

Beta Distribution Guided Aspect-aware Graph for Aspect Category Sentiment Analysis with Affective Knowledge

Bin Liang^{1,2*}, Hang Su^{1,2*}, Rongdi Yin^{1,2}, Lin Gui^{3†}, Min Yang⁴,
Qin Zhao¹, Xiaoqi Yu⁵, and Ruifeng Xu^{1,6†}

¹School of Computer Science and Technology,
Harbin Institute of Technology, Shenzhen, China

²Joint Lab of HITSZ and China Merchants Securities, Shenzhen, China

³Department of Computer Science, University of Warwick, Coventry, U.K

⁴SIAT, Chinese Academy of Sciences, Shenzhen, China

⁵China Merchants Securities Co., Ltd., ⁶Peng Cheng Laboratory, Shenzhen, China

{bin.liang, suhang}@stu.hit.edu.cn, rongdiyin@gmail.com

{zhaqin, xuruifeng}@hit.edu.cn, lin.gui@warwick.ac.uk

min.yang@siat.ac.cn, yuxiaoqi@cmschina.com.cn

Abstract

In this paper, we investigate the Aspect Category Sentiment Analysis (ACSA) task from a novel perspective by exploring a Beta Distribution guided aspect-aware graph construction based on external knowledge. That is, we are no longer entangled about how to laboriously search the sentiment clues for coarse-grained aspects from the context, but how to preferably find the words highly sentiment-related to the aspects in the context and determine their importance based on the public knowledge base, so as to naturally learn the aspect-related contextual sentiment dependencies with these words in ACSA. To be specific, we first regard each aspect as a pivot to derive aspect-aware words that are highly related to the aspect from external affective commonsense knowledge. Then, we employ Beta Distribution to educe the aspect-aware weight, which reflects the importance to the aspect, for each aspect-aware word. Afterward, the aspect-aware words are served as the substitutes of the coarse-grained aspect to construct graphs for leveraging the aspect-related contextual sentiment dependencies in ACSA. Experiments on 6 benchmark datasets show that our approach significantly outperforms the state-of-the-art baseline methods.

1 Introduction

Aspect category sentiment analysis (ACSA) aims to detect the sentiment polarity for a coarse-grained aspect category from a given sentence. Different from the task of target-dependent or aspect term sentiment analysis, whose target or term explicitly occurs in the sentence, the aspect category in

Sentence: This place is pricey, but the pizza is yummy.
Aspect category: RESTAURANT#PRICES FOOD#QUALITY

Figure 1: Example of aspect-related sentiment expressions. Words in the same color represent sentiment-related words.

ACSA does not necessarily occur in the sentence. Here, the aspect category (hereinafter also referred to as aspect) generally consists of an entity E and an attribute A (i.e. $E\#A$) or only an entity E . As shown in Figure 1, in sentence “*This place is pricey, but the pizza is yummy.*”, there are two aspects mentioned in the sentence: “*RESTAURANT#PRICES*” (negative) and “*FOOD#QUALITY*” (positive).

Many existing research efforts focus on ACSA with deep learning methods to attend the significant information for the aspect category in sentiment prediction (Wang et al., 2016; Cheng et al., 2017; Liang et al., 2019a,b; Li et al., 2020a; Chen et al., 2020; Li et al., 2020b; Liang et al., 2020a). Despite promising progress made by existing methods, they are generally entangled about how to search the sentiment clues of coarse-grained aspects from the context. However, making sense of the aspect-oriented sentiment words from the context purely based on the implicit aspects is a daunting task. This mostly due to 1) aspect categories generally do not manifest in the context, and 2) multiple aspects and sentiment polarities may be mentioned in the same context. On the contrary, we can exploit the aspect-related words that explicitly occur in the sentence to model the contextual sentiment information for the aspect. As the examples shown in Figure 1, there are some aspect-related words (e.g. “*place*”, “*pricey*”, “*pizza*” and “*yummy*”) in the sentence, allowing us to explicitly leverage the

* The first two authors contribute equally to this work

† Corresponding Authors

sentiment dependencies with these words for identifying the sentiment polarities of the aspects.

Motivated by this, we investigate the ACSA task from a novel perspective by proposing to construct aspect-aware graph(s) for the context with respect to the corresponding aspect. More concretely, we regard the distinct aspect as the distinct pivot and then search the aspect-related words from external knowledge, called aspect-aware words, which are served as the substitutes of the coarse-grained aspect to construct graph of the context for the specific aspect. That is, external knowledge is deployed as a bridge between implicit aspect category and the context, so as to skillfully and actively build connections between highly aspect-related context words and the specific aspect by means of a graph construction. In addition, inspired by many previous graph-based methods (Yao et al., 2019; Qin et al., 2020; Liang et al., 2020b; Qin et al., 2021b,a; Zhang et al., 2021; Liang et al., 2021a), weights of edges in a graph are important for graph information aggregation. Moreover, based on our empirical study (as shown in Figure 3 and 4), the contributions of aspect-aware words to the aspect are obviously different. For example, the aspect-aware word “*place*” is more important than “*pizza*” to the aspect entity “*RESTAURANT*”. Following that the main challenge of the idea evolves into how to determine the importance of aspect-aware words for the corresponding aspect, which can be leveraged as the weights of edges in a graph for learning the sentiment clues of the aspect.

In the light of the knowledge base, vividly, a word can connect or not to an aspect via various routes, the successful connection probability (corresponding to the weight of an edge in a graph) can be naturally regard as a Binomial Distribution. We hence examine the weights of edges via modeling all the probabilities of successful connection possibility based on the prior knowledge (routes and connection information) of external knowledge by means of Beta Distribution (Gupta and Nadarajah, 2004), which is the Conjugate prior distribution of Binomial Distribution. In this way, all the probabilities of aspect-aware words that connecting to the aspect could be investigated, so as to determine the optimum confidence probability (weight) of the aspect-aware word, called aspect-aware weight. Subsequently, we construct aspect-aware graph(s) for each context with respect to the aspect based on the aspect-aware words paired with their weights.

