate that our MGFN model outperforms the state-of-the-art methods on SemEval 2014 and Twitter datasets.

2 Methodology

In this section, we elaborate on the details of our proposed model. The overall framework of MGFN is shown in Figure 2. It contains four components: 1) Text Encoding Module encodes the contextualized representations of input sentence. 2) Graph Construction Module constructs a novel syntax-aware latent graph (SaLG) and a semantic graph (SeG), respectively. 3) Multi-Graph Fusion Module adaptively integrates semantic information from SeG into SaLG via an adaptive fusion gate. 4) Affective Refinement Module introduces a novel affective refinement strategy to encourage MGFN to pay more attention to the opinion expressions of aspect words.

2.1 Text Encoding Module

Given a n-word sentence $s = \{w_1, w_2, \cdots, w_{\tau+1}, \cdots, w_{\tau+m}, \cdots, w_n\}$ with the aspect $a = \{w_{\tau+1}, \cdots, w_{\tau+m}\}$, we utilize the pre-trained language model BERT (Devlin et al., 2019) to obtain contextualized representation for each word. For the BERT encoder, we first construct a BERT-based sentence-aspect pair $\mathbf{x} = ([\mathrm{CLS}]\ s\ [\mathrm{SEP}]\ a\ [\mathrm{SEP}])$ as input. The output contextualized representation $\boldsymbol{H} = \mathrm{BERT}(\mathbf{x})$. $\boldsymbol{H} = [\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_n] \in \mathbb{R}^{n \times d}$, where d denotes the dimensionality of BERT embeddings and $\boldsymbol{h}_i$ is the contextual representation of the i-th word.

pendent word (Wang et al., 2020).

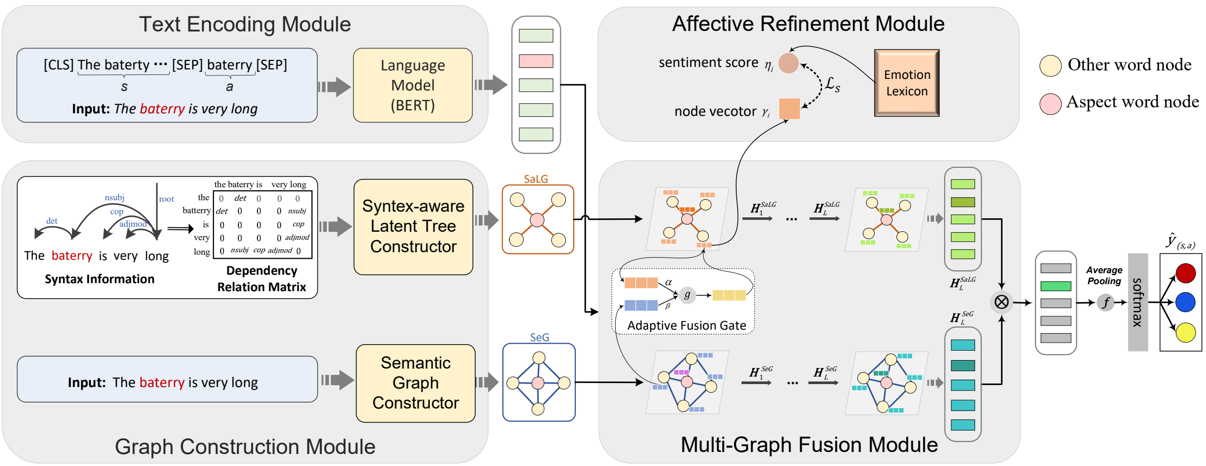


Figure 2: The overall architecture of MGFN, which is composed primarily of four modules.

2.2 Graph Construction Module

2.2.1 Syntax-aware latent graph

In order to capture syntax dependency relation label information, we construct a novel syntax-aware latent graph (SaLG) by implicitly labeling the edges with different dependency relations.

We construct dependency relation matrix $\boldsymbol{R} \in \mathbb{R}^{n\times n}$ from off-the-shelf dependency parser to utilize the dependency relation label information. Each $\boldsymbol{r}_{ij} \in \boldsymbol{R}$ represents the syntax dependency relation label between i-th and j-th words:

$$\boldsymbol{r}_{ij} = \begin{cases} deprel & \text{if } \operatorname{link}(i,j) = 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $\operatorname{link}(i,j)$ shows that i-th and j-th words have a dependence link, and $deprel$ is dependency relation label (e.g., *nsubj, dobj*). A new dependency relation dictionary $\boldsymbol{V}^r$ is built based on the frequency of $deprel$ in corpus to encode dependency relations:

$$\boldsymbol{V}^r = \{deprel : \operatorname{toId}(p(deprel))\} \tag{2}$$

$$p(deprel) = \frac{N(deprel)}{N} \tag{3}$$

where $\operatorname{toId}(\cdot)$ can map each kind of $deprel$ into a corresponding non-repeating integer ID according to its frequency calculated by p. $N(deprel)$ is the number of $deprel$, N is the total number of all kinds of $deprel$. By using the constructed $\boldsymbol{V}^r$ as lookup table, each relation r_{ij} can be embedded into high-dimensional word embedding vector $\boldsymbol{e}_{ij} \in \mathbb{R}^{1\times d_e}$. Subsequently, syntactic relation type-aware matrix $\tilde{\boldsymbol{A}} \in \mathbb{R}^{n\times n}$ is defined as:

$$\tilde{\boldsymbol{A}}_{ij} = \operatorname{softmax}(\boldsymbol{W}^a \boldsymbol{e}_{ij} + \boldsymbol{b}^a) \tag{4}$$

