A Joint Model for Aspect-Category Sentiment Analysis with Shared Sentiment Prediction Layer

Abstract

Aspect-category sentiment analysis (ACSA) aims to predict the aspect categories mentioned in texts and their corresponding sentiment polarities. Some joint models have been proposed to address this task. Given a text, these joint models detect all the aspect categories mentioned in the text and predict the sentiment polarities toward them at once. Although these joint models obtain promising performances, they train separate parameters for each aspect category and therefore suffer from data deficiency of some aspect categories. To solve this problem, we propose a novel joint model which contains a shared sentiment prediction layer. The shared sentiment prediction layer transfers sentiment knowledge between aspect categories and alleviates the problem caused by data deficiency. Experiments conducted on SemEval-2016 Datasets demonstrate the effectiveness of our model.

1 Introduction

Aspect-category sentiment analysis (ACSA) is a subtask of aspect-based sentiment analysis (ABSA) (Pontiki et al., 2014; Pontiki et al., 2015; Pontiki et al., 2016). ACSA aims to identify all the aspect categories mentioned in texts and their corresponding sentiment polarities. An aspect category (or simply aspect) is an entity E and attribute A pair, denoted by E#A. For example, in the text "The place is small and cramped but the food is fantastic.", the aspect categories mentioned in the text are *AMBIENCE#GENERAL* and *FOOD#QUALITY*, and their sentiment polarities are negative and negative, respectively.

Many methods have been proposed to address the ACSA task. However, most existing methods (Zhou et al., 2015; Movahedi et al., 2019; Wang et al., 2016; Ruder et al., 2016; Cheng et al., 2017; Xue and Li, 2018; Tay et al., 2018) divide the ACSA task into two subtasks: aspect category detection (ACD) which detects aspect categories in a text and sentiment classification (SC) which categorizes the sentiment polarities with respect to the detected aspect categories, and perform these two tasks separately. Such two-stage approaches lead to error propagation, that is, errors caused by aspect category detection would affect sentiment classification. To avoid error propagation, previous studies (Schmitt et al., 2018; Hu et al., 2019; Wang et al., 2019) have proposed some joint models, which jointly model the detection of aspect categories and the classification of their polarities. Further more, given a text, these joint models detect all the aspect categories mentioned in the text and predict the sentiment polarities toward them at once. Although these joint models obtain promising performances, they train separate parameters for each aspect category and therefore suffer from data deficiency of some aspect categories. For example, the english laptops domain dataset from SemEval-2016 task 5: Aspect-based Sentiment Analysis (Pontiki et al., 2016) has a quarter of the aspect categories whose sample size are less than or equal to 2 (see Table 2). Previous joint models will under-fit on the aspect categories with deficient samples.

To solve the problem caused by data deficiency mentioned above, we propose a novel joint model, which contains a shared sentiment prediction layer. Our model is based on the observation that the sentiment expressions and their polarities of different aspect categories are transferable. For instance, in Table 1, the three aspect categories *LAPTOP#QUALITY*, *LAPTOP#PRICE*, and *LAP-TOP#OPERATION_PERFORMANCE* have the same sentiment word "surprised" and the consistent po-

Aspect Category	Text	Polarity
LAPTOP#QUALITY	...I was **surprised** at the overall quality	positive
LAPTOP#PRICE	and the price...	positive
LAPTOP#OPERATION_PERFORMANCE	...I was **surprised** with the performance	positive
LAPTOP#QUALITY	and quality of this HP Laptop...	positive

Table 1: Different aspect categories have the same sentiment word which have the same sentiment polarity.

larity. The shared sentiment prediction layer transfers sentiment knowledge between aspect categories and alleviates the problem caused by data deficiency.

In summary, the main contributions of our work are as follows:

- We propose a novel joint model for the aspect category sentiment analysis (ACSA) task, which contains a shared sentiment prediction layer. The shared sentiment prediction layer transfers sentiment knowledge between aspect categories and alleviates the problem caused by data deficiency.