Based on it, an aspect-aware graph convolutional network (AAGCN) structure is proposed to draw contextual sentiment dependencies to the aspect for ACSA. The main contributions of our work are summarized as follows:

(i) The ACSA task is approached from a novel perspective that learning how to find the aspect-aware words that highly related to the aspect and educe their importance to the aspect, so as to construct a graph with these words for learning the contextual sentiment features in ACSA.

(ii) A novel scenario of modeling all the importance probabilities of aspect-aware words with Beta Distribution is deployed to educe the aspect-aware weights for constructing the knowledge enhanced aspect-aware graph.

(iii) An aspect-aware graph convolutional network is proposed to draw contextual sentiment dependencies to the aspect for sentiment detection and achieves state-of-the-art performance.

2 Related Work

Previous studies in ACSA task largely pay attention to straightforwardly extract the contextual sentiment for coarse-grained aspect categories. Wang et al. (2016) proposed an attention-based LSTM model for selectively attending the regions of the context representations. Xue and Li (2018) exploited a gated convolutional neural network to selectively extract aspect-specific sentiment information for sentiment prediction. Xing et al. (2019) explored an aspect-aware LSTM to incorporate aspect information into LSTM cells for ACSA. In multi-task learning methods, Li et al. (2020b) adopted aspect category detection task to aggregate the sentiment for the aspect from the context. Chen et al. (2020) modeled document-level sentiment preference with cooperative graph attention networks for document-level ACSA. Cai et al. (2020) explored a hierarchical graph convolutional network to model the inner- and inter-relations for aspects in sentiment prediction.

In addition, to enhance the learning ability of the model, there are a series of studies that incorporate external knowledge into the framework (Ma et al., 2018; Zhang et al., 2020; Tian et al., 2020; Tong et al., 2020; Liang et al., 2021b). Among them, Tian et al. (2020) modeled sentiment information at the word, polarity, and aspect level into pre-trained sentiment representation in sentiment analysis based on the automatically-mined

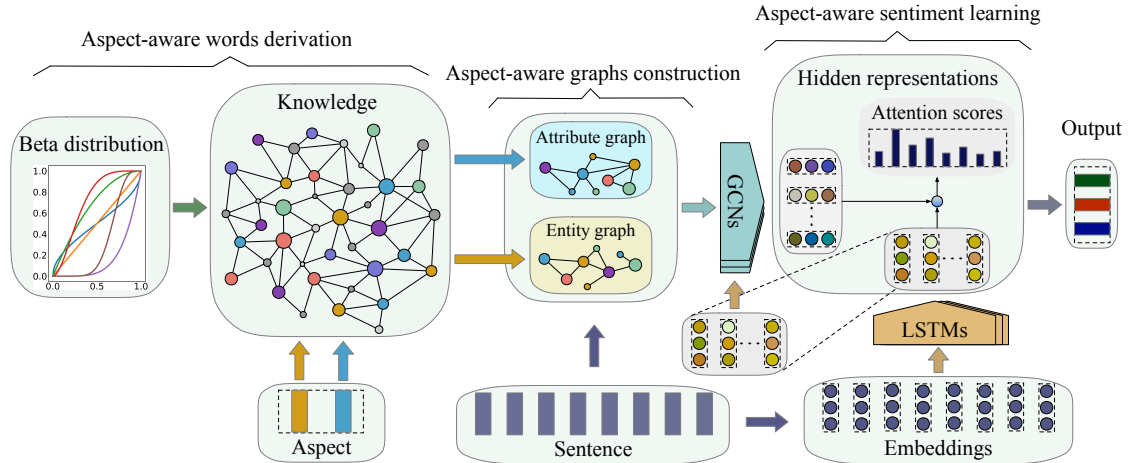


Figure 2: The architecture of our proposed AAGCN.

knowledge. Zhang et al. (2020) exploited semantic and emotion lexicons as a bridge to enable knowledge transfer across different targets in cross-target stance detection. In targeted aspect-based sentiment analysis, Ma et al. (2018) exploited affective knowledge to extend the classic LSTM cell for simultaneously learning a target-specific attention and a global attention.

3 Methodology

In this section, we describe our proposed aspect-aware graph convolutional network (AAGCN) in detail. As illustrated in Figure 2, our proposed model consists of three primary components: 1) Aspect-aware words derivation, which generates a distinct series of affective words for the distinct aspect from external knowledge. 2) Aspect-aware graphs construction, which constructs aspect-aware graphs of the context based on aspect-aware words. 3) Aspect-aware sentiment learning, which extracts the aspect-related sentiment dependencies based on aspect-aware graphs and context representations.

3.1 Task Definition

Given a sentence s consists of n words $s = \{w_1, w_2, \dots, w_n\}$ and the corresponding aspect a , which may not occur in $(\{w_i\} | i = 1, 2, \dots, n)$. The goal of aspect category sentiment analysis is to detect the sentiment polarity (i.e. *Positive*, *Negative*, or *Neutral*) of the aspect from the context. Here, each aspect may consist of an entity E and an attribute A (i.e. $E\#A$) or only an entity E .

3.2 Aspect-aware Words Derivation

To construct contextual sentiment dependency graph for the aspect that does not occur in the sen-

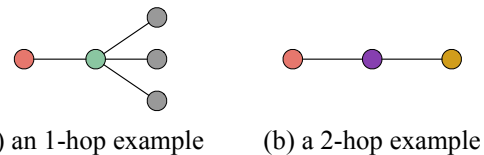


Figure 3: Examples of aspect-aware words with different hops.

tence, we explore a novel scenario that regarding an aspect as the pivot and deriving the aspect-aware words by searching the words that are highly associated with the aspect from the external affective knowledge within a certain number of hops. To be specific, if words contain direct relations with the aspect, then these words are the 1-hop aspect-aware words. Correspondingly, if words contain relations with the 1-hop aspect-aware words, then those words are the 2-hops aspect-aware words, etc. In addition, we seek the aspect-aware words for the entity E and the attribute A respectively if an aspect consists of $E\#A$ since the roles of E and A are generally different in sentiment detection.