Utilizing $\tilde{\boldsymbol{A}}$ as initial edge weight matrix, the syntax-aware latent tree with n nodes is derived by tree inducer (Zhou et al., 2021), where each node is the word of input sentence. Firstly, we define the variant of Laplacian matrix $\widehat{\boldsymbol{L}}$ of the syntax-aware latent tree which further accounts for the dependencies headed by the root symbol:

$$\widehat{\boldsymbol{L}}_{ij} = \begin{cases} \psi_i + \sum_{i'=1}^{n} \tilde{\boldsymbol{A}}_{i'j} & \text{if } i = j \\ -\tilde{\boldsymbol{A}}_{ij} & \text{otherwise} \end{cases} \tag{5}$$

where $\psi_i = \exp(\boldsymbol{W}^r \boldsymbol{h}_i + \boldsymbol{b}^r)$ is the score of i-th node to be selected as structure root. $\widehat{\boldsymbol{L}}$ can be used to simplify calculation of the sum of weights. Subsequently, the marginal probability $\boldsymbol{A}_{ij}^{SaLG}$ of the syntax-aware latent tree is calculated by $\widehat{\boldsymbol{L}}_{ij}$:

$$\boldsymbol{A}_{ij}^{SaLG} = \begin{cases} \tilde{\boldsymbol{A}}_{ij}[\widehat{\boldsymbol{L}}^{-1}]_{jj} & i = 1 \text{ and } j \neq 1 \\ \tilde{\boldsymbol{A}}_{ij}[\widehat{\boldsymbol{L}}^{-1}]_{ji} & i \neq 1 \text{ and } j = 1 \\ \tilde{\boldsymbol{A}}_{ij}[\widehat{\boldsymbol{L}}^{-1}]_{jj} \\ \quad -\tilde{\boldsymbol{A}}_{ij}[\widehat{\boldsymbol{L}}^{-1}]_{ji} & i \neq 1 \text{ and } j \neq 1 \\ 0 & i = 1 \text{ and } j = 1 \end{cases} \tag{6}$$

where $\boldsymbol{A}^{SaLG}$ can be seen as the weighted adjacency matrix of SaLG transformed from syntax-aware latent tree.

We adopt a root constraint strategy (Zhou et al., 2021) to keep SaLG be rooted at aspect:

$$\mathcal{L}_r = -\sum_{i=1}^{N} p_i^r \log \widehat{\boldsymbol{P}}_i^r + (1 - p_i^r) \log(1 - \widehat{\boldsymbol{P}}_i^r) \tag{7}$$

where, $\widehat{\boldsymbol{P}}_i^r = \psi_i[\widehat{\boldsymbol{L}}^{-1}]_{i1}$ is the probability of i-word headed by the root of latent structure. $p_i^r \in \{0, 1\}$ represents whether i-th word is the aspect.

2.2.2 Semantic Graph

The semantic graph (SeG) offers semantic information. The adjacency matrix $\boldsymbol{A}^{SeG} \in \mathbb{R}^{n \times n}$ of SeG is obtained via a multi-head self-attention mechanism for calculating the semantic similarity:

$$\boldsymbol{A}^{SeG} = \frac{\sum_{k=1}^{K} \boldsymbol{A}^{SeG,k}}{K} \quad (8)$$

$$\boldsymbol{A}^{SeG,k} = \mathrm{softmax}(\frac{\boldsymbol{H}\boldsymbol{W}^{Q} \times (\boldsymbol{H}\boldsymbol{W}^{K})^{T}}{\sqrt{D_H}}) \quad (9)$$

where K is the number of attention heads. $A^{sem,k}$ is attention scores matrix of k-th head. $\sqrt{D_H}$ is the dimensionality of contextual representation $\boldsymbol{H}$.

2.3 Multi-Graph Fusion Module

Since SaLG fails to fully focus on the opinion expressions, we design a multi-graph fusion module with adaptive fusion gate to offer semantic information guide, adaptively fusing semantic information from SeG into SaLG during iterative interaction.