- Experiments conducted on SemEval-2016 Datasets demonstrate the effectiveness of our model.

2 Related Work

Existing methods for Aspect-Category Sentiment Analysis (ACSA) can be divided into two categories: two-stage methods and joint models.

Two-stage methods perform the ACD task and the SC task separately. Zhou et al. (2015) and Movahedi et al. (2019) perform the ACD task. Zhou et al. (2015) propose a semi-supervised word embedding algorithm to obtain word embeddings on a large set of reviews, which are then used to generate deeper and hybrid features to predict the aspect category. Movahedi et al. (2019) utilize topic attention to attend to different aspects of a given text. Many methods (Wang et al., 2016; Ruder et al., 2016; Cheng et al., 2017; Xue and Li, 2018; Tay et al., 2018; Liang et al., 2019; Jiang et al., 2019; Xing et al., 2019; Sun et al., 2019; Zhu et al., 2019; Lei et al., 2019) have been proposed for the SC task. Wang et al. (2016) first propose aspect embedding (AE) and use an Attention-based Long Short-Term Memory Network (AT-LSTM) to generate aspect-specific text representations for sentiment classification based on aspect embedding. Ruder et al. (2016) propose a hierarchical bidirectional LSTM (H-LSTM) to modeling the interdependencies of sentences in a review. Tay et al. (2018) propose a method named Aspect Fusion LSTM (AF-LSTM) to model word-aspect relationships. Xue and Li (2018) propose a model, namely Gated Convolutional network with Aspect Embedding (GCAE), which incorporates aspect information into the neural model by gating mechanisms. Jiang et al. (2019) proposed new capsule networks to model the complicated relationship between aspects and contexts. All the two-stage methods have the problem of error propagation.

Joint models jointly model the detection of aspect categories and the classification of their polarities. Only a few joint models (Schmitt et al., 2018; Hu et al., 2019; Wang et al., 2019) have been proposed for ACSA. Schmitt et al. (Schmitt et al., 2018) propose two joint models: End-to-end LSTM and End-to-end CNN, which produce all the aspect categories and their corresponding sentiment polarities at once. Hu et al. (2019) propose constrained attention networks (CAN), which extends AT-LSTM to multi-task settings and introduces orthogonal and sparse regularizations to constrain the attention weight allocation. As a result, the CAN achieves better sentiment classification performance. However, to train the CAN, we need to annotate the multi-aspect sentences with overlapping or nonoverlapping. Wang et al. (2019) propose the aspect-level sentiment capsules model (AS-Capsules), which utilizes the correlation between aspect category and sentiment through shared components including capsule embedding, shared encoders, shared attentions and a shared recurrent neural network. These joint models train separate parameters for each aspect category, which results in that these models under-fit on the aspect categories with deficient samples.
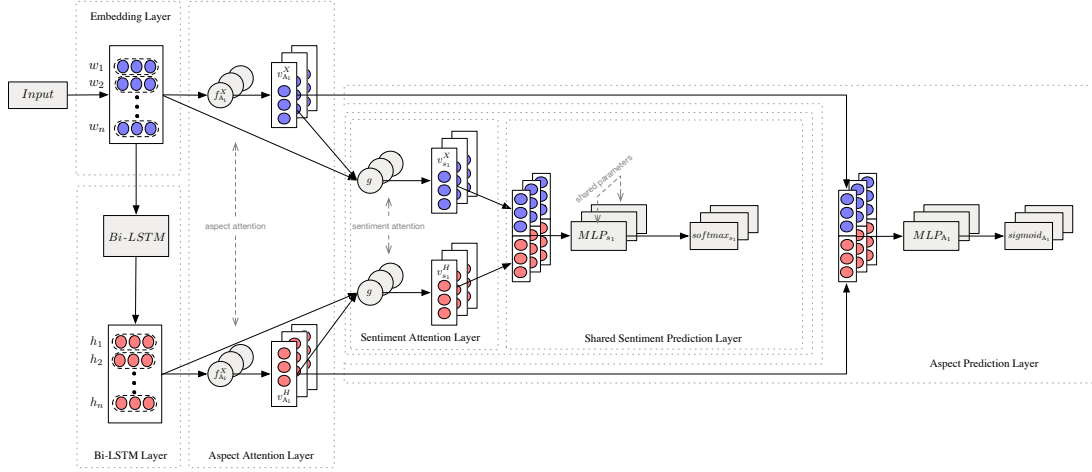
Figure 1: Overall architecture of the proposed method.