In this scenario, intuitively, the main challenge is to determine the affective importance of each aspect-aware word with respect to the aspect. Overall, the hop number is the roughly important impact. However, as shown in Figure 3 (a) and (b), the yellow dot with 2-hops, which contains only a unique link, is more important than the green one that simultaneously connects to many other irrelevant words. For each word it either connects to the aspect within κ -hop or not, there is a potential Beta Distribution for each aspect-aware word that reveals the distribution of the correlation degree to the aspect. Thus based on the priori knowledge

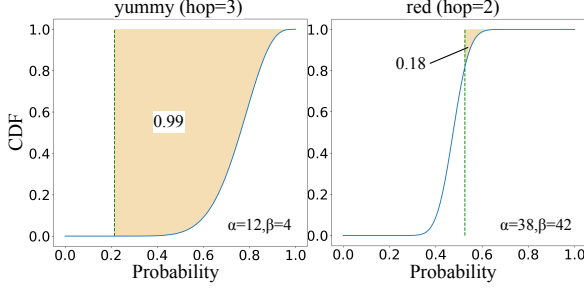


Figure 4: Examples of aspect-aware weight probability distributions. CDF=Cumulative Distribution Function.

learned by the external knowledge, we employ Beta Distribution, which is generally adopted to model all success probabilities of an experiment, to educe the importance $\rho(w_i)$ of each aspect-aware word:

$$\rho(w_i) = 1 - \text{CDF}(f(\mu_i; \alpha, \beta)) \quad (1)$$

$$\mu_i = \gamma_1 \frac{C_i^a - C_i^s}{C_i^a} + \gamma_2 \frac{N^\kappa}{N} \quad (2)$$

where $\text{CDF}(f(\mu_i; \alpha, \beta))$ represents the Cumulative Distribution of $f(\mu_i; \alpha, \beta)$. Here μ_i represents the unrelated probability of the aspect-aware word w_i towards the aspect, C_i^a is the neighbor count of w_i in the knowledge and C_i^s is the count of aspect-aware neighbors. N^κ is the vocabulary size of κ -hop aspect-aware words and N is the vocabulary size of the whole corpus. γ_1 and γ_2 are the coefficients to control the influence of the unrelated neighbors and the hop number. That is, we consider the influence of both the unrelated neighbors and the hop number when deriving the aspect-aware weight. Since as the examples depicted in Figure 4, the aspect-aware word “yummy” is more important than “red” with respect to the aspect “food”, although its hop number is greater. $f(\theta; \alpha, \beta)$ denotes the Beta Distribution of all importance probabilities θ , which is defined as:

$$f(\theta; \alpha, \beta) \triangleq \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)} \quad (3)$$

$$B(\alpha, \beta) \triangleq \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta \quad (4)$$

where $B(\cdot)$ is Beta function for normalization. Here α and β denote the parameters of the Beta Distribution towards the aspect which are learned by the prior knowledge from the external knowledge:

$$\alpha = C_i^s + 1, \quad \beta = C_i^a - C_i^s + 1 \quad (5)$$

Based on it, we can derive a decent aspect-aware weight for each aspect-aware word. In addition, we

set the aspect-aware weights of the aspect itself and each irrelevant word as 1 and 0 respectively.

3.3 Aspect-aware Graph Construction

In this section, we describe the novel solution of constructing contextual dependency graphs with respect to the aspects, granted that aspects do not occur in the sentence. Based on the aspect-aware words and their aspect-aware weights, we compute the edge weight of each word pair of the aspect-aware graph as follow:

$$A_{i,j} = \rho(w_i) + \rho(w_j) \quad (6)$$

Here inspired by many previous graph-based studies (Zhang et al., 2019; Huang and Carley, 2019; Liang et al., 2020b), we also employ dependency tree of the sentence to better capture the syntactical relations¹. That is, we add 1 to the edge weight of $A_{i,j}$ if w_i and w_j contain dependency in the dependency tree of the sentence. Then we construct the undirected graph to enrich the affective and dependency information: $A_{i,j} = A_{j,i}$, and also set a self-loop for each word: $A_{i,i} = 1$.

3.4 Aspect-aware Sentiment Learning

For each sentence, we first retrieve the embedding of each word in the sentence from the embedding lookup table $\mathbf{V} \in \mathbb{R}^{m \times N}$. Thus for a sentence with n words, we can get the corresponding embedding matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$, where $\mathbf{x}_i \in \mathbb{R}^m$ is the word embedding of w_i , which are fine-tuned during the training process. m is the dimension of the embedding. Afterward, the embedding matrix \mathbf{X} is fed as input into the bidirectional LSTM (Bi-LSTM) layers to derive the hidden contextual representations of the sentence:

$$\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n\} = \text{Bi-LSTM}(\mathbf{X}) \quad (7)$$

where $\mathbf{h}_t \in \mathbb{R}^{2m}$ represents the hidden representation at time step t derived by the Bi-LSTM layers.