The hidden state representation of SaLG and SeG at l-th layer is updated through stacked common graph convolutional (C-GCN) blocks:

$$\boldsymbol{H}_l^{SaLG} = \sigma(\boldsymbol{A}^{SaLG}\boldsymbol{W}_l^{c}\boldsymbol{H}_{l-1}^{SaLG} + \boldsymbol{b}_l^{c}) \quad (10)$$

$$\boldsymbol{H}_l^{SeG} = \sigma(\boldsymbol{A}^{SeG}\boldsymbol{W}_l^{c}\boldsymbol{H}_{l-1}^{SeG} + \boldsymbol{b}_l^{c}) \quad (11)$$

where $\boldsymbol{H}_l^{SaLG}$ and $\boldsymbol{H}_l^{SeG}$ are SaLG and SeG representations at the l-th layer. $\boldsymbol{H}_{l-1}^{SaLG}$ and $\boldsymbol{H}_{l-1}^{SeG}$ are inputs of preceding layer of the C-GCN block and $\boldsymbol{H}$ is the initial input of the first block. $\boldsymbol{W}_l^{c}$ and $\boldsymbol{b}_l^{c}$ are the shared trainable parameters. Meanwhile, an adaptive fusion gate is adopted to adaptively integrate $\boldsymbol{H}_l^{SaLG}$ and $\boldsymbol{H}_l^{SeG}$ for each node:

$$\boldsymbol{H}_l^{SaLG} = \mathrm{ReLU}(\boldsymbol{W}_l(\alpha\boldsymbol{H}_l^{SaLG} + \beta\boldsymbol{H}_l^{SeG})) \quad (12)$$

$$\alpha = \rho \cdot \sigma(\mathrm{g}(\boldsymbol{H}_l^{SaLG})) \quad (13)$$

$$\beta = 1 - \alpha \quad (14)$$

where α and β are the dynamic fusion proportions. $\mathrm{g}(\cdot)$ is a self-gating function (Bo et al., 2021) with a shared convolutional kernel. $\rho \in [0, 1]$ is the hyper-parameter of prior knowledge. $l \in [1, L]$.

We use control factor $\omega = \sigma(\mathrm{g}(\boldsymbol{H}_{l-1}))$ to retain the information of preceding layer of C-GCN block to relieve the over-smoothing problem:

$$\boldsymbol{H}_l^{SaLG} = \omega \cdot \boldsymbol{H}_l^{SaLG} + (1 - \omega) \cdot \boldsymbol{H}_{l-1}^{SaLG} \quad (15)$$

Capture significant sentiment feature. The latent-specific attention mechanism is utilized to capture significant sentiment features of SaLG:

$$\boldsymbol{\varepsilon} = \mathrm{softmax}(\boldsymbol{H}_L^{SaLG}\boldsymbol{H}_L^{SeG^{\top}}) \quad (16)$$

where $\boldsymbol{\varepsilon}$ is semantic-aware latent weight based on the output representation of the last C-GCN block. Then we can obtain a more richer sentiment representations $\boldsymbol{z} = \boldsymbol{\varepsilon}\boldsymbol{H}_L^{SeG}$. To make feature aspect-oriented, a mask mechanism is utilized to get aspect-oriented sentiment feature representation $\boldsymbol{z}_i^{A} = m_i\boldsymbol{z}_i$:

$$m_i = \begin{cases} 0, & 1 \leq i < \tau + 1, \tau + m < t \leq n \\ 1, & \tau + 1 \leq t \leq \tau + m \end{cases} \quad (17)$$

where $\tau + 1 \leq t \leq \tau + m$ denotes the aspect words.

2.4 Affective Refinement Module

In order to guide MGFN to determine the significant affective clues from surrounding contexts, we propose a novel affective refinement strategy to better correlate the aspect and opinion words.

We use SenticNet6 (Cambria et al., 2020) to get the affective score η_i for each word of input sentence in order to obtain a lexicon vector $\boldsymbol{lex} \in \mathbb{R}^{n \times 1} = [\eta_1, \eta_2, \cdots, \eta_n]$, where $\eta_i = 0$ if i-th word is not in SenticNet6. Meanwhile, the hidden state representation $\boldsymbol{H}_l^{SaLG}$ at l-th layer is mapped into the intermediate vector $\boldsymbol{\gamma}^{SaLG} \in \mathbb{R}^{n \times 1} = [\gamma_1, \gamma_2, \cdots, \gamma_n]$, where each low-dimensional node representation γ_i is given by:

$$\gamma_i = \boldsymbol{W}^{SaLG}\boldsymbol{H}_{l,i}^{SaLG} + \boldsymbol{b}^{SaLG} \quad (18)$$

Through minimizing the loss function $\mathcal{L}_s$ of affective refinement strategy, ideally, our model will pay more attention to the opinion expressions of aspect words:

$$\mathcal{L}_s = (\boldsymbol{\gamma}^{SaLG} - \boldsymbol{lex})^2 \quad (19)$$

2.5 Model Training

Softmax classifier. To deal with multi-word aspect, we apply average pooling on aspect nodes of $\boldsymbol{z}^{A}$, and calculate the sentiment probability distribution $\hat{y}_{(s,a)}$ by a linear layer with softmax function:

$$\hat{y}_{(s,a)} = \mathrm{softmax}(\boldsymbol{W}^{p}\mathrm{AvePooling}(\boldsymbol{z}^{A}) + \boldsymbol{b}^{p}) \quad (20)$$

where (s, a) is a sentence-aspect pair.