3 Proposed Model

We first formulate the problem. There are N predefined aspect categories $A = \{A_1, A_2, \cdots, A_N\}$ and M predefined sentiment polarities $P = \{P_1, P_2, \cdots, P_M\}$ in the dataset. Given a sentence or a review, denoted by $S = \{w_1, w_2, \cdots, w_n\}$, the task aims to predict the aspect categories and the corresponding sentiment polarities, i.e., aspect-sentiment pairs$\{< A_j, P_k >\}$, expressed in the text. The overall model architecture is illustrated in Figure 1, which contains six modules: embedding layer, Bi-LSTM layer, aspect attention layer, sentiment attention layer, aspect category prediction layer, and shared sentiment prediction layer. Then, we display the details of each module and introduce the training objective function.

3.1 Embedding Layer

The input to our model is a text consisting of n words $\{w_1, w_2, ..., w_n\}$. With a embedding matrix U, the input text is converted to a sequence of vectors $X = \{x_1, x_2, ..., x_n\}$. Where $U \in R^{d_w \times |V|}$, d_w is the dimension of the word embeddings, and $|V|$is the vocabulary size.

3.2 Bidirectional LSTM Layer

The word embeddings of the text are then fed into a Bidirectional LSTM (Graves et al., 2013) network (Bi-LSTM) with two LSTM (Hochreiter and Schmidhuber, 1997) networks. We can obtain two hidden representations, and then concatenate the forward hidden state and backward hidden state of each word. Formally, given the sequence of vectors $X = \{x_1, x_2, \cdots, x_n\}$, Bi-LSTM outputs hidden states $H = \{h_1, h_2, \cdots, h_n\}$. At each time step $i = 1, 2, \cdots, n$, the hidden state h_i of the Bi-LSTM is computed by:

$$\overrightarrow{h_i} = \overrightarrow{LSTM}(\overrightarrow{h}_{i-1}, x_i) \tag{1}$$

$$\overleftarrow{h_i} = \overleftarrow{LSTM}(\overleftarrow{h}_{i+1}, x_i) \tag{2}$$

$$h_i = [\overrightarrow{h_i}, \overleftarrow{h_i}] \tag{3}$$

where $\overrightarrow{h_i} \in R^{d_s}, \overleftarrow{h_i} \in R^{d_s}$, $h_i \in R^{2d_s}$, and d_s denotes the size of the hidden state of LSTM.

3.3 Aspect Attention Layer

This layer applies an attention mechanism on the outputs of both the embedding layer and the Bi-LSTM layer and generates aspect-specific representations for the ACD task. Different aspect categories have

different attention parameters. The process can be formulated as follows:

$$v_{A_j}^X = f_{A_j}^X(X), j = 1, \cdots, N \tag{4}$$

$$v_{A_j}^H = f_{A_j}^H(H), j = 1, \cdots, N \tag{5}$$

where f(·) is an attention mechanism (Yang et al., 2016) and can be defined as follows:

$$f(V) = v = \Sigma_{i=1}^n \alpha_i v_i \tag{6}$$

$$u_i = tanh(W_a v_i + b_a) \quad for \quad i = 1, 2, \cdots, n \tag{7}$$

$$\alpha_i = \frac{exp(u_i^T u_w)}{(\Sigma_{j=1}^n exp(u_j^T u_w)} \quad for \quad i = 1, 2, \cdots, n \tag{8}$$

where $V = \{v_1, \cdots, v_i, \cdots, v_n\}$ is a sequence of vectors and $v_i \in R^d$. $W_a \in R^{m \times d}, b_a \in R^m$, and $u_w \in R^m$ are the parameters of the attention mechanism. m is the dimensionality of the attention context vector, and d is the dimensionality of the input vector. Note that the vector v generated by $f(V)$ is a weighted sum of vectors in V and is in the same semantic space with them.

