Aspect Sentiment Classification with Document-level Sentiment Preference Modeling

Xiao Chen[1], **Changlong Sun**[2], **Jingjing Wang**[1,*], **Shoushan Li**[1], **Luo Si**[2], **Min Zhang**[1], **Guodong Zhou**[1]
[1]School of Computer Science and Technology, Soochow University, China
[2]Alibaba Group, China
xchen1206@stu.suda.edu.cn, changlong.scl@taobao.com
{djingwang, lishoushan, minzhang, gdzhou}@suda.edu.cn
luo.si@alibaba-inc.com

Abstract

In the literature, existing studies always consider Aspect Sentiment Classification (ASC) as an independent sentence-level classification problem aspect by aspect, which largely ignore the document-level sentiment preference information, though obviously such information is crucial for alleviating the information deficiency problem in ASC. In this paper, we explore two kinds of sentiment preference information inside a document, i.e., contextual sentiment consistency w.r.t. the same aspect (namely intra-aspect sentiment consistency) and contextual sentiment tendency w.r.t. all the related aspects (namely inter-aspect sentiment tendency). On the basis, we propose a Cooperative Graph Attention Networks (CoGAN) approach for cooperatively learning the aspect-related sentence representation. Specifically, two graph attention networks are leveraged to model above two kinds of document-level sentiment preference information respectively, followed by an interactive mechanism to integrate the two-fold preference. Detailed evaluation demonstrates the great advantage of the proposed approach to ASC over the state-of-the-art baselines. This justifies the importance of the document-level sentiment preference information to ASC and the effectiveness of our approach capturing such information.

1 Introduction

Aspect Sentiment Classification (ASC), a fine-grained sentiment classification task in the field of sentiment analysis (Pang and Lee, 2007; Li et al., 2010), aims to identify the sentiment polarity (e.g., *positive*, *negative* or *neutral*) for each aspect discussed inside a sentence. For example, the sentence "*The restaurant has quite low price but the food tastes not good*" would be assigned with a *positive* polarity for the aspect *price* and with a *negative* polarity for the aspect *food*. Over the past decade, the ASC task has been drawing more and more interests (Tang et al., 2016b; Wang et al., 2018) due to its wide applications, such as e-commerce customer service (Jing et al., 2015), public opinion mining (Wang et al., 2019c) and Question Answering (Wang et al., 2019a).

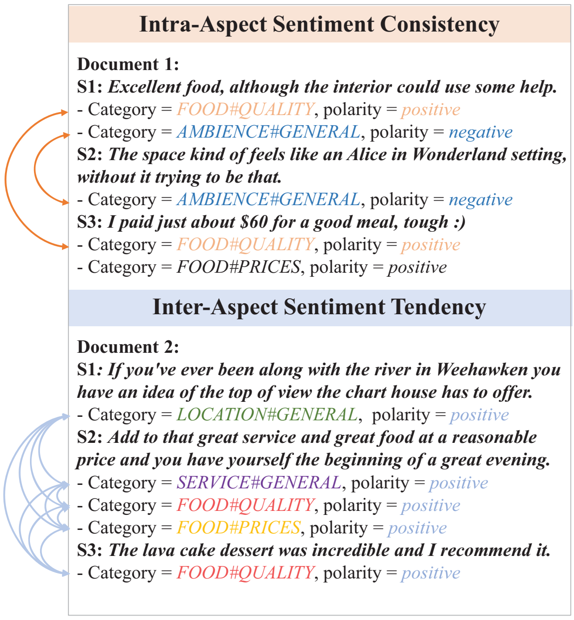

Figure 1: Two documents from SemEval 2016 (Pontiki et al. (2016)) datasets, where aspect category is defined as the entity *E* and attribute *A* pair (i.e., *E#A*). Red lines denote the intra-aspect sentiment consistency and blue lines denote the inter-aspect sentiment tendency.

In the literature, given the ASC datasets (Pontiki et al. (2016)) where aspects (i.e., entity and attribute) are manually annotated comprehensively sentence by sentence, previous studies model the aspect sentiment independently sentence by sentence, which suffer from the problem of ignoring the document-level sentiment preference information. In this study, we argue that such document-level sentiment preference information is crucial to

*Corresponding Author: Jingjing Wang.

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 3667–3677
July 5 - 10, 2020.

remedy the information deficiency issue in ASC. Especially, we explore two kinds of sentiment preference information inside a document.

On one hand, we assume that the sentences in a document involving the same aspect tend to have the same sentiment polarity on this aspect. For instance, in **Document 1**, both the sentence **S1** and **S2** involve aspect *AMBIENCE#GENERAL*. Although it is difficult to infer the *negative* sentiment for aspect *AMBIENCE#GENERAL* through the clause *"without it trying to be that"* in **S2**, we can infer that the sentiment of aspect *AMBIENCE#GENERAL* is more likely to be *negative* according to **S1**. This is because it is easier to infer *negative* for aspect *AMBIENCE#GENERAL* through the clause *"interior could use some help"* in **S1**. Therefore, a well-behaved approach should capture the contextual sentiment consistency w.r.t. the same aspect (namely intra-aspect consistency for short) information.

On the other hand, we assume that the sentences in a document tend to have the same sentiment polarity on all the related aspects. For the example of **Document 2** where the sentence **S2** involves multiple aspects, it is really hard to precisely predict the sentiment polarity for each aspect. However, when taken the context into consideration, the sentiment polarity for each aspect in **S2** is largely possible to be *positive*, since all the neighboring sentences express the *positive* sentiment polarity for their aspects. Therefore, a well-behaved model should capture the contextual sentiment tendency w.r.t. all the related aspects (namely inter-aspect tendency for short) information.