Our training goal is to minimize the following overall objective function:

$$\mathcal{L}(\Theta) = \lambda \mathcal{L}_C + \mu_1 \mathcal{L}_r + \mu_2 \mathcal{L}_s \tag{21}$$

where Θ represents all trainable parameters of model. λ, μ_1 and μ_2 are the hyper-parameters. The cross-entropy loss L_C for main classification task is defined as follows:

$$\mathcal{L}_C = \sum_{(s,a)\in\mathcal{D}} y_{(s,a)} \log \hat{y}_{(s,a)} \tag{22}$$

where $\mathcal{D}$ contains all sentence-aspect pairs and $y_{(s,a)}$ is the real distribution of sentiment.

3 Experimental Setup

3.1 Datasets

Our model is evaluated the performance on three benchmark datasets. The Laptop (LAP14) and Restaurant (REST14) datasets are made public from SemEval2014 ABSA challenge (Pontiki et al., 2014). Furthermore, the Twitter dataset is a collection of tweets from (Dong et al., 2014). All three datasets have three sentiment polarities: *positive*, *negative* and *neutral*. Each dataset provides aspect terms and corresponding polarities. Detailed statistics of the datasets can be found in Table 1 .

3.2 Implementation Details

The Stanford parser[2] is utilized to get syntactic dependency relations. We employ the uncased english version of the BERT model[3] in PyTorch. The dropout rate is 0.3. The number of layers of graph convolutional block is 2. Our model is trained with a batch size of 16 and uses Adam optimizer with a learning rate of $2e-5$. The coefficients μ_1 and μ_2 are set to (0.04, 0.04), (0.05, 0.06) and (0.06, 0.08) for three datasets. The hyper-parameter λ is 0.5, and ρ is 0.2. We repeat each experiment three times and average the results. We use accuracy (Acc.) and macro-f1 (F1.) as the main evaluation metrics.

4 Experimental Results

4.1 Baselines

We compare our MGFN with state-of-the-art baselines which are described as follows:

- **CDT** (Sun et al., 2019b) used GCNs to learn aspect representation over a dependency tree.

[2]https://stanfordnlp.github.io/CoreNLP/
[3]https://github.com/huggingface/transformers

Dataset	#Positve		#Negative		#Neutral	
	Train	Test	Train	Test	Train	Test
LAP14	976	337	851	128	455	167
REST14	2164	727	807	196	637	196
TWITTER	1507	172	1528	169	3016	336

Table 1: Satistics of three datasets.

- **BERT-SRC** (Devlin et al., 2019) is the vanilla BERT model for classification.
- **R-GAT** (Wang et al., 2020) designed a new aspect-oriented dependency tree and encoded the new tree by relational GAT.
- **KumaGCN** (Chen et al., 2020) combined external dependency parse graph and latent graph to generate task-specific representation.
- **DGEDT** (Tang et al., 2020) proposed a dependency graph enhanced dual-transformer network.
- **BATAE-GRU** (Wang and Wang, 2021) used an attention-based model to relate the aspect.
- **DualGCN** (Li et al., 2021b) proposed a dual-graph GCN to address disadvantages of attention and dependency tree based methods.
- **ACLT** (Zhou et al., 2021) designed an aspect-centric latent tree to shorten the distance between aspects and opinion words.
- **BERT4GCN** (Xiao et al., 2021) utilized outputs from intermediate layers of BERT and positional information to augment GCN.
- **CPA-SA** (Huang et al., 2022) designed two asymmetrical contextual position weight functions to adjust the weight of aspect.
- **IMA** (Wang et al., 2022) combined interaction matrix and global attention mechanism to measure relationships between words.
- **HGCN** (Xu et al., 2022) synthesize information from constituency tree and dependency tree to enrich the representation.

Baselines and MGFN are all BERT-based. We present the reported results of those baselines. However, for CDT method, we implement it under BERT setting using its open implementation. The source code and BERT settings of kumaGCN are not provided, so we use the results reported by ACLT in order to be fair for other models.