Based on it, we feed the aspect-aware graph(s) of the sentence and the hidden contextual representations \mathbf{H} into the aspect-aware GCN to draw contextual sentiment dependencies to the aspect. For the aspect that consists of $E\#A$, we employ a novel interactive GCN block to capture the potential interaction between entity and attribute. Each node in the l -th GCN block is updated according to the hidden representations of its neighborhoods in

¹We employ spaCy toolkit to derive dependency tree of the sentence: <https://spacy.io/>.

the adjacency matrices of entity and attribute graph, the process is defined as:

$$\mathbf{f}_i^l = \text{ReLU}(\tilde{\mathbf{A}}_i^e \mathbf{g}_i^{l-1} \mathbf{W}_e^l + \mathbf{b}_e^l) \quad (8)$$

$$\mathbf{g}_i^l = \text{ReLU}(\tilde{\mathbf{A}}_i^a \mathbf{f}_i^l \mathbf{W}_a^l + \mathbf{b}_a^l) \quad (9)$$

where \mathbf{g}^{l-1} is the hidden representation evolved from the preceding GCN block. $\tilde{\mathbf{A}}$ is a normalized symmetric adjacency matrix:

$$\tilde{\mathbf{A}}_i = \mathbf{A}_i / (\mathbf{E}_i + 1) \quad (10)$$

where $\mathbf{E}_i = \sum_{j=1}^n A_{i,j}$ is the degree of \mathbf{A}_i . Here, the original input nodes of the first GCN block are retrieved from the hidden representations learned by Bi-LSTM layers, i.e. $\mathbf{g}^0 = \mathbf{H}$. In addition, for the aspect that only consists of E , the aspect-aware GCN updates with Eq. (8). Then inspired by (Zhang et al., 2019), we adopt a retrieval-based attention mechanism to capture the significant contextual aspect-related sentiment clue:

$$\zeta_t = \sum_{i=1}^n \mathbf{h}_t^\top \mathbf{g}_i^L, \quad \eta_t = \frac{\exp(\zeta_t)}{\sum_{i=1}^n \exp(\zeta_i)} \quad (11)$$

Hence, the final representation of the aspect-aware sentiment features is formulated as follow:

$$\mathbf{y} = \text{softmax}(\mathbf{W}_o \mathbf{r} + \mathbf{b}_o) \quad (12)$$

$$\mathbf{r} = \sum_{t=1}^n \eta_t \mathbf{h}_t \quad (13)$$

where $\text{softmax}(\cdot)$ is the softmax function to obtain the output distribution.

3.5 Model Training

The objective of our task is to train the classifiers by minimizing the cross-entropy loss between predicted and ground-truth distribution:

$$\mathcal{L} = - \sum_{i=1}^S \sum_{j=1}^C \hat{y}_i^j \cdot \log(y_i^j) + \lambda \|\Theta\|_2 \quad (14)$$

Where S is the training size, C is the number of classes. \hat{y} is the ground-truth distribution of sentiment. λ is the weight of the L_2 regularization term. Θ denotes all trainable parameters.

4 Experiments

4.1 Dataset and Experiment Setting

We conduct experiments on 6 benchmark datasets to verify the effectiveness of our proposed model².

²We remove sentences contain no aspect and aspects express conflict sentiment polarities in the same sentence.

Dataset		Positive	Negative	Neutral	
E#A	REST15	Train	1058	344	49
		Test	400	319	42
	LAP15	Train	1101	763	106
		Test	540	328	79
REST16	Train	1460	661	95	
	Test	506	187	42	
LAP16	Train	1634	1081	188	
	Test	479	272	46	
E	REST14	Train	2179	839	500
		Test	657	222	94
	MAMS	Train	1929	2084	3077
		Test	245	263	393
Dev	241	259	388		

Table 1: Statistics of the experimental datasets.

In which, aspects from Semeval 2015 (Pontiki et al., 2015) (REST15 and LAP15), and Semeval 2016 (Pontiki et al., 2016) (REST16 and LAP16) consist of E#A. i.e. two restaurant domain datasets (REST15 and REST16) and two laptop domain datasets (LAP15 and LAP16). To show the generalizability of coping with aspects that consist of E, we conduct experiments on a dataset from Semeval 2014 (Pontiki et al., 2014) (REST14), and a dataset with multiple aspects multiple sentiments in one sentence (MAMS) (Jiang et al., 2019). Each sample consists of the sentence, aspect, and the sentiment polarity towards the aspect. The statistics of the datasets are shown in Table 1. Following (Cai et al., 2020), for the datasets without development sets, we randomly select 10% of the training set as the development data to tune the hyper-parameters³.

For non-BERT models, we use GloVe (Pennington et al., 2014) to initialize each word into 300-dimensional embedding. The hidden vector dimension is 300. The GCN blocks number is 2. The coefficients of γ_1 and γ_2 are 0.4 and 0.6, λ is 0.00001, which are the optimal hyper-parameters in the pilot studies. Adam is utilized as the optimizer with a learning rate of 0.001 and a mini-batch of 16. We apply a dropout of 0.3 after the embedding layer. For BERT-based models, we use the pre-trained uncased BERT-base (Devlin et al., 2019) with 768-dimensional embedding⁴, and the learning rate is 0.00002. SenticNet (Cambria et al., 2020), which contains affective commonsense relations between words, is employed to derive aspect-aware words in this work. We set the max hop number to 5. The

³The source code of this work is released at <https://github.com/BinLiang-NLP/AAGCN-ACSA>.

⁴Since the baselines are BERT-base based, we construct our model based on BERT-base for a fair comparison.