To well accommodate the above two kinds of document-level sentiment preference information, we propose a Cooperative Graph Attention Networks (CoGAN) approach to ASC. Specifically, two graph attention networks are constructed to model the two-fold sentiment preference with the attention weight to measure the preference-degree. Furthermore, considering that the two-fold preference can jointly influence the sentiment polarities for aspects, we propose an interactive mechanism to jointly model the two-fold preference for obtaining better aspect-related sentence representation. Detailed evaluation shows our proposed CoGAN approach significantly outperforms the state-of-the-art baselines, including the three top-performed systems from SemEval-2015 Task 12 and SemEval-2016 Task 5 (Pontiki et al., 2015, 2016).

2 Related Work

In this section, we first review the Aspect Sentiment Classification (ASC) task, and then introduce the related studies on graph-based neural networks.

Aspect Sentiment Classification. The ASC task aims to predict the sentiment polarity for each aspect discussed inside a sentence. Existing studies mainly focus on utilizing various approaches (e.g., attention mechanism and memory network) to align each aspect and the sentence for learning aspect-related sentence representation. Wang et al. (2016) propose an attention-based LSTM in order to explore the potential correlation of aspects and sentiment polarities in ASC. Wang et al. (2018) propose a hierarchical attention network to incorporate both words and clauses information for ASC. He et al. (2018a) propose an attention-based approach to incorporate the aspect-related syntactic information for ASC. Tang et al. (2016b) and Chen et al. (2017) design deep memory networks to align the aspect and sentence for ASC. Lin et al. (2019) propose a semantic and context-aware memory network to integrate aspect-related semantic parsing information for performing ASC. Wang et al. (2019a) and Wang et al. (2019b) leverage reinforcement learning grounded approaches to select aspect-relevant words for ASC. Recently, a few studies have recognized the information deficiency problem in ASC and attempted to using external information to improve the performance of ASC. He et al. (2018b) and Chen and Qian (2019) incorporate the knowledge from document-level sentiment classification to improve the performance of ASC. Ma et al. (2018) propose an extension of LSTM to integrate the commonsense knowledge into the recurrent encoder for improving the performance of ASC. In addition, it is worthwhile to note that Hazarika et al. (2018) also investigate the inter-aspect sentiment dependency for ASC, but is limited to capture this information inside a single sentence.

In summary, all the above studies ignore the document-level sentiment preference information, which can be leveraged to effectively mitigate the information deficiency problem in ASC.

Graph-based Neural Networks. In recent years, graph-based neural networks have received more and more attentions. As a pioneer, Kipf and Welling (2017) present a simplified graph neural network model, called graph convolutional networks (GCN), which has been exported to several tasks such as scene recognition (Yuan et al., 2019), semi-supervised node classification (Zhang et al., 2019b), text-to-SQL parsing (Bogin et al., 2019) and relation extraction (Sahu et al., 2019). On this basis, some other improved Graph-based Neural Networks are proposed. Morris et al. (2019) propose a generalization of Graph-based Neural Networks, so-called k-dimensional GNNs (k-GNNs), which can take higher-order graph structures at multiple scales into account. Cao et al. (2019) propose a novel Multi-channel Graph Neural Network model to learn alignment-oriented knowledge graph embeddings by robustly encoding two knowledge graphs via multiple channels. More recently, there exist several studies also adopting graph-based neural networks to ASC. For instance, Hou et al. (2019) and Zhang et al. (2019a) build GCN over the dependency tree of a sentence to exploit syntactical information and word dependencies for learning better aspect-related sentence representation for ASC.

Different from all the above studies, this paper proposes a novel Cooperative Graph Attention Networks approach to capture the document-level sentiment preference information in ASC. To our best knowledge, this is the first attempt to incorporate this information for the ASC task.

3 Cooperative Graph Attention Networks (CoGAN)

In this section, we formulate the Aspect Sentiment Classification (ASC) task as follows. In each document D with sentences[1] $\{s_1, s_2, ..., s_I\}$, given a sentence $s_i, i \in \{1, 2, ..., I\}$ and its aspect $a_k, k \in \{1, 2, ..., K\}$, the ASC task aims to predict the sentiment polarity ℓ for aspect a_k by automatically learning the aspect-related sentence representation r_i of sentence s_i. Here, I is the number of sentence s_i , and K is the number of aspect a_k inside the document.

In this paper, we propose a Cooperative Graph Attention Networks (CoGAN) approach with two types of Graph Attention Networks (GAN) to incorporate the two-fold preference information respectively. Figure 2 shows the overall architecture of the CoGAN approach which consists of five major blocks: **1)** Encoding Block; **2)** Intra-Aspect Consistency Modeling Block; **3)** Inter-Aspect Tendency Modeling Block; **4)** Interaction Block. **5)** Softmax Decoding Block. Before introducing our CoGAN approach, we first give an overview of the basic Graph Attention Network (GAN).

[1]Like Pontiki et al. (2015), all aspects of every sentence are unrolled in a document. For instance, a sentence with two aspects occurs twice in succession, once with each aspect.

3.1 Basic Graph Attention Network

Graph Attention Network (GAN) (Velickovic et al., 2017) is a new graph neural network architecture including attention mechanism, which enables specifying different attention weights to different vertices in a neighborhood. In principle, GAN can aggregate the features of neighboring nodes and also can propagate the information of a vertex to its nearest neighbors. From this regard, GAN is capable of sufficiently modeling local contextual information for learning the representation of each vertex. Formally, given a graph $G(V, E)$ where V and E denote the vertices and edges respectively, GAN updates each new vertex vector $\hat{h}_i$ of vertex v_i by considering neighboring vertices' vectors $\{h_j\}_{j=1}^{I}$ with the following formulas:

$$
\begin{aligned}
\hat{h}_i &= \tanh(\sum_{j=1}^{I} \alpha_{ij} \boldsymbol{W} h_j + b) \\
\alpha_{ij} &= \frac{\exp(f(w^\top [\boldsymbol{W} h_i; \boldsymbol{W} h_j]))}{\sum_{t=1}^{I} \exp(f(w^\top [\boldsymbol{W} h_i; \boldsymbol{W} h_t]))}
\end{aligned} \tag{1}
$$

where α_{ij} is the attention weight (i.e., the edge weight) between vertex v_i and vertex v_j. $f(\cdot)$ is a LeakyReLU activation function. $[;]$ denotes vector concatenation. $\boldsymbol{W} \in \mathbb{R}^{d\times d}$ and $w \in \mathbb{R}^{2d}$ are the trainable parameters.