3.4 Aspect Category Prediction Layer

Aspect category prediction layer takes as input the concatenation of the aspect-specific representations at the embedding layer and the Bi-LSTM layer for the ACD task and predicts whether the text mentions the aspect categories. Formally, for the j-th aspect category:

$$v_{A_j} = [v_{A_j}^X, v_{A_j}^H] \tag{9}$$

$$\widehat{y}_{A_j} = \sigma(\widehat{W}_{A_j} ReLU(W_{A_j} v_{A_j} + b_{A_j}) + \widehat{b}_{A_j}) \tag{10}$$

$$\sigma(x) = \frac{1}{(1 + e^{-x})} \tag{11}$$

where $W_{A_j}, b_{A_j}, \widehat{W}_{A_j}$, and $\widehat{b}_{A_j}$ are the parameters of the j-th aspect category. If $\widehat{y}_{A_j}$ is greater than the specified threshold τ, we judge that the j-th aspect category is mentioned by the text.

3.5 Sentiment Attention Layer

This layer generates aspect-specific text representations for the SC task based on aspect-specific text representations for the ACD task. For the j-th aspect category, its aspect-specific text representations for the SC task can be computed as follows:

$$v_{s_j}^X = g(X, v_{A_j}^X) \tag{12}$$

$$v_{s_j}^H = g(H, v_{A_j}^H) \tag{13}$$

where X and H are the outputs of the embedding layer and the Bi-LSTM layer, respectively. $v_{A_j}^X$ and $v_{A_j}^H$ are the aspect-specific text representations of the j-th aspect category for the ACD task at the outputs of the embedding layer and the Bi-LSTM layer respectively. g(·) is an attention mechanism (Vaswani et al., 2017) and can be defined as follows:

$$v_s = g(X, v_q) \tag{14}$$

$$\beta_i = \frac{exp(x_i^T v_q)}{\Sigma_{j=1}^n x_j^T v_q} \quad for \quad i = 1, 2, \cdots, n \tag{15}$$

$$v_s = \Sigma_{i=1}^{n} \beta_i x_i \tag{16}$$

where X is a sequence vectors $\{x_1, x_2, \cdots, x_n\}$, and v_q is the query vector of the attention. We use the dot product to compute attention weights because it does not import extra aspect-specific parameters. Since the query vector and the key vector of the attention are in the same semantic space in our model, the dot product is reasonable.

3.6 Shared Sentiment Prediction Layer

The aspect-specific text representation of the j-th aspect for the SC task is generated by concatenating the aspect-specific text representations of the j-th aspect category for the SC task at the outputs of the embedding layer and the Bi-LSTM layer. The representation are then fed to a fully connected layer with the ReLU activation function and then the output of the fully connected layer is fed to another fully connected layer with the softmax activation function to generate sentiment probability distribution. Formally, for the j-th aspect category:

$$v_{s_j} = [v_{s_j}^X, v_{s_j}^H] \tag{17}$$

$$\widehat{y}_{s_j} = softmax(\widehat{W}_s ReLU(W_s v_{s_j} + b_s) + \widehat{b}_s) \tag{18}$$

$$softmax(x)_i = \frac{exp(x_i)}{\Sigma_{k=1}^{M} exp(x_k)} \tag{19}$$

where W_s, b_s, $\widehat{W}_s$, and $\widehat{b}_s$ are the shared parameters of all aspect categories.

3.7 Loss

For the aspect category detection task, as each prediction is a binary classification problem, the loss function is defined by:

$$L_A(\theta) = -\Sigma_{j=1}^{N} y_{A_j} log \widehat{y}_{A_j} + (1 - y_{A_j}) log(1 - \widehat{y}_{A_j}) \tag{20}$$

For the sentiment classification task, the loss function is defined by:

$$L_s(\theta) = -\Sigma_{j=1}^{N} \Sigma_{k=1}^{M} y_{s_{j_k}} log(\widehat{y}_{s_{j_k}}) \tag{21}$$

if the j-th aspect category is not mentioned in the text, $y_{s_{j_k}} = 0$ for $k = 1, 2, \cdots, M$.