Model	REST15		LAP15		REST16		LAP16		REST14		MAMS	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
TC-LSTM (Tang et al., 2016)	76.39 [‡]	58.70 [‡]	74.13	60.08	83.55 [‡]	60.26 [‡]	77.12	58.23	80.27	65.59	-	-
ATAE-LSTM (Wang et al., 2016)	78.48 [‡]	59.77 [‡]	75.32	63.02	84.19 [‡]	62.89 [‡]	78.39	62.45	82.12 [‡]	70.15 [‡]	70.63 [§]	-
GCAE (Xue and Li, 2018)	77.55 [‡]	57.43 [‡]	75.30	62.87	84.66 [‡]	60.89 [‡]	78.27	61.03	81.34 [‡]	67.51	72.10 [§]	-
AA-LSTM (Xing et al., 2019)	-	-	-	-	-	-	-	-	83.45 [‡]	75.00 [‡]	-	-
CapsNet (Jiang et al., 2019)	78.14	61.57	74.71	61.75	83.79	61.36	76.31	61.07	81.17 [‡]	69.63	73.99 [§]	-
AS-Capsules (Wang et al., 2019)	-	-	-	-	-	-	-	-	82.18 [‡]	-	75.12 [‡]	-
GIN (Yin et al., 2020)	81.17 [‡]	62.38 [‡]	75.93	63.18	87.05 [‡]	65.03 [‡]	78.92	62.93	-	-	-	-
MIMLLN (Li et al., 2020b)	78.27	60.59	75.30	61.39	85.76	63.52	78.57	62.63	81.60 [‡]	71.25	76.43 [‡]	-
AAGCN-one (w/o distribution)	81.22	63.70	77.16	62.53	85.87	66.93	78.50	63.35	82.96	74.90	76.16	74.82
AAGCN-hop (w/o distribution)	81.31	64.08	76.92	62.97	86.12	66.80	79.06	63.61	83.19	75.15	76.45	75.21
AAGCN-BD (binomial)	81.39	64.78	77.26	63.45	86.67	67.43	79.22	64.10	83.75	75.30	76.88	75.54
AAGCN-PD (poisson)	81.45	64.27	76.73	63.68	87.22	67.80	79.39	63.97	83.92	75.67	77.02	75.83
AAGCN-GD (gamma)	82.11	65.30	77.58	64.62	87.83	69.57	80.50	64.82	84.06	76.01	77.13	76.15
AAGCN-c (ConceptNet)	82.36	66.82	78.73	65.12	88.02	70.67	81.02	65.48	84.72	76.17	77.42	76.54
AAGCN (ours)	82.79	67.43	80.02	65.87	88.32	72.55	81.76	65.96	85.86	77.36	77.52	76.89
BERT (Devlin et al., 2019)	82.41 [‡]	64.35 [‡]	81.57	66.23	88.60 [‡]	73.62 [‡]	82.18	64.33	87.48 [‡]	77.53	78.29 [§]	-
BERT-QA (Sun et al., 2019)	82.53	64.89	82.73	62.37	89.83	72.86	81.93	62.13	87.52 [‡]	78.14	79.13 [‡]	-
CapsNet-BERT (Jiang et al., 2019)	81.89	61.85	82.19	59.75	86.50	62.12	80.53	61.03	86.56 [‡]	78.84	79.46 [§]	-
GIN-BERT (Yin et al., 2020)	83.96 [‡]	66.03 [‡]	82.97	65.29	89.47 [‡]	74.87 [‡]	82.76	63.77	-	-	-	-
MIMLLN-BERT (Li et al., 2020b)	82.76	65.10	82.98	62.36	88.12	73.05	82.57	63.26	89.25 [‡]	79.03	81.20 [‡]	-
Hier-GCN-BERT (Cai et al., 2020)	-	64.23 [‡]	-	62.13 [‡]	-	74.55 [‡]	-	54.15 [‡]	-	-	-	-
AAGCN-BERT-c (ConceptNet)	87.18	71.02	85.23	71.52	92.05	79.86	84.27	68.75	91.02	81.09	81.28	80.79
AAGCN-BERT (ours)	87.92	71.75	85.82	72.39	92.83	80.77	85.24	69.68	91.50	82.52	81.93	81.36

Table 2: Main experimental results (%). Acc. represents accuracy, F1 represents Macro-F1 score. Best results are in bold face. The results with ‡ are retrieved from (Yin et al., 2020), with † are retrieved from (Li et al., 2020b), with ‡ is retrieved from (Xing et al., 2019), with † are retrieved from (Cai et al., 2020), with § are retrieved from (Jiang et al., 2019).

reported results are averaged scores of 10 runs to obtain statistically stable results.

4.2 Comparison Models

We compare our proposed AAGCN with various models, including (1) Non-BERT ACSA models: TC-LSTM (Tang et al., 2016), ATAE-LSTM (Wang et al., 2016), GCAE (Xue and Li, 2018), AA-LSTM (Xing et al., 2019), CapsNet (Jiang et al., 2019), AS-Capsules (Wang et al., 2019), GIN (Yin et al., 2020), MIMLLN (Li et al., 2020b). (2) BERT-based models: BERT (Devlin et al., 2019), BERT-QA (Sun et al., 2019), CapsNet-BERT (Jiang et al., 2019), CoGAN (Chen et al., 2020), GIN-BERT (Yin et al., 2020), MIMLLN-BERT (Li et al., 2020b), Hier-GCN-BERT (Cai et al., 2020).

We also provide various variants of our proposed AAGCN:

(1) To verify the effectiveness of our proposed model based on different pre-trained models, we provide AAGCN with GloVe (AAGCN) and BERT (AAGCN-BERT). AAGCN-BERT takes “[CLS] sentence [SEP] aspect [SEP]” as input.

(2) To show the generalizability of our method, another external knowledge (ConceptNet (Speer

et al., 2017)), which contains concept relations between words, is employed to produce aspect-aware words. Then two comparison models are derived, i.e. AAGCN-c and AAGCN-BERT-c.

(3) To evaluate the significance of the Distribution exploited in our proposed method, we design two variants of our model without Distribution. That is, “AAGCN-one” and “AAGCN-hop”, whose aspect-aware weights are respectively computed as $\rho(w_i) = 1$ and $\rho(w_i) = \frac{1}{\kappa_i}$.

(4) To demonstrate the effectiveness of Beta Distribution for determining aspect-aware weights, we also perform other three related Distributions with the proposed AAGCN. Including Binomial Distribution (AAGCN-BD), whose aspect-aware weight is defined as $\rho(w_i) = 1 - \sum_{r=0}^{C_i^a - C_i^s} \binom{C_i^a}{r} \mu_i^r (1 - \mu_i)^{C_i^a - r}$, Poisson Distribution (AAGCN-PD), whose aspect-aware weight is defined as $\rho(w_i) = 1 - \sum_{r=0}^{C_i^a - C_i^s} \frac{\mu_i^r}{r!} e^{-\mu_i}$, and Gamma Distribution (AAGCN-GD), whose aspect-aware weight is defined as $\rho(w_i) = 1 - \text{CDF}(\text{gamma}(\mu_i; \alpha, \beta))$.