In the following, we will illustrate the five main components of our CoGAN approach respectively.

3.2 Encoding Block

As a text encoding mechanism, BERT (Devlin et al., 2019) can be fine-tuned to create state-of-the-art models for a range of NLP tasks, e.g., text classification and natural language inference. In our approach, we use BERT-base[2] (uncased) model to encode both the aspect and the sentence as follows.

• **Aspect Encoding.** Since an aspect a_k consists of an entity e_{entity} and an attribute $e_{attribute}$ (Pontiki et al., 2015), we process the entity-attribute pair (e_{entity}, $e_{attribute}$) into the input pair format of BERT as:

$$[\text{CLS}] \;\; e_{entity} \;\; [\text{SEP}] \;\; e_{attribute} \;\; [\text{SEP}]$$

Then, we feed the entity-attribute pair into BERT and regard the mark "[CLS]" representation as the aspect vector $e_k \in \mathbb{R}^d$ of the aspect a_k.

[2]https://github.com/google-research/bert

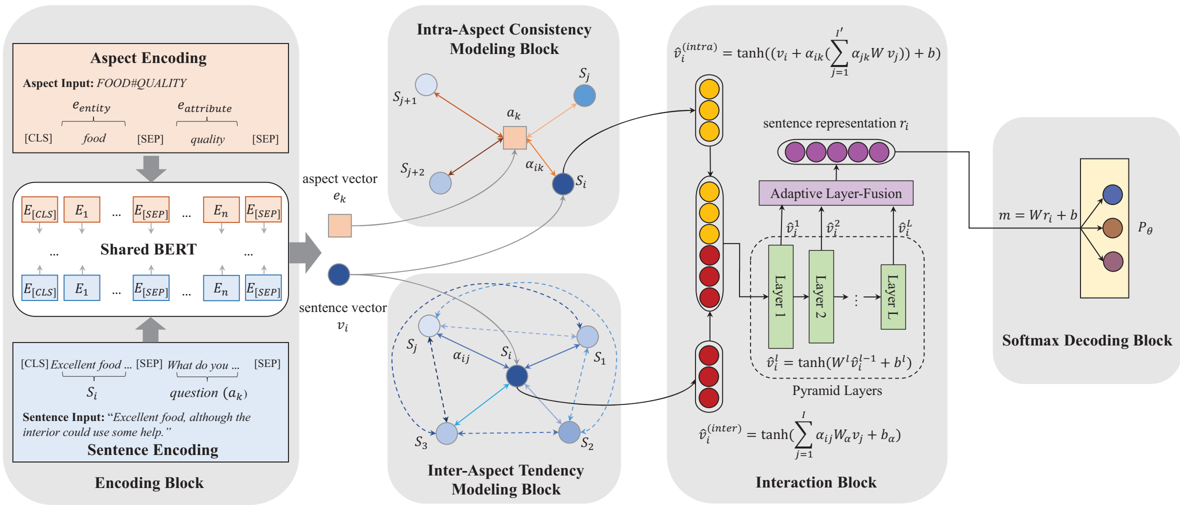

Figure 2: The overall framework of our proposed Cooperative Graph Attention Networks (CoGAN).

• **Sentence Encoding.** We borrow the approach proposed by Sun et al. (2019) to generate the aspect-related sentence representation, which has achieved promising performance for the ASC task. Following Sun et al. (2019), we first process the sentence s_i and its corresponding aspect a_k into the input pair format of BERT as:

$$[\text{CLS}]\ \ s_i\ \ [\text{SEP}]\ \ question(a_k)\ \ [\text{SEP}]$$

where $question(\cdot)$ denotes the construction of auxiliary question sentence for aspect a_k proposed by Sun et al. (2019). For example, the auxiliary sentence for aspect *FOOD#PRICE* is constructed as "*what do you think of the food and price?*". Then, we similarly feed the above pair into BERT (shared with aspect encoding) and obtain the aspect-related sentence vector $v_i \in \mathbb{R}^d$ of the sentence s_i. Further, we fine-tune BERT and update both the aspect vector e_k and sentence vector v_i according to Eq.(8).

3.3 Intra-Aspect Consistency Modeling Block

In our approach, we propose a consistency-aware GAN to model the intra-aspect consistency. Given a document $\mathcal{D}$ with sentences $\{s_1, s_2, ..., s_I\}$, the consistency-aware GAN is denoted as a bipartite graph $G(S \bigcup A, E_{sa})$. Here, S and A are two disjoint sets of vertices, denoting the sentence vertices and the aspect vertices respectively. E_{sa} is the set of the edge between the sentence $s_i \in S$ and its corresponding aspect $a_k \in A$ in the document $\mathcal{D}$.

On the basis, the intra-aspect consistency is formulated as that sentence vertices $\{s_i\}_{i=1}^{I'}$ sharing the same neighboring aspect vertex a_k and located in the same document tend to have the same sentiment for this aspect a_k. Here, I' denotes the number of sentences sharing the same aspect a_k. Nevertheless, there still possibly exist some sentiment inconsistency cases.

Considering all the scenarios above, we use the graph attention mechanism (Velickovic et al., 2017) to measure the preference-degree, where the attention weight (preference-degree) is computed as the edge weight between the sentence vertex s_i and the aspect vertex a_k in a document. Specifically, according to Eq.(1), the attention weight α_{ik} between sentence s_i and aspect a_k is computed as follows:

$$\alpha_{ik} = \frac{\exp(f(w^\top [\boldsymbol{W_v} v_i; \boldsymbol{W_e} e_k]))}{\sum_{t=1}^{I'} \exp(f(w^\top [\boldsymbol{W_v} v_t; \boldsymbol{W_e} e_k]))} \quad (2)$$

where $\boldsymbol{W_v}, \boldsymbol{W_e} \in \mathbb{R}^{d \times d}$ and $w \in \mathbb{R}^{2d}$ are the trainable parameters.