4.2 Overall Performance Comparison

Table 2 shows main experimental results of the baselines and our model. We can observe that:

Model	LAP14		REST14		Twitter	
	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)
BERT-SRC (Devlin et al., 2019)	78.99	75.03	84.46	76.98	73.55	72.14
CDT (Sun et al., 2019b)	79.70	75.61	86.36	80.16	77.50	<u>76.54</u>
R-GAT (Wang et al., 2020)	78.21	74.07	86.60	<u>81.35</u>	76.15	74.88
DGEDT (Tang et al., 2020)	79.80	75.60	86.30	80.00	<u>77.90</u>	75.40
KumaGCN (Chen et al., 2020)	79.57	75.61	84.91	77.22	74.33	73.42
BERT4GCN (Xiao et al., 2021)	77.49	73.01	84.75	77.11	74.73	73.76
BATAE-GRU (Wang and Wang, 2021)	78.59	74.78	84.11	76.09	74.34	72.76
ACLT (Zhou et al., 2021)	79.68	75.83	85.71	78.44	75.48	74.51
DualGCN (Li et al., 2021b)	<u>81.80</u>	<u>78.10</u>	<u>87.13</u>	81.16	77.40	76.02
CPA-SA (Huang et al., 2022)	75.18	71.5	82.64	73.38	-	-
IMA (Wang et al., 2022)	77.44	73.48	82.81	73.66	-	-
HGCN (Xu et al., 2022)	79.59	-	86.45	-	-	-
Our MGFN	**81.83**	**78.26**	**87.31**	**82.37**	**78.29**	**77.27**

Table 2: Main experimental results of aspect-based sentiment classification on three public datasets. The best results are in bold, and the second-best results are underlined.

1) Our MGFN model achieves the state-of-the-art performances over all baselines on three datasets. Compared to the state-of-the-art graph-based model DualGCN, our model makes especially 1.21% and 1.25% in terms of F1 improvements on REST14 and Twitter respectively. Our MGFN slightly outperforms DualGCN (0.16%) on LAP14 dataset. 2) The state-of-the-art latent graph based model ACLT does not outperform DualGCN, indicating that latent graph needs to be further improved. 3) The dependency parse tree based models (e.g., CDT, and DualGCN) usually outperform syntax information free models (e.g., BERT-SRC, CPA-SA), which means syntactic dependency relation information is effective. Therefore, our MGFN proposes a novel SaLG to leverage richer syntax dependency relations. 4) The KumaGCN combines latent graph and syntactic dependency graph, but has still poor performance. In contrast, our MGFN leverages affective semantic information of words to improve the experimental results successfully.

4.3 Ablation Study

We conduct an ablation study by removing modules and loss terms, shown in Table 3. We remove the syntax dependency relation label (w/o Syn. Information), which leads to performance degradation. MGFN w/o adaptive fusion gate is that we do not fuse SeG into SaLG during iterations. We observe

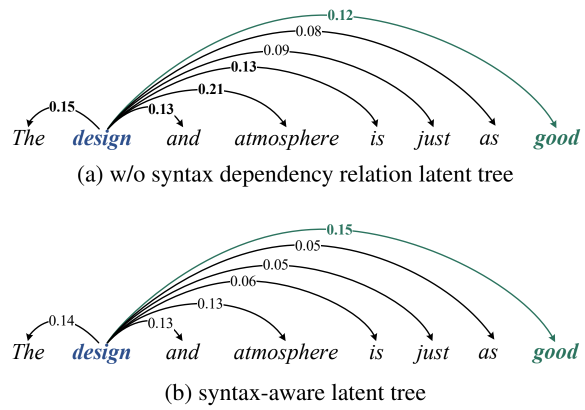


Figure 3: A review from REST14 dataset to illustrate different trees. The aspect words are in blue.

that both w/o SaLG, w/o SeG and w/o adaptive fusion gate result in performance drops, showing that adaptively integrating semantic information into SaLG improves performance of MGFN as far as possible. MGFN w/o $\mathcal{L}_r\&\mathcal{L}_s$ is we remove both root constraint strategy and affective refinement strategy, MGFN w/o $\mathcal{L}_r$ or $\mathcal{L}_s$ is we remove one of these strategies, both leading to performance drops.

5 Discuss and Analysis

5.1 Effect of Syntax-aware Latent Graph

To investigate the effect of SaLG, we utilize the latent tree w/o syntax dependency relation informa-

Model	LAP14		REST14		Twitter	
	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)	Acc.(%)	F1.(%)
Our MGFN	**81.83**	**78.26**	**87.31**	**82.37**	**78.29**	**77.27**
w/o Syn. Informaiton	81.06	76.58	86.86	81.73	77.55	76.06
w/o SaLG	80.22	76.23	86.32	79.92	76.25	75.32
w/o SeG	80.38	76.41	86.60	80.32	76.63	75.92
w/o Adaptive Fusion Gate	80.53	76.69	86.87	81.15	76.81	75.98
w/o $\mathcal{L}_r$ & $\mathcal{L}_s$	80.22	76.23	86.68	79.83	77.4	75.87
w/o $\mathcal{L}_r$	81.17	78.02	87.02	80.6	77.55	76.58
w/o $\mathcal{L}_s$	80.38	76.38	86.70	80.11	77.51	75.99