We also set several varieties of our proposed AAGCN to analyze the impact of different components in the ablation study. “w/o $\rho + \mathcal{D}$ ” denotes constructing fully connected graph for each sen-

Model	REST15		LAP15		REST16		LAP16		REST14		MAMS		
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	
AAGCN	w/o $\rho+\mathcal{D}$	72.17	55.67	71.26	58.47	77.32	58.43	73.80	58.39	76.03	60.34	70.14	69.79
	w/o ρ	73.60	57.17	72.83	59.51	79.20	60.08	74.66	60.17	78.21	61.53	71.28	70.53
	w/o \mathcal{D}	82.11	66.76	79.22	64.75	88.19	71.03	80.62	65.19	85.01	76.67	75.83	75.62
	complete	82.79	67.43	80.02	65.87	88.32	72.55	81.76	65.96	85.86	77.36	77.52	76.89

Table 3: Experimental results of ablation study.

tence, that is, each word pair contains an edge. “w/o ρ ” denotes without aspect-aware words and “w/o \mathcal{D} ” denotes without dependency tree.

4.3 Experiment Results

As shown in Table 2, the experimental results on 6 datasets demonstrate that our proposed model performs consistently better than the comparison models for both non-BERT and BERT-based models and for both $E\#A$ and E aspects. This verifies the effectiveness of our proposed model in ACSA.

Compared with models without employing Distributions to derive aspect-aware weights, the performance is overall improved in any distribution. This denotes that exploring Distributions to model the successful connection probability between words and the corresponding aspect is more adaptive to derive more valuable aspect-aware weights from external knowledge. In addition, the results produced by different distributions show that our proposed AAGCN, which explores Beta Distribution to determine aspect-aware weights, outstandingly outperforms several related distributions. This implies that deploying Beta Distribution to model all the probabilities of successful connection probability for aspect-aware words based on the priori knowledge learned from external knowledge derives more sound aspect-aware weights and leads to an improved ACSA performance.

For different external knowledge scenarios, both AAGCN and AAGCN-c perform overall better than the baselines, which demonstrates the generalizability of our proposed method in deriving aspect-aware words. In addition, compared with models based on ConceptNet, models with SenticNet reveal considerable superiorities for both non-BERT and BERT-based conditions. This indicates that SenticNet, which contains affective relations can advance the model to leverage sentiment information and achieves better performance in ACSA.

4.4 Ablation Study

To investigate the impact of different components in our proposed model bring to the performance,

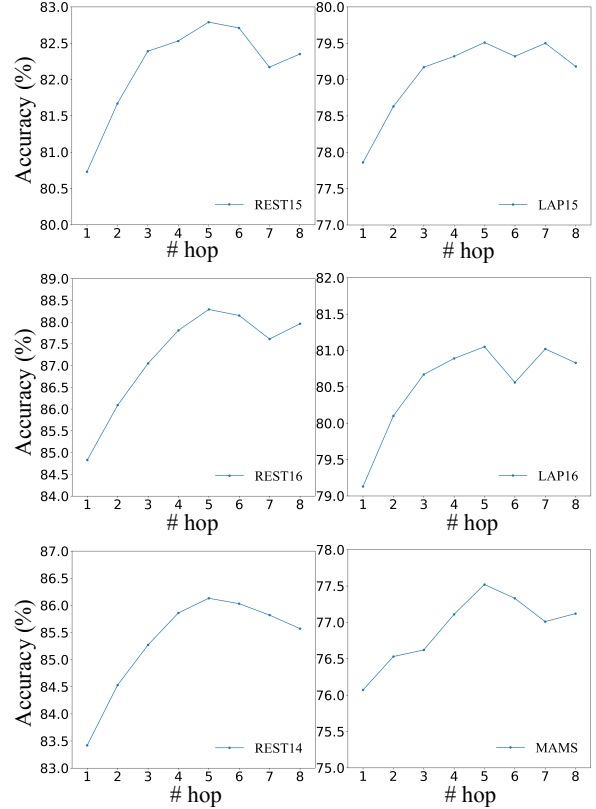


Figure 5: Performance of using different numbers of hop on 6 datasets.

we conduct an ablation study and report the results in Table 3. Note that both fully connected graph and removal of the aspect-aware words reduces the performance seriously. This verifies the significance and effectiveness of recognizing aspect-aware words from the context for constructing graph in ACSA task. Additionally, model that without employing dependency tree leads to slightly poor performance, which implies that incorporating syntactical relations into the graph can further lead to the improved ACSA performance.

4.5 Impact of Hop Numbers

To investigate the impact of different hop numbers when deriving aspect-aware words from external knowledge, we vary them from 1 to 8 and report the results in Figure 5. Note that as the hop number increases from 1 to 5 the performance

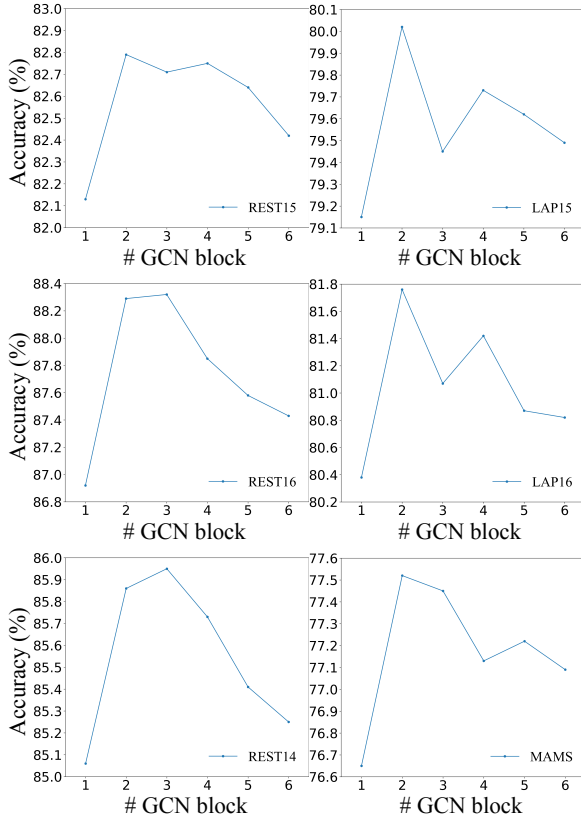


Figure 6: Performance of using different numbers of GCN blocks on 6 datasets.

improves steadily on all datasets, and the curves erratically fluctuate when the hop is greater than 5. This implies that the significant learning advantage brought by aspect-aware graph relies on an appropriate amount of aspect-aware words, while excessively extending the hop numbers for searching aspect-aware words may bring noise. Thus we set the hop number as 5 in our model.