As a vertex in $G(S \bigcup A, E_{sa})$, the sentence s_i is then encoded as the aspect-related sentence representation $\hat{v}_i^{(intra)}$ according to the following proposed formula by modifying Eq.(1).

$$\hat{v}_i^{(intra)} = \tanh((v_i + \alpha_{ik}(\sum_{j=1}^{I'} \alpha_{jk} \boldsymbol{W} v_j)) + b) \quad (3)$$

where $\sum_{j=1}^{I'} \alpha_{jk} \boldsymbol{W} v_j$ is the vector representation of the aspect vertex a_k, which is weighted added to the sentence vector v_i for enhancing the aspect-related sentence representation. $\boldsymbol{W} \in \mathbb{R}^{d \times d}$, $b \in \mathbb{R}^d$ are trainable parameters.

3.4 Inter-Aspect Tendency Modeling Block

In our approach, we leverage a tendency-aware GAN to model the inter-aspect tendency. Given a document $\mathcal{D}$ with sentences $\{s_1, s_2, ...s_I\}$, the tendency-aware GAN is denoted as an undirected graph $G(S, E_{ss})$. Here, S is the set of sentence vertices. E_{ss} is the set of the edge between sentence $s_i \in S$ and sentence $s_j \in S$ in the document $\mathcal{D}$.

On the basis, the inter-aspect tendency assumption is formulated as that the sentence vertex s_i tend to have the same sentiment with the neighboring sentence vertices $\{s_j\}_{j=1}^{I}$ inside a same document.

Similar to intra-aspect consistency modeling block, according to Eq.(1), the following formula is applied to compute the attention weight α_{ij} between the sentence vertex s_i and the sentence vertex s_j from the same document:

$$\alpha_{ij} = \frac{\exp(f(w^{\top}[\boldsymbol{W_1}v_i; \boldsymbol{W_2}v_j]))}{\sum_{t=1}^{I}\exp(f(w^{\top}[\boldsymbol{W_1}v_i; \boldsymbol{W_2}v_t]))} \quad (4)$$

where $\boldsymbol{W_1}, \boldsymbol{W_2} \in \mathbb{R}^{d\times d}$ and $w \in \mathbb{R}^{2d}$ are the trainable parameters.

As a vertex in $G(S, E_{ss})$, according to Eq.(1), sentence s_i is encoded as the new sentence representation $\hat{v}_i^{(inter)}$, i.e.,

$$\hat{v}_i^{(inter)} = \tanh(\sum_{j=1}^{I}\alpha_{ij}\boldsymbol{W_\alpha}v_j + b_\alpha) \quad (5)$$

where $\boldsymbol{W_\alpha} \in \mathbb{R}^{d\times d}$, $b_\alpha \in \mathbb{R}^{d}$ are the parameters.

3.5 Interaction Block

Since the above two-fold preference can jointly affect the sentiment for aspect a_k in s_i, we make the two-fold preference pairwisely interact with each other for cooperatively boosting the performance. Especially, after obtaining the two sentence representations $\hat{v}_i^{(inter)}$ and $\hat{v}_i^{(intra)}$ of sentence s_i from the above two-fold preference modeling blocks, we propose an interactive mechanism to make an interaction between the two vectors instead of simply concatenating them. This is because a simple vector concatenation does not account for any interactions between the latent features of the two-fold preference, which is insufficient for cooperatively modeling the two-fold preference. In detail, this interactive mechanism leverages two strategies to learn the sentence representation.

• **Pyramid Layers.** As proposed in He et al. (2016), the model using a small number of hidden units for higher layers can learn more abstractive features. Inspired by this, we add pyramid hidden layers (see Figure 2) on the concatenated vector for interacting the latent features of the two-fold preference, where the bottom layer is the widest and each successive layer has a smaller number of neurons. More specifically, the sentence vector $\hat{v}_i^l \in \mathbb{R}^{2d\cdot(\frac{1}{2})^{l-1}}$ of the l-th layer is defined as:

$$\hat{v}_i^l = \tanh(\boldsymbol{W^l}\hat{v}_i^{l-1} + b^l) \quad (6)$$

where $\hat{v}_i^1 = [\hat{v}_i^{(inter)}; \hat{v}_i^{(intra)}]$ and adding one layer will make the dimension of the sentence vector half. $\boldsymbol{W^l} \in \mathbb{R}^{2d\cdot(\frac{1}{2})^{l-1}\times 2d\cdot(\frac{1}{2})^{l-2}}$ and b^l are the trainable parameters. $l \in [1, L]$ denotes the layer index.

• **Adaptive Layer-Fusion.** To sufficiently fuse the sentence representations at different level of abstractions, an adaptive fusion mechanism is proposed to fuse the representations of all layers for computing the final sentence vector $r_i \in \mathbb{R}^d$ of vertex s_i as follows:

$$r_i = \tanh(\boldsymbol{W_r}(\prod_{l=1}^{L}\alpha_i\hat{v}_i^l) + b_r) \quad (7)$$

where $\prod$ denotes the concatenation of multiple vectors. W_r and $b_r \in \mathbb{R}^d$ are the trainable parameters. L is the number of added layers and set to be 4 fine-tuned according to the development data. $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_L]$ is a normalized weights vector to weigh each layer, which is learned during training.

3.6 Softmax Decoding Block

After obtaining the final sentence vector r_i of sentence s_i, we feed it to a softmax classifier $m = \boldsymbol{W}r_i + b$, where $m \in \mathbb{R}^C$ is output vector; $\boldsymbol{W}$ and b are the trainable parameters.

Then, the probability of labeling sentence with sentiment polarity $\ell \in [1, C]$ is computed by $p_\theta(\ell|r_i) = \frac{\exp(m_\ell)}{\sum_{\eta=1}^{C}\exp(m_\eta)}$. Finally, the label with the highest probability stands for the predicted sentiment polarity for the aspect a_k.