We jointly train our model for the two tasks. The parameters in our model are then trained by minimizing the combined loss function:

$$L_s(\theta) = L_A(\theta) + \eta L_s(\theta) + \lambda ||\theta||_2^2 \tag{22}$$

where η is the weight of sentiment classification loss, λ is the L2 regularization factor and θ contains all the parameters except for Bi-LSTM layer's parameters. Furthermore, to avoid over-fitting, we adopt the dropout strategy to enhance our model.

4 Experiments

4.1 Datasets

We conduct experiments on four public datasets from SemEval-2016 task 5: Aspect-based Sentiment Analysis (Pontiki et al., 2016):

CH-CAME-SB1 is a Chinese sentence-level dataset about digital cameras domain.

CH-PHNS-SB1 is a Chinese sentence-level dataset about mobile phones domain.

EN-REST-SB2 is an English review-level dataset about restaurants domain.

EN-LAPT-SB2 is an English review-level dataset about laptops domain.

We randomly split the original training set into training, validation sets in the ratio 9:1. We use quartiles to measure the distribution of the sample size of aspects in these datasets. Detailed statistics are summarized in Table 2. Particularly, for the three datasets, CH-CAME-SB1, CH-PHNS-SB1, and EN-LAPT-SB2, the sample size of 50% of the aspects are no more than 7.

Dataset	#aspect	#polarity	#train	#val	#test	#min	#Q1	#Q2	#Q3
CH-CAME-SB1	75	2	1090	169	481	1.0	1.0	6.0	13.0
CH-PHNS-SB1	81	2	1152	181	529	1.0	1.8	4.5	23.3
EN-LAPT-SB2	88	4	355	40	80	1.0	2.0	7.0	20.0
EN-REST-SB2	12	4	301	34	90	20.0	36.5	68.5	177.0

Table 2: Statistics of the datasets. #aspect and #polarity represent the number of predefined aspects and sentiment polarities, respectively. #train, #dev, and #test represent the sample size of training sets, validation sets, and test sets, respectively. #min indicates the minimum value of the aspect sample size. #Q1, #Q2, #Q3 are the first quartile, the second quartile, and the third quartile of the aspect sample size, respectively.

4.2 Evaluation Metrics

We use micro-averaged F1-scores as the evaluation metric for both the ACSA and the ACD:

$$F_1 = \frac{2 * P * R}{P + R} \tag{23}$$

where precision and recall are defined as:

$$P = \frac{|S \cap G|}{|S|} \tag{24}$$

$$P = \frac{|S \cap G|}{|G|} \tag{25}$$

Here S is the set of aspect-sentiment pairs or aspect category annotations (in ACSA and ACD, respectively) that a model returns for all the test texts, and G is the set of the gold (correct) aspect-sentiment pairs or aspect category annotations. To evaluate the SC task, we use the gold aspect category annotations to select sentiment polarities model predicts and calculated the accuracy.

4.3 Comparison Methods

We select the following methods for comparison.

End-to-end LSTM (Schmitt et al., 2018) performs the ACSA task, which jointly models the detection of aspects and the classification of their polarities in an end-to-end trainable neural network.

End-to-end CNN (Schmitt et al., 2018) is an CNN version of End-to-end LSTM, which replaces the Bi-LSTM in End-to-end LSTM with a convolutional neural network (CNN) described in (Kim, 2014).

AS-Capsules (Wang et al., 2019) utilizes the correlation between aspect category and sentiment through shared components including capsule embedding, shared encoders, shared attentions and a shared recurrent neural network.

SemEval-2016 Best is the best model for each subtask of SemEval-2016 task 5: Aspect based Sentiment Analysis (Pontiki et al., 2016).