4.6 Impact of GCN Blocks

To analyze the impact of the layer number of GCN blocks over the performance of our proposed model, we conduct experiments by varying the layers from 1 to 6 and show the results in Figure 6. Note that 2-layer GCN blocks performs overall better, thus we set the layer number of GCN blocks as 2 in our experiments. Comparatively, 1-layer GCN block performs unsatisfactorily, which potentially indicates that 1-layer GCN block is insufficient to leverage precise aspect-related sentiment information from the context. In addition, the performance fluctuates with the increasing layer number of GCN blocks and evidently tends to decline when the layer number is greater than 4. This implies that roughly increasing the depth of GCN block is

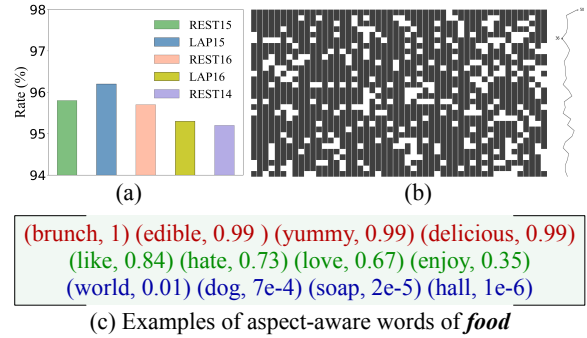


Figure 7: Covering rate of aspect-aware words (a). Examples of aspect-aware words distribution (b), white spaces denote aspect-aware words in the context.

vulnerable to slash the learning ability of the model due to the sharp increase of model parameters.

4.7 Analysis of Aspect-aware Words

To investigate the appearance of aspect-aware words in the sentence, we report the covering rates of aspect-aware words on different datasets in Figure 7 (a). Note that the coverage rate of aspect-aware words in all datasets exceeds 95%. That is, more than 95% sentences contain aspect-aware words. This validates the hypothesis that aspect-related words generally serve as sentiment descriptions of the corresponding aspect in the sentence, and verifies the convincingness and significance of our proposed method in ACSA task. Further, we randomly select 50 sentences from REST15 dataset and show the distribution of aspect-aware words in Figure 7 (b). Note that almost all the sentences contain an appropriate amount of aspect-aware words. This impliedly indicates that aspect-aware words are generally occur as key clues in the sentences. We show some typical aspect-aware words paired with their weights derived for aspect word “*food*” in Figure 7 (c). Note that 1) the words that highly associated with “*food*” are with great weights (the red examples), 2) the common sentiment words are with average weights (the green examples), 3) the irrelevant words are with small weights (the blue examples). This qualitatively verifies that our proposed method of deploying Beta Distribution to derive aspect-aware weight is effective in ACSA.

4.8 Case Study

To qualitatively demonstrate how contextual aspect-aware words work in ACSA task, we visualize the aspect-aware weights in Figure 8. Although the aspect (both E and A) of Example (a) is non-existent in the sentence, the sentiment clue of the aspect can

(a)	FOOD	0	0	0.98	0	0.57	0	0.86
	PRICES	0	0	0	0	0.01	0	0.99
		But	the	pizza	is	way	to	expensive
(b)	RESTAURANT	0	1	0	0.95	0	0.99	0.61
	PRICES	0	0.01	0	0.46	0	0.31	0.99
		The	place	is	pricey	the	food	is
(c)	FOOD	0	0.76	0	0.93	0	1	0.79
	QUALITY	0	0.01	0	0.86	0	0.15	0.99
		The	pizza	is	overpriced	and	soggy	

Figure 8: Examples of contextual aspect-aware words. The values in the box represent aspect-aware weights.

be easily learned with the help of the aspect-aware words. Example (b) and (c) are two instances containing multiple aspects, in which, the entity “food” occurs in the sentence of Example (b) while none of the aspect occurs in the sentence in Example (c). Note that the significant contextual words with respect to the distinct aspect can be extracted and distinguished for learning aspect-related sentiment expression with the help of aspect-aware words.

5 Conclusion

In this paper, we investigate the Aspect Category Sentiment Analysis (ACSA) task from a novel perspective that learning how to preferably find the aspect-aware words that are highly related to the aspects, and educe their weights with Beta Distribution based on the public knowledge. The aspect-aware words paired with their weights are deployed to construct aspect-aware graph(s) of the context for learning the contextual sentiment dependencies in ACSA with a graph convolutional structure. Experimental results on 6 benchmark datasets demonstrate the effectiveness of our proposed method.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (61632011, 61876053, 62006062, 62176076), the Guangdong Province Covid-19 Pandemic Control Research Funding (2020KZDZX1224), the Shenzhen Foundational Research Funding (JCYJ20180507183527919 and JCYJ20200109113441941), China Postdoctoral Science Foundation (2020M670912), Joint Lab of HITSZ and China Merchants Securities. Youth Innovation Promotion Association of CAS China (No. 2020357), and Shenzhen Science and Technology Innovation Program (Grant No. KQTD20190929172835662).