3.7 Model Training

We use cross-entropy loss function to train our model end-to-end given a set of training data (s_i, a_k, y_i) from corpus $\mathcal{C}$, where s_i is the i-th sentence to be predicted, a_k is its corresponding aspect and y_i is the ground-truth sentiment polarity for aspect a_k. The objective of learning θ is to minimize the loss function as follows:

$$J(\theta) = \mathbb{E}_{(s_i,a_k,y_i)\sim\mathcal{C}}[-\log p_\theta(y_i|r_i)] + \frac{\delta}{2}||\theta||_2^2 \quad (8)$$

where $\mathbb{E}$ denotes the expectation-maximization. $\sim$ denotes the sampling operation. θ denotes all the trainable parameters of our CoGAN approach. δ is a L_2 regularization.

4 Experimentation

4.1 Experimental Settings

Data Settings. We conduct experiments on four datasets[3], i.e., two datasets (*restaurant15* and *laptop15*) released by SemEval-2015 Task 12 (Pontiki et al., 2015) and the other two datasets (*restaurant16* and *laptop16*) released by SemEval-2016 Task 5 (Pontiki et al., 2016), to verify the effectiveness of our proposed approach. Wherein, each dataset averagely consists of about 442 documents and one document averagely contains 4.9 sentences. Moreover, each sentence is annotated with one or multiple aspects and a sentiment polarity (i.e., *positive*, *negative* or *neutral*) for each aspect. Additionally, we set aside 10% from the training set as the development data to tune the hyper-parameters.

Implementation Details. In our experiments, all hyper-parameters are tuned according to the development set. Specifically, BERT is optimized by the Adam optimizer (Kingma and Ba, 2015), where $\beta_1 = 0.9$ and the initial learning rate is 10^{-4}. Other parameters of BERT are following (Devlin et al., 2019). For our CoGAN approach, we adopt another Adam optimizer with an initial learning rate of 10^{-3} and $\beta_1 = 0.95$ for cross-entropy training. The regularization weight of parameters is 10^{-5}. The dropout rate is 0.25. All matrix and vector parameters of the layers are initialized by the Glorot uniform (Glorot and Bengio, 2010).

Evaluation Metrics. The performance is evaluated using standard *Accuracy* (Acc.) and *Macro-F1* (F1) (Wang et al., 2017). Moreover, t-test is used to evaluate the significance of the performance difference (Yang and Liu, 1999).

Baselines. We give the following baseline approaches for comparison in order to comprehensively evaluate the performance of our approach. **1) TC-LSTM.** This approach extends LSTM by considering the aspect information where a forward LSTM and a backward one towards the aspect are adopted (Tang et al., 2016). **2) ATAE-LSTM.** This approach models the aspect-related context words via attention-based LSTM (Wang et al., 2016). **3) RAM.** This approach captures importance of context words for a specific aspect with a deep memory network and the results of multiple attentions are non-linearly combined with a recurrent neural network (Chen et al., 2017). **4) IAN.** This approach is an interactive learning approach, which models the contexts and aspects via LSTM and then interactively learns attentions in the contexts and aspects (Ma et al., 2017). **5) Clause-Level ATT.** This approach employs hierarchical attention to incorporate the clause information for ASC (Wang et al., 2018). **6) LSTM+synATT+TarRep.** This approach employs syntax-aware attention to learn aspect-related representation for ASC. This is a state-of-the-art approach proposed by He et al. (2018a). **7) BERT.** This approach transforms ASC from a single sentence classification task to a sentence pair classification task. In our implementation, we regard the pair of sentence and its aspect as the input pair of BERT-base model (Devlin et al., 2018) for performing ASC. **8) CADMN.** This approach employs attention model to attend on relevant aspects for enhancing the aspect representation. This is a state-of-the-art approach proposed by Song et al. (2019). **9) IMN.** This approach is a multi-task learning approach, which employs a novel message passing mechanism to better exploit the correlation among the tasks related to ASC. This is a state-of-the-art approach proposed by He et al. (2019). **10) BERT-QA.** This approach is an extension of the above BERT baseline proposed by Sun et al. (2019). In this study, we adopt BERT-pair-QA-M in our implementation. This is another state-of-the-art approach for ASC. **11) Sentiue.** This is the best-performed system in SemEval-2015 Task 12 (Saias, 2015), which achieves the best accuracy scores in both the *laptop15* and *restaurant15* domains. **12) XRCE.** This is the best-performed system in SemEval-2016 Task 5 (Pontiki et al., 2016), which achieves the best accuracy score in the *restaurant16* domain. **13) IIT-TUDA.** This is also the best-performed system in SemEval-2016 Task 5 (Pontiki et al., 2016), while achieving the best accuracy score in the *laptop16* domain. **15) CoGAN w/o Intra-Aspect Consistency.** Our approach only modeling Inter-Aspect Tendency. **16) CoGAN w/o Inter-Aspect Tendency.** Our approach only modeling Intra-Aspect Consistency. **17) CoGAN w/o Interactive Mechanism.** Our approach only concatenating the two vectors $\hat{v}_i^{(inter)}$ and $\hat{v}_i^{(intra)}$ instead of using interaction block to integrate them.

4.2 Experimental Results

Table 1 shows the performance comparison of different approaches. From the table, we can see that, all state-of-the-art approaches, such as **Clause-**

[3]Detail statistics can be seen in Pontiki et al. (2015, 2016).