Our Model – w/o Share was added to show the effectiveness of the shared sentiment prediction layer, which trains a separate sentiment prediction layer for each aspect.

4.4 Implementation Details

We implement all models in Keras. We set $\lambda = 0.01$ and gradient clipping norm to 5. Adam (Kingma and Ba, 2014) optimizer is applied to minimize the loss. We apply a dropout of $p = 0.5$ after the embedding layer and the Bi-LSTM layer. Hidden layer size for Bi-LSTM is 100. We use 300-dimensional word embeddings. We use GloVe (Pennington et al., 2014) embeddings which are pre-trained on an unlabeled corpus whose size is about 840 billion for English and Skip-Gram (Mikolov et al., 2013) embeddings which are pre-trained on the Baidu Encyclopedia dataset for Chinese. If an aspect is not mentioned, its corresponding sentiment label is set to a zero vector. We set threshold $\tau = 0.25$ for aspect category

Models	CH-CAME-SB1	CH-PHNS-SB1	EN-LAPT-SB2	EN-REST-SB2
End-to-end cnn	34.96	19.87	36.52	66.03
End-to-end lstm	41.52	26.30	37.94	63.75
AS-Capsules	38.85	27.05	33.47	63.99
Our Model	**42.01**	**28.98**	**50.05**	68.24
– w/o Share	36.23	22.72	49.43	**68.28**

Table 3: Results of the ACSA task in terms of micro-averaged F1-scores(%).

Models	CH-CAME-SB1	CH-PHNS-SB1	EN-LAPT-SB2	EN-REST-SB2
SemEval-2016 Best	80.45	73.34	75.05	81.93
End-to-end cnn	70.55	64.15	69.42	80.78
End-to-end lstm	75.12	67.36	72.00	80.03
AS-Capsules	76.96	71.56	69.05	76.794
Our Model	**82.54**	**76.50**	**75.91**	82.43
– w/o Share	69.09	59.86	70.53	**83.33**

Table 4: Results of the SC task in terms of accuracy(%).

detection. While batch size is 32 on CAME-SB1 and CH-PHNS-SB1, batch size is 10 on EN-LAPT-SB2 and EN-REST-SB2. The sentiment classification loss weight is 1 on CH-CAME-SB1, CH-PHNS-SB1, and EN-LAPT-SB2, and is 0.6 on EN-REST-SB2. To reduce the randomness of results, we train each model three times and report their averaged scores.

4.5 Results

Table 3, Table 4, and Table 5 show our experimental results on the ACSA, SC, and ACD tasks, respectively. The best results are marked in bold.

Table 3 shows the experimental results on the ACSA task, which show the overall performance of our joint model. We observe that our proposed joint model outperforms the baseline models on all datasets, which demonstrates the effectiveness of our model.

The experimental results on the SC task are in Table 4. First, we observe that our model surpasses all baseline models on all datasets, which indicates the effectiveness of our model predicting the sentiment polarities toward given aspect categories. Second, our model outperforms its variant (– w/o Share) by 13.45%, 16.64% and 5.38% on CH-CAME-SB1, CH-PHNS-SB1, and EN-LAPT-SB2 datasets, respectively. The reason is that the three datasets have many aspect categories which only have a few instances and benefit from parameter sharing. This shows that our shared parameter prediction layer can alleviate the problem caused by data deficiency. Meanwhile, our model obtains worse perfromance than its variant (– w/o Share) on the EN-REST-SB2 dataset. The possible reason is that the sample size of the aspect categories in the EN-REST-SB2 dataset is enough to train independent sentiment prediction parameters, and parameter sharing brings some noise between aspect categories.

Models	CH-CAME-SB1	CH-PHNS-SB1	EN-LAPT-SB2	EN-REST-SB2
SemEval-2016 Best	36.3	22.5	60.4	**83.9**
End-to-end cnn	47.83	26.64	42.25	76.20
End-to-end lstm	**52.98**	33.81	43.81	76.24
AS-Capsules	48.72	33.66	40.99	78.29
Our Model	51.81	**36.67**	**62.91**	81.65
– w/o Share	52.16	36.54	62.72	81.45

Table 5: Results of the ACD task in terms of micro-averaged F1-scores(%).