References

- Hongjie Cai, Yaofeng Tu, Xiangsheng Zhou, Jianfei Yu, and Rui Xia. 2020. Aspect-category based sentiment analysis with hierarchical graph convolutional network. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 833–843.
- 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 *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*, page 105–114.
- Xiao Chen, Changlong Sun, Jingjing Wang, Shoushan Li, Luo Si, Min Zhang, and Guodong Zhou. 2020. Aspect sentiment classification with document-level sentiment preference modeling. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3667–3677.
- Jiajun Cheng, Shenglin Zhao, Jiani Zhang, Irwin King, Xin Zhang, and Hui Wang. 2017. Aspect-level sentiment classification with heat (hierarchical attention) network. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 97–106.
- 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: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- Arjun K Gupta and Saralees Nadarajah. 2004. *Handbook of beta distribution and its applications*. CRC press.
- Binxuan Huang and Kathleen Carley. 2019. Syntax-aware aspect level sentiment classification with graph 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)*, pages 5469–5477.
- Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao, and Min Yang. 2019. A challenge dataset and effective models for aspect-based sentiment analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6280–6285.
- Yuncong Li, Cunxiang Yin, and Sheng-hua Zhong. 2020a. Sentence constituent-aware aspect-category sentiment analysis with graph attention networks. In *CCF International Conference on Natural Language Processing and Chinese Computing*, pages 815–827. Springer.

- Yuncong Li, Cunxiang Yin, Sheng-hua Zhong, and Xu Pan. 2020b. Multi-instance multi-label learning networks for aspect-category sentiment analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3550–3560.
- Bin Liang, Jiachen Du, Ruifeng Xu, Binyang Li, and Hejiao Huang. 2019a. Context-aware embedding for targeted aspect-based sentiment analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4678–4683.
- Bin Liang, Yonghao Fu, Lin Gui, Min Yang, Jiachen Du, Yulan He, and Ruifeng Xu. 2021a. Target-adaptive graph for cross-target stance detection. In *Proceedings of the Web Conference 2021*, pages 3453–3464.
- Bin Liang, Rongdi Yin, Jiachen Du, Lin Gui, Yulan He, Min Yang, and Ruifeng Xu. 2021b. [Embedding refinement framework for targeted aspect-based sentiment analysis](#). *IEEE Transactions on Affective Computing*.
- Bin Liang, Rongdi Yin, Lin Gui, Jiachen Du, Yulan He, and Ruifeng Xu. 2020a. Aspect-invariant sentiment features learning: Adversarial multi-task learning for aspect-based sentiment analysis. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*, page 825–834.
- Bin Liang, Rongdi Yin, Lin Gui, Jiachen Du, and Ruifeng Xu. 2020b. Jointly learning aspect-focused and inter-aspect relations with graph convolutional networks for aspect sentiment analysis. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 150–161.
- Yunlong Liang, Fandong Meng, Jinchao Zhang, Jinan Xu, Yufeng Chen, and Jie Zhou. 2019b. A novel aspect-guided deep transition model for aspect based sentiment analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5569–5580.
- Yukun Ma, Haiyun Peng, and Erik Cambria. 2018. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In *Proceedings of the Thirty-Second Conference on Artificial Intelligence (AAAI-18)*, pages 5876–5883.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphee De Clercq, Veronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Núria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495.
- 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 2014)*, pages 27–35.
- Libo Qin, Zhouyang Li, Wanxiang Che, Minheng Ni, and Ting Liu. 2021a. Co-gat: A co-interactive graph attention network for joint dialog act recognition and sentiment classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13709–13717.
- Libo Qin, Fuxuan Wei, Tianbao Xie, Xiao Xu, Wanxiang Che, and Ting Liu. 2021b. GL-GIN: Fast and accurate non-autoregressive model for joint multiple intent detection and slot filling. 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)*, pages 178–188.
- Libo Qin, Xiao Xu, Wanxiang Che, and Ting Liu. 2020. Towards fine-grained transfer: An adaptive graph-interactive framework for joint multiple intent detection and slot filling. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 1807–1816.
- Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge.
- Chi Sun, Luyao Huang, and Xipeng Qiu. 2019. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. 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)*, pages 380–385.
- Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2016. Effective LSTMs for target-dependent sentiment classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3298–3307.

- Hao Tian, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu. 2020. SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4067–4076.
- Meihan Tong, Bin Xu, Shuai Wang, Yixin Cao, Lei Hou, Juanzi Li, and Jun Xie. 2020. Improving event detection via open-domain trigger knowledge. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5887–5897.
- Yequan Wang, Minlie Huang, xiaoyan zhu, and Li Zhao. 2016. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615.
- Yequan Wang, Aixin Sun, Minlie Huang, and Xiaoyan Zhu. 2019. Aspect-level sentiment analysis using as-capsules. In *The World Wide Web Conference*, page 2033–2044.
- Bowen Xing, Lejian Liao, Dandan Song, Jingang Wang, Fuzheng Zhang, Zhongyuan Wang, and Heyan Huang. 2019. Earlier attention? aspect-aware lstm for aspect-based sentiment analysis. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 5313–5319. International Joint Conferences on Artificial Intelligence Organization.
- Wei Xue and Tao Li. 2018. Aspect based sentiment analysis with gated convolutional networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 2514–2523.
- Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7370–7377.
- Rongdi Yin, Hang Su, Bin Liang, Jiachen Du, and Ruifeng Xu. 2020. Extracting the collaboration of entity and attribute: Gated interactive networks for aspect sentiment analysis. In *Natural Language Processing and Chinese Computing*, pages 802–814.
- Bowen Zhang, Min Yang, Xutao Li, Yunming Ye, Xiaofei Xu, and Kuai Dai. 2020. Enhancing cross-target stance detection with transferable semantic-emotion knowledge. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3188–3197.
- 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)*, pages 4567–4577.
- Dong Zhang, Suzhong Wei, Shoushan Li, Hanqian Wu, Qiaoming Zhu, and Guodong Zhou. 2021. Multi-modal graph fusion for named entity recognition with targeted visual guidance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14347–14355.