Approaches	*Restaurant15* Acc.	F1	*Laptop15* Acc.	F1	*Restaurant16* Acc.	F1	*Laptop16* Acc.	F1
TC-LSTM (Tang et al., 2016)	0.747†	0.634†	0.745†	0.622†	0.813	0.629	0.766	0.578
ATAE-LSTM (Wang et al., 2016)	0.752†	0.641†	0.747†	0.637†	0.821	0.644	0.781	0.591
RAM (Chen et al., 2017)	0.767†	0.645†	0.759†	0.639†	0.839	0.661	0.802	0.627
IAN (Ma et al., 2017)	0.755†	0.639†	0.753†	0.625†	0.836	0.652	0.794	0.622
Clause-Level ATT (Wang et al., 2018)	0.809†	0.685†	0.816†	0.667†	0.841	0.667	0.809	0.634
LSTM+synATT+TarRep (He et al., 2018a)	0.817‡	0.661‡	0.822	0.649	0.846‡	0.675‡	0.813	0.628
BERT (Devlin et al., 2018)	0.811	0.647	0.809	0.683	0.884	0.729	0.811	0.670
CADMN (Song et al., 2019)	-	-	-	-	0.879♯	0.700♯	-	-
IMN (He et al., 2019)	0.856♮	0.718♮	0.831	0.654	0.892	0.710	0.802	0.623
BERT-QA (Sun et al., 2019)	0.824	0.650	0.827	0.595	0.896	0.715	0.812	0.596
Sentiue (Saias, 2015)	0.787†	0.660†	0.793†	0.634†	-	-	-	-
XRCE (Brun et al., 2016)	-	-	-	-	0.881*	-	-	-
IIT-TUDA (Kumar et al., 2016)	-	-	-	-	-	-	0.828§	-
CoGAN w/o Intra-Aspect Consistency	0.857	0.707	0.846	0.722	0.907	0.769	0.839	0.706
CoGAN w/o Inter-Aspect Tendency	0.854	0.716	0.841	0.708	0.915	0.770	0.811	0.676
CoGAN w/o Interactive Mechanism	0.864	0.704	0.839	0.698	0.908	0.788	0.839	0.700
CoGAN	**0.872**	**0.732**	**0.851**	**0.745**	**0.920**	**0.816**	**0.842**	**0.707**

Table 1: Comparison of all the approaches. The results with symbol † are retrieved from Wang et al. (2018); those with ‡ are from He et al. (2018a); those with ♯ are from Song et al. (2019); those with ♮ are from He et al. (2019); those with ∗ are from Brun et al. (2016) and those with § are from Kumar et al. (2016). The symbol - denotes both the results and codes are not reported by these papers.

Level ATT, **CADMN** and **IMN**, perform better than **TC-LSTM**. This result demonstrates the effectiveness of using a proper attention mechanism to learn the aspect-related sentence representation for performing the ASC task.

The BERT-based approaches, i.e., **BERT** and **BERT-QA**, perform better than the above approaches on almost all datasets. This result encourages to utilize the pre-trained BERT model as the aspect and sentence encoder for the ASC task.

Furthermore, our approach **CoGAN w/o Intra-Aspect Consistency** and **CoGAN w/o Inter-Aspect Tendency** outperform most of the above state-of-the-art approaches. This encourages to model the intra-aspect consistency or inter-aspect tendency information for the ASC task.

In comparison, when incorporating both the two-fold sentiment preference information, our approach **CoGAN** outperforms all the above baseline approaches and even significantly outperforms (p-value < 0.05) all three top-performed systems from SemEval-2015 Task 12 and SemEval-2016 Task 5, i.e., **Sentiue**, **XRCE** and **IIT-TUDA** on all four datasets. Impressively, compared to **TC-LSTM**, our approach achieves the average improvement of 11.6% (*Accuracy*), 14.3%(*Macro-F1*) on the two *restaurant* datasets and 9.1% (*Accuracy*), 12.6%(*Macro-F1*) on the two *laptop* datasets. Significance test shows that these improvements are all significant (p-value < 0.01). These results highlight the importance of incorporating both the intra-aspect consistency and inter-aspect tendency information in a document for the ASC task.

In addition, it is worthwhile to note that **CoGAN** outperforms **CoGAN w/o Interactive Mechanism**, which encourages to employ the proposed interactive mechanism to cooperatively integrate the two-fold sentiment preference information.

5 Analysis and Discussion

5.1 Case Study

We provide a qualitative analysis of our **CoGAN** approach on the test sets of the *restaurant16* and *laptop16* datasets respectively. Figure 3 shows two documents, along with their predicted sentiment for each aspect, and probabilities of the ground-truth label by different approaches. From this figure, we can see that: **1)** For the example of Document 1, it is difficult to infer the sentiment for aspect *LAPTOP#MISCELLANEOUS* (to classify) in **S1** since the long sentence **S1** involves syntactic complications. Despite this, **S8** expresses explicit *negative* polarity for the same aspect *LAPTOP#MISCELLANEOUS*. Considering this intra-aspect consistency information, our **CoGAN** approach can still give the correct *negative* for aspect *LAPTOP#MISCELLANEOUS*, while both **BERT**

Document 1 (Intra-Aspect Consistency)	Document 2 (Inter-Aspect Tendency)
S1: ***I would've given 5 stars, had it not been for the hours of updates I've had to do to this upon arrival.***	**S1:** ***hidden little jem.***
- Category = *LAPTOP#GENERAL* — ***positive***	- Category = *RESTAURANT#GENERAL* — ***To classify***
- Category = *LAPTOP#MISCELLANEOUS* — ***To classify***	**S2:** ***Never too crowded and always great service.***
S2: ***It's portable, reliable, and great for what I use it for.***	- Category = *SERVICE#GENERAL* — ***positive***
- Category = *LAPTOP#GENERAL* — ***positive***	- Category = *RESTAURANT#MISCELLANEOUS* — ***positive***
……	**S3:** ***I think I have probably tried each item on their menu at least once it is all excellent.***
S8: ***Negatives: As aforementioned, my only con was the updates.***	- Category = *FOOD#QUALITY* — ***positive***
- Category = *LAPTOP#MISCELLANEOUS* — ***negative***	……

BERT	IMN	CoGAN(Our Approach)	BERT	IMN	CoGAN(Our Approach)
(*positive*) ✗	(*positive*) ✗	(*negative*) ✓	(*negative*) ✗	(*neutral*) ✗	(*positive*) ✓
P(*negative*)=0.21	P(*negative*)=0.26	P(*negative*)=0.82	P(*positive*)=0.25	P(*positive*)=0.28	P(*positive*)=0.87

Figure 3: Examples from the test data with their polarities predicted by different approaches (i.e., BERT, IMN and our approach). ✓ (or ✗) denotes that the predicted sentiment polarity is correct (or wrong).