	Text	Label	Our model	-w/o Share
Train set	...The only objection I have is that after you buy it the windows 7 system is a starter and charges for the upgrade...	negtive		
	...The flaws are, this computer is not for computer gamers because of the OS X...	negtive		
Test set	...The OS is easy, and offers all kinds of surprises...	positive	positive	negtive
	...The free upgrade to Mountain Lion FAILED...	negtive	negtive	negtive

Table 6: Impact of the shared sentiment prediction layer on the sentiment prediction of the aspect category *OS#MISCELLANEOUS*.

Table 5 shows the results on the ACD task. Although we did not specifically optimize our model for the ACD task, our model still achieves competitive performance. Specifically, our model outperforms all baselines on the CH-PHNS-SB1 and EN-LAPT-SB2 datasets.

4.6 Case Studies

To have an intuitive understanding of our proposed shared sentiment prediction layer for the SC task, we use the EN-LAPT-SB2 dataset to illustrate the impact of knowledge transferring. The selected aspect category from the dataset is *OS#MISCELLANEOUS*. There are only two samples with both negative polarities in the training set, while there are two samples in the test set, whose polarities are negative and positive, respectively. Table 6 shows that our model can correctly predict the polarity of the sample with positive sentiment in the test set. After removing the shared sentiment prediction layer, – w/o Share fails to predict the polarity of the sample with positive sentiment, which confirms the importance of the shared sentiment prediction layer.

5 Conclusion

In this work, we propose a novel joint model which contains a shared sentiment prediction layer. The shared sentiment prediction layer transfers sentiment knowledge between aspect categories and alleviates the problem caused by data deficiency. Experiments conducted on four datasets from SemEval-2016 task 5 demonstrate the effectiveness of our model. Furture work could consider introducing extra component that prevents the shared sentiment prediction layer from transfering aspect-specific sentiment knowledge.

References

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.

Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. 2013. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Mengting Hu, Shiwan Zhao, Li Zhang, Keke Cai, Zhong Su, Renhong Cheng, and Xiaowei Shen. 2019. Can: Constrained attention networks for multi-aspect 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 4593–4602.

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 6281–6286.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar, October. Association for Computational Linguistics.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Zeyang Lei, Yujiu Yang, Min Yang, Wei Zhao, Jun Guo, and Yi Liu. 2019. A human-like semantic cognition network for aspect-level sentiment classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6650–6657.

Yunlong Liang, Fandong Meng, Jinchao Zhang, Jinan Xu, Yufeng Chen, and Jie Zhou. 2019. 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 5572–5584.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Sajad Movahedi, Erfan Ghadery, Heshaam Faili, and Azadeh Shakery. 2019. Aspect category detection via topic-attention network. *arXiv preprint arXiv:1901.01183*.

Jeffrey Pennington, Richard Socher, and Christopher D 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, 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, Dublin, Ireland, August. Association for Computational Linguistics.

Maria Pontiki, Dimitrios Galanis, Harris 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, Dimitrios Galanis, Harris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, AL-Smadi Mohammad, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphee De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30.

Sebastian Ruder, Parsa Ghaffari, and John G Breslin. 2016. A hierarchical model of reviews for aspect-based sentiment analysis. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 999–1005.

Martin Schmitt, Simon Steinheber, Konrad Schreiber, and Benjamin Roth. 2018. Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1109–1114.

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.

Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

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*, pages 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 sentiment analysis. *arXiv preprint arXiv:1905.07719*.

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 (Volume 1: Long Papers)*, pages 2514–2523.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489.

Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao. 2015. Representation learning for aspect category detection in online reviews. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Peisong Zhu, Zhuang Chen, Haojie Zheng, and Tieyun Qian. 2019. Aspect aware learning for aspect category sentiment analysis. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 13(6):1–21.