Table 3: Ablation study experimental results

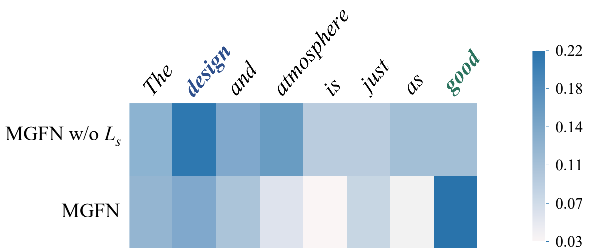
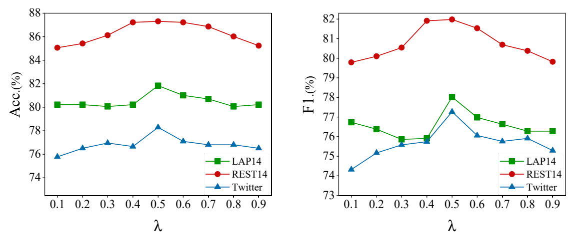
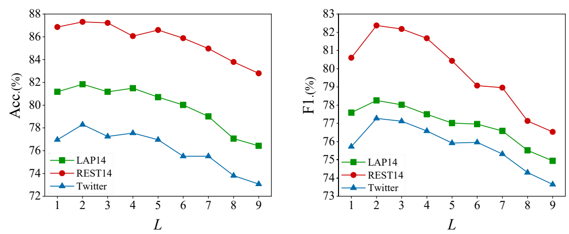


Figure 4: Attention visualization of learned latent weights by MGFN and MGFN w/o $\mathcal{L}_s$ models. "*design*" is the aspect word.



(a) (b)

Figure 5: The impact of different λ.



(a) (b)

Figure 6: The impact of the number of common graph convolutional block.

tion to compare with our novel syntax-aware latent tree, shown in Figure 3. Specifically, in Figure 3 (a), the edge weight from aspect "*design*" to opinion word "*good*" is only 0.12, while the weights to neighbour words are much higher (*e.g.* 0.15 *for "The", and* 0.21 *for "atmosphere", etc.*). However, in Figure 3 (b), the weight between "*design*" and "*good*" increases to 0.15, slightly higher than neighbour words. Utilizing syntactic dependency relation label information, aspect pays more attention to opinion word "*good*" in our SaLG.

5.2 Impact of Affective Refinement Strategy

In order to verify the effectiveness of the affective refinement strategy, we visualize the attention weight ε in Eq. (16) of the example review. In Figure 4, we observe that the MGFN w/o $\mathcal{L}_s$ model assigns higher attention on "*The*", "*and*" and "*atmosphere*" incorrectly when $\mathcal{L}_s$ is not utilized. In comparison, for our MGFN model, the aspect "*design*" can assign the highest attention on "*good*" obviously, since opinion word "*good*" contains the highest sentiment score in lexicon vector of example review.

5.3 Hype-parameter Analysis

To investigate the effect of the hype-parameter, we vary the λ from 0.1 to 0.9, shown in Figure 5. The hyper-parameter λ represents the proportion of main classification task in total objective function. From Figure 5, the performance reaches its highest when λ equals to 0.5. If λ is less than 0.5, the main task cannot be trained fully. However, if λ is more than 0.5, the proposed constraint strategies fail to work well. Therefore, it is important to set an appropriate λ to balance the performance of main classification task and two constraint strategies.

Sentence	ACLT	MGFN w/o L_s	MGFN
The [**menu**]$_{neg}$ is limited but the [**dishes**]$_{pos}$ are excellent.	(neg✓,pos✓)	(neg✓,pos✓)	(neg✓,pos✓)
For my user experience, the [**speed**]$_{pos}$ is better than the [**battery life**]$_{neg}$.	(pos✓,pos✗)	(pos✓,neg✓)	(pos✓,neg✓)
I had great interest in this restaurant due to its [**atmosphere**]$_{pos}$, but the [**service**]$_{neg}$ was disappointing.	(neg✗,neg✓)	(neu✗,neg✓)	(pos✓,neg✓)

Table 4: Case study experimental results of three different models

5.4 Impact of Number of C-GCN Blocks

To investigate the impact of number L of C-GCN blocks, we vary the L from 1 to 9, shown in Figure 6. Our model with 2 C-GCN blocks achieves the best performance. When L is less than 2, our MGFN is not enough to fully integrate semantic information from SeG into SaLG. When L is excessive, the performance of our model decreases due to vanishing gradient and over-smoothing. However, the performance of MGFN does not degrade sharply because of our control factor ω.

5.5 Case Study

We conduct a case study by classifying a few examples using different models, shown in Table 4. We use boldface in brackets to show aspects of each sentence and subscripts to indicate corresponding golden sentiment polarities. For the first sentence, aspects "*menu*" and "*dishes*" are both next to their own opinion words, so all models easily assign correct sentiment polarities. In the second sentence, aspects "*speed*" and "*battery life*" are adjacent to opinion expression "*better*". The ACLT model can not identify the dependency relation type information, which results in wrong prediction of aspect "*battery life*". Besides, for the third sentence, aspect "*atmosphere*" is closer to opinion expression "*disappointing*", which leads to incorrect predictions by ACLT and MGFN w/o $\mathcal{L}_s$ models. While our MGFN includes an affective refinement strategy and can capture the significant affective cue of true opinion expression "*great interest*".