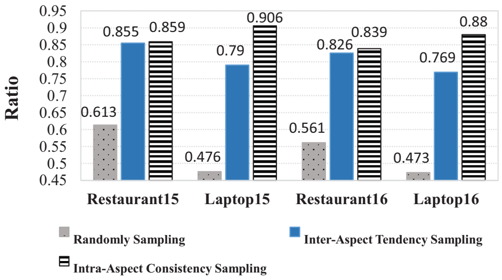

Figure 4: Ratios that two sentences have an identical sentiment polarity for their corresponding aspects.

and **IMN** give wrong predictions. This justifies the effectiveness of the intra-aspect consistency information for ASC. **2)** For the example of Document 2, it is rather difficult to infer the sentiment for aspect *RESTAURANT#GENERAL* (to classify) in **S1**, since the sentence **S1** "*hidden little jem.*" is too short and can not provide sufficient information to predict the *positive* polarity for aspect *RESTAURANT#GENERAL*. Despite this, **CoGAN** considering the inter-aspect tendency information can still give the *positive* for aspect *RESTAURANT#GENERAL*. This is reasonable because take the whole context into consideration, this restaurant has good reputations due to its service and food.

5.2 Effectiveness Study

To better illustrate the effectiveness of modeling the intra-aspect consistency and inter-aspect tendency information, we systematically investigate both sentiment preference phenomena in all the four evaluation datasets respectively. Specifically, we sample 200 sentence[4] pairs inside each dataset and calculate the ratio that the two sentences in the pair have the same sentiment for their corresponding aspects. Especially, we propose three sampling strategies as follows. **1)** Randomly Sampling: randomly selecting sentence pairs inside each dataset. **2)** Inter-Aspect Tendency Sampling: randomly selecting the sentence pairs under the premise that each two sentences should be located in the same document. **3)** Intra-Aspect Consistency Sampling: randomly selecting the sentence pairs under the premise that each two sentences should be located in the same document and should have the same aspect. Figure 4 shows the statistical results of the three sampling strategies on all four datasets. From this figure, we can see that **Inter-Aspect Tendency Sampling** and **Intra-Aspect Consistency Sampling** impressively outperform **Randomly Sampling** by 27.9 % and 34% respectively. Moreover, the average ratio of the two highest sampling strategies is up to 84.1%. This is the reason for the effectiveness of our CoGAN approach to ASC, and encourages to leverage CoGAN for incorporating the two-fold sentiment preference information.

[4]Sentences are repeated in a document according to the unrolled aspects. For instance, a sentence with two aspects will be repeated twice, each sentence with only one aspect.

5.3 Error Analysis

We randomly analyzed 100 error cases and roughly categorized them into 5 classes briefly introduced as follows. (1) 29% of errors are due to the occurrence of negation words, e.g., "*Nothing really came across as outstanding.*". CoGAN incorrectly predicts positive polarity, inspiring us to optimize CoGAN for capturing negation scope better. (2) 27% are due to incorrectly recognizing neutral instances. The shortage of neutral training examples makes it hard to recognize neutral instances, inspiring us to use data augmentation to enlarge the scale of neutral data. (3) 24% are due to the implicit sentiment expression, e.g., "*There is definitely more to* *say...*". CoGAN incorrectly predicts positive polarity instead of negative. (4) 12% are due to too short sentences (e.g. with less than 5 words), inspiring us to incorporate external ConceptNet knowledgebase to enhance the semantic representation. (5) 8% are due to comparative opinions, e.g., "*I've had better frozen pizza*". CoGAN incorrectly predicts positive, inspiring us to investigate whether incorporating syntactic information can remedy this issue.

6 Conclusion

In this paper, we propose a novel Cooperative Graph Attention Networks (CoGAN) approach to Aspect Sentiment Classification (ASC). The main idea of the proposed approach is to incorporate two kinds of sentiment preference information (i.e., the intra-aspect consistency and inter-aspect tendency) in a document for remedying the information deficiency problem in ASC. Experimental results on four datasets from SemEval-2015 and 2016 demonstrate that our approach significantly outperforms a number of competitive baselines, including all the three best-performed systems in the shared tasks of both SemEval-2015 and 2016.

In our future work, we would like to improve the performance of the ASC task by using unlabeled data since our graph-based neural network approach is easy to add unlabeled data. Moreover, we would like to apply our approach to other sentiment analysis tasks, e.g., aspect-oriented opinion summarization and multi-label emotion detection.

Acknowledgments

We thank our anonymous reviewers for their helpful comments. This work was supported by three NSFC grants, i.e., No.61751206, No.61672366 and No.61525205. This work was also supported by a Project funded by China Postdoctoral Science Foundation No.2019M661930, a Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), and a joint research project of Alibaba Group and Soochow University.

References

Ben Bogin, Jonathan Berant, and Matt Gardner. 2019. Representing schema structure with graph neural networks for text-to-sql parsing. In *Proceedings of ACL-2019*, pages 4560–4565.

Caroline Brun, Julien Perez, and Claude Roux. 2016. XRCE at semeval-2016 task 5: Feedbacked ensemble modeling on syntactico-semantic knowledge for aspect based sentiment analysis. In *Proceedings of SemEval@NAACL-HLT-2016*, pages 277–281.

Yixin Cao, Zhiyuan Liu, Chengjiang Li, Zhiyuan Liu, Juanzi Li, and Tat-Seng Chua. 2019. Multi-channel graph neural network for entity alignment. In *Proceedings of ACL-2019*, pages 1452–1461.

Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of EMNLP-2017*, pages 452–461.

Zhuang Chen and Tieyun Qian. 2019. Transfer capsule network for aspect level sentiment classification. In *Proceedings of ACL-2019*, pages 547–556.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT-2019*, pages 4171–4186.

Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of AISTATS-2010*, pages 249–256.

Devamanyu Hazarika, Soujanya Poria, Prateek Vij, Gangeshwar Krishnamurthy, Erik Cambria, and Roger Zimmermann. 2018. Modeling inter-aspect dependencies for aspect-based sentiment analysis. In *Proceedings of NAACL-2018*, pages 266–270.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of CVPR-2016*, pages 770–778.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018a. Effective attention modeling for aspect-level sentiment classification. In *Proceedings of COLING-2018*, pages 1121–1131.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018b. Exploiting document knowledge for aspect-level sentiment classification. In *Proceedings of ACL-2018*, pages 579–585.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In *Proceedings of ACL-2019 Volume 1: Long Papers*, pages 504–515.

Xiaochen Hou, Jing Huang, Guangtao Wang, Kevin Huang, Xiaodong He, and Bowen Zhou. 2019. Selective attention based graph convolutional networks for aspect-level sentiment classification. *CoRR*, abs/1910.10857.

Ranzhe Jing, Yang Yu, and Zhangxi Lin. 2015. How service-related factors affect the survival of B2T providers: A sentiment analysis approach. *J. Org. Computing and E. Commerce*, 25(3):316–336.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of ICLR-2015*.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *Proceedings of ICLR-2017*.

Ayush Kumar, Sarah Kohail, Amit Kumar, Asif Ekbal, and Chris Biemann. 2016. IIT-TUDA at semeval-2016 task 5: Beyond sentiment lexicon: Combining domain dependency and distributional semantics features for aspect based sentiment analysis. In *Proceedings of SemEval@NAACL-HLT-2016*, pages 1129–1135.

S. Li, C. Huang, G. Zhou, and Sophia Yat Mei Lee. 2010. Employing personal/impersonal views in supervised and semi-supervised sentiment classification. In *Proceedings of ACL*.

Peiqin Lin, Meng Yang, and Jianhuang Lai. 2019. Deep mask memory network with semantic dependency and context moment for aspect level sentiment classification. In *Proceedings of IJCAI-2019*, pages 5088–5094.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. In *Proceedings of IJCAI-2017*, pages 4068–4074.

Yukun Ma, Haiyun Peng, and Erik Cambria. 2018. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. In *Proceedings of AAAI-2018*, pages 5876–5883.

Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. 2019. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of AAAI-2019*, pages 4602–4609.

B. Pang and L. Lee. 2007. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2).

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia V. Loukachevitch, Evgeniy V. Kotelnikov, Núria Bel, Salud María Jiménez Zafra, and Gülsen Eryigit. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of SemEval@NAACL-HLT-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 SemEval@NAACL-HLT-2015*, pages 486–495.

Sunil Kumar Sahu, Fenia Christopoulou, Makoto Miwa, and Sophia Ananiadou. 2019. Inter-sentence relation extraction with document-level graph convolutional neural network. In *Proceedings of ACL-2019*, pages 4309–4316.

José Saias. 2015. Sentiue: Target and aspect based sentiment analysis in semeval-2015 task 12. In *Proceedings of SemEval@NAACL-HLT-2015*, pages 767–771.

Kaisong Song, Wei Gao, Lujun Zhao, Jun Lin, Changlong Sun, and Xiaozhong Liu. 2019. Cold-start aware deep memory network for multi-entity aspect-based sentiment analysis. In *Proceedings of IJCAI-2019*, pages 5197–5203.

Chi Sun, Luyao Huang, and Xipeng Qiu. 2019. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of NAACL-HLT-2019*, 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*, pages 3298–3307.

Duyu Tang, Bing Qin, and Ting Liu. 2016b. Aspect level sentiment classification with deep memory network. In *Proceedings of EMNLP-2016*, pages 214–224.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2017. Graph attention networks. *CoRR*, abs/1710.10903.

Jingjing Wang, Jie Li, Shoushan Li, Yangyang Kang, Min Zhang, Luo Si, and Guodong Zhou. 2018. Aspect sentiment classification with both word-level and clause-level attention networks. In *Proceedings of IJCAI-2018*, pages 4439–4445.

Jingjing Wang, Shoushan Li, and Guodong Zhou. 2017. Joint learning on relevant user attributes in microblog. In *Proceedings of IJCAI*, pages 4130–4136.

Jingjing Wang, Changlong Sun, Shoushan Li, Xiaozhong Liu, Luo Si, Min Zhang, and Guodong Zhou. 2019a. Aspect sentiment classification towards question-answering with reinforced bidirectional attention network. In *Proceedings of ACL-2019*, pages 3548–3557.

Jingjing Wang, Changlong Sun, Shoushan Li, Jiancheng Wang, Luo Si, Min Zhang, Xiaozhong Liu, and Guodong Zhou. 2019b. Human-like decision making: Document-level aspect sentiment classification via hierarchical reinforcement learning. In *Proceedings of EMNLP-2019*, pages 5580–5589.

Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of EMNLP-2016*, pages 606–615.

Yunjie Wang, Hui Li, and Chen Lin. 2019c. Modeling sentiment evolution for social incidents. In *Proceedings of CIKM-2019*, pages 2413–2416.

Yiming Yang and Xin Liu. 1999. A re-examination of text categorization methods. In *Proceedings of SIGIR-1999*, pages 42–49.

Yuan Yuan, Zhitong Xiong, and Qi Wang. 2019. ACM: adaptive cross-modal graph convolutional neural networks for RGB-D scene recognition. In *Proceedings of AAAI-2019*, pages 9176–9184.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019a. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *CoRR*, abs/1909.03477.

Yingxue Zhang, Soumyasundar Pal, Mark Coates, and Deniz Üstebay. 2019b. Bayesian graph convolutional neural networks for semi-supervised classification. In *Proceedings of AAAI-2019*, pages 5829–5836.