6 Related Work

Aspect-based Sentiment analysis: Sentiment analysis is one of the most active research areas in natural language processing (Liao et al., 2021; Tang et al., 2022), and is widely studied in QA system (Ma et al., 2021), stance detection (AlDayel and Magdy, 2021; Hardalov et al., 2021), recommendation system (Aljunid and Huchaiah, 2021; Abbasi-Moud et al., 2021), and event detection (Ma et al., 2022). Aspect-based Sentiment analysis (ABSA) is first proposed by Hu and Liu (2004) to refine sentiment analysis, which aims to detect fine-grained sentiments towards different aspects. Early efforts on ABSA utilizes attention-based neural models to model semantic interactions (Wang et al., 2016; Chen et al., 2017). Some other efforts (Wang et al., 2016; Nguyen and Nguyen, 2018; Huang et al., 2021) try to explicitly establish the syntactic dependency connections between words.

Graph neural networks: Recently, Graph neural networks (GNNs) (Huang et al., 2019; Kim et al., 2019) have received growing attention and successfully used in many applications such as action recognition (Zhang et al., 2022), relation extraction (Bastos et al., 2021; Zhang et al., 2021) and scene image generation (Li et al., 2021a). Yao et al. (2019) innovatively utilized graph convolution networks (GCNs) for text classification in natural language process field. For ABSA, Zhang et al. (2019) used GCNs to encode dependency information of syntactic dependency parse tree. Tang et al. (2020) proposed a dependency graph enhanced dual-transformer network(DGEDT) to allow the dependency graph to guide the representation learning of the transformer encoder. Wang et al. (2020) constructed the aspect-oriented dependency trees by which reshaped the ordinary dependency parse tree to root it at aspect using manual rules.Li et al. (2021b) used the probability matrix with all dependency structures of input sentence from off-the-shelf dependency parser to alleviate inaccurate parse problem and integrated syntactic and semantic information.

More recently, several teams have explored to construct latent graph that can adaptively capture the relation between words of the sentence in an end-to-end fashion. Chen et al. (2020) constructed a latent graph sampled from the HardKuma distribution, and combined a dependency parse graph with it to generate task-specific representation. Zhou et al. (2021) utilized a variant of Kirchhoff's Matrix-Tree Theorem to induce the task-specific aspect-centric latent dependency tree.

7 Conclusion

In this paper, we propose an MGFN model to address the disadvantages of latent graph based models for aspect-based sentiment analysis. We construct a novel SaLG to leverage the richer syntax dependency relation label information, and adaptively fuse the semantic information from SeG into SaLG to facilitate the learning of sentiment representation. Moreover, to capture more significant affective clues from surrounding contexts, we propose an affective refinement strategy in multi-graph fusion module. This strategy can guide MGFN to pay more attention to the opinion expressions of aspects. Extensive experiments on three datasets show that our model achieves the best performance.

Limitations

Our MGFN model is designed for English datasets, thus it is only applicable to English remarks. Moreover, as we construct two graphs for every sentence and fuse the information of different kinds of graphs, the scale of graphs cannot be too large. That is, for a long text, our proposed MGFN cannot be applied to long texts.

Acknowledgements

This work was supported by the National Natural Science Foundation of China(No. 61976051), the Major Key Project of PCL (No.PCL2021A09, PCL2021A02, PCL2022A03), and Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies (2022B1212010005).

References

Zahra Abbasi-Moud, Hamed Vahdat-Nejad, and Javad Sadri. 2021. Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Syst. Appl.*

Abeer AlDayel and Walid Magdy. 2021. Stance detection on social media: State of the art and trends. *Inf. Process. Manag.*

Mohammed Fadhel Aljunid and Manjaiah Doddaghatta Huchaiah. 2021. An efficient hybrid recommendation model based on collaborative filtering recommender systems. *CAAI Trans. Intell. Technol.*, 6(4):480–492.

Anson Bastos, Abhishek Nadgeri, Kuldeep Singh, Isaiah Onando Mulang', Saeedeh Shekarpour, Johannes Hoffart, and Manohar Kaul. 2021. RECON: relation extraction using knowledge graph context in a graph neural network. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, pages 1673–1685. ACM / IW3C2.

Marouane Birjali, Mohammed Kasri, and Abderrahim Beni Hssane. 2021. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowl. Based Syst.*, 226:107134.

Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. 2021. Beyond low-frequency information in graph convolutional networks. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI*.

Erik Cambria, Yang Li, Frank Z. Xing, Soujanya Poria, and Kenneth Kwok. 2020. Senticnet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management*.

Chenhua Chen, Zhiyang Teng, and Yue Zhang. 2020. Inducing target-specific latent structures for aspect sentiment classification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP*, pages 5596–5607. Association for Computational Linguistics.

Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 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, NAACL*.

Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014. Adaptive recursive neural network for target-dependent twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL*.

Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2021. Cross-domain label-adaptive stance detection. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP*.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004*.

Bo Huang, Ruyan Guo, Yimin Zhu, Zhijun Fang, Guohui Zeng, Jin Liu, Yini Wang, Hamido Fujita, and Zhicai Shi. 2022. Aspect-level sentiment analysis with aspect-specific context position information. *Knowl. Based Syst.*, 243.

Lianzhe Huang, Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2019. Text level graph neural network for text classification. 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*.

Yuan Huang, Zhixing Li, Wei Deng, Guoyin Wang, and Zhimin Lin. 2021. D-BERT: incorporating dependency-based attention into BERT for relation extraction. *CAAI Trans. Intell. Technol.*, 6(4):417–425.

Jongmin Kim, Taesup Kim, Sungwoong Kim, and Chang D. Yoo. 2019. Edge-labeling graph neural network for few-shot learning. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*.

Rongjie Li, Songyang Zhang, Bo Wan, and Xuming He. 2021a. Bipartite graph network with adaptive message passing for unbiased scene graph generation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*.

Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard H. Hovy. 2021b. Dual graph convolutional networks for aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, ACL*, pages 6319–6329. Association for Computational Linguistics.

Qing Liao, Heyan Chai, Hao Han, Xiang Zhang, Xuan Wang, Wen Xia, and Ye Ding. 2021. An integrated multi-task model for fake news detection. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–1.

Kaixin Ma, Filip Ilievski, Jonathan Francis, Yonatan Bisk, Eric Nyberg, and Alessandro Oltramari. 2021. Knowledge-driven data construction for zero-shot evaluation in commonsense question answering. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI*.

Xiaobo Ma, Yongbin Liu, and Chunping Ouyang. 2022. Capturing semantic features to improve chinese event detection. *CAAI Trans. Intell. Technol.*, 7(2):219–227.

Huy-Thanh Nguyen and Minh-Le Nguyen. 2018. Effective attention networks for aspect-level sentiment classification. In *10th International Conference on Knowledge and Systems Engineering, KSE*.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014*.

Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019a. 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*.

Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019b. 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*.

Hao Tang, Donghong Ji, Chenliang Li, and Qiji Zhou. 2020. Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL*, pages 6578–6588. Association for Computational Linguistics.

Jingyao Tang, Yun Xue, Ziwen Wang, Shaoyang Hu, Tao Gong, Yinong Chen, Haoliang Zhao, and Luwei Xiao. 2022. Bayesian estimation-based sentiment word embedding model for sentiment analysis. *CAAI Trans. Intell. Technol.*, 7(2):144–155.

Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. Relational graph attention network for aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL*, pages 3229–3238. Association for Computational Linguistics.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016. Recursive neural conditional random fields for aspect-based sentiment analysis. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP*.

Xiaodi Wang, Xiaoge Pan, Tian Yang, Jianhua Xie, and Mingwei Tang. 2022. Aspect-based sentiment analysis using interaction matrix and global attention neural network. *The Computer Journal*.

Yuan Wang and Qian Wang. 2021. BATAE-GRU: attention-based aspect sentiment analysis model. In *ISEEIE 2021: International Symposium on Electrical, Electronics and Information Engineering, Seoul Republic of Korea, February 19 - 21, 2021*.

Zeguan Xiao, Jiarun Wu, Qingliang Chen, and Congjian Deng. 2021. BERT4GCN: using BERT intermediate layers to augment GCN for aspect-based sentiment classification. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP*.

Lvxiaowei Xu, Xiaoxuan Pang, Jianwang Wu, Ming Cai, and Jiawei Peng. 2022. Learn from structural scope: Improving aspect-level sentiment analysis with hybrid graph convolutional networks. *CoRR*, abs/2204.12784.

Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Graph convolutional networks for text classification. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI*.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019. 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, EMNLP-IJCNLP*.

Jiaxu Zhang, Gaoxiang Ye, Zhigang Tu, Yongtao Qin, Qianqing Qin, Jinlu Zhang, and Jun Liu. 2022. A spatial attentive and temporal dilated (SATD) GCN for skeleton-based action recognition. *CAAI Trans. Intell. Technol.*, 7(1):46–55.

Ningyu Zhang, Xiang Chen, Xin Xie, Shumin Deng, Chuanqi Tan, Mosha Chen, Fei Huang, Luo Si, and Huajun Chen. 2021. Document-level relation extraction as semantic segmentation. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI*.

Pinlong Zhao, Linlin Hou, and Ou Wu. 2020. Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification. *Knowl. Based Syst.*, 193:105443.

Yuxiang Zhou, Lejian Liao, Yang Gao, Zhanming Jie, and Wei Lu. 2021. To be closer: Learning to link up aspects with opinions. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP*.