Aspect Sentiment Classification with Aspect-Specific Opinion Spans

Lu Xu[*1, 2], **Lidong Bing**[2], **Wei Lu**[1], **Fei Huang**[2]

[1]StatNLP Research Group, Singapore University of Technology and Design
[2]DAMO Academy, Alibaba Group
`xu_lu@mymail.sutd.edu.sg, luwei@sutd.edu.sg`
`{l.bing, f.huang}@alibaba-inc.com`

Abstract

Aspect sentiment classification, predicting the sentiment polarity of given aspects, has drawn extensive attention. Previous attention-based models emphasize using aspect semantics to help extract opinion features for classification. However, these works are either not able to capture opinion spans as a whole or capture variable-length opinion spans. In this paper, we present a neat and effective multiple CRFs based structured attention model that is capable of extracting aspect-specific opinion spans. The sentiment polarity of the target is then classified based on the extracted opinion features and contextual information. The experimental results on four datasets demonstrate the effectiveness of the proposed model, and our further analysis shows that our model can capture aspect-specific opinion spans.[1]

1 Introduction

Aspect Based Sentiment Analysis (ABSA) (Pang and Lee, 2008; Liu, 2012) is an extensively studied sentiment analysis task on a fine-grained semantic level, i.e., opinion targets explicitly mentioned in sentences. Previous ABSA studies focused on a few sub-tasks, such as Aspect Sentiment Classification (ASC) (Wang et al., 2016; Chen et al., 2017; Ma et al., 2018), Aspect Term Extraction (ATE) (Li et al., 2018b; He et al., 2017), Aspect and Opinion Co-Extraction (Liu et al., 2013; Wang et al., 2017; Xu et al., 2018; Dai and Song, 2019), E2E-ABSA (a joint task of ASC and ATE) (Li et al., 2019a; He et al., 2019; Li et al., 2019b), Aspect Sentiment Triplet Extraction (ASTE) (Peng et al., 2019; Xu et al., 2020), etc. ASC analyzes the sentiment polarity of given aspects/targets in a review.

For example, consider the review sentence "***Food** is usually very good, though occasionally I worry about freshness of **raw vegetables** in side orders.*" This review mentions two aspects: **Food** and **raw vegetables**, and for ASC, the objective is to give a positive sentiment on **Food** and a negative sentiment on **raw vegetables**. Most of the previous works (Wang et al., 2016; Chen et al., 2017; Liu and Zhang, 2017; Yang et al., 2017; Li et al., 2018c; He et al., 2018; Li and Lu, 2019; Hu et al., 2019) adopt attention mechanism (Bahdanau et al., 2015) to capture the semantic relatedness among the context words and the aspect, and learn aspect-specific features for sentiment classification.

However, it is challenging for attention-based approaches to consider an opinion span as a whole during feature extraction because they are over-reliant on neural models to learn the context-structural information and perform feature extraction over individual hidden representations. Previous work (Wang and Lu, 2018) engage structured attention networks (Kim et al., 2017), which extend the previous attention mechanism to incorporate structure dependencies, to model the interaction among context words, and perform soft-selections of word spans. In particular, they introduce two hand-coded regularizers to constrain the soft-selection process to attend to few short opinion spans. However, such regularizers disturb the structure dependencies, and their method is not capable of emphasizing aspect-specific opinion spans for sentiment classification.

To better capture opinion features for aspect sentiment classification, we propose the MCRF-SA model, which introduces multiple conditional random fields (CRF) (Lafferty et al., 2001) to structured attention model. While exploiting the advantages of structured attention mechanisms, our model avoids the regularizers by the complementarity among multiple CRFs. We also improve the

---

[1]Our code is released at `https://github.com/xuuuluuu/Aspect-Sentiment-Classification`

previous position decay function (Li et al., 2018a; Tang et al., 2019) to reduce the importance of context words that are further away from the aspect so as to emphasize aspect-specific opinion spans. Our multi-CRF layer with the effective decay function extracts aspect-specific features from different representation sub-spaces to overcome the previous limitations. The experimental results on the four datasets demonstrate the effectiveness of our model, and the analysis shows that the behaviors are in alignment with our intuition.

2 Model Description

Given a context sequence $\mathbf{w^c} = \{w_1, w_2, \ldots, w_n\}$ and a aspect sequence $\mathbf{w^a} = \{w_i, ..., w_j\}$ ($1 \leq i \leq j \leq n$) which is a sub-sequence of $\mathbf{w^c}$, the goal of ASC is to predict sentiment polarity $y \in \{positive, negative, neutral\}$ over the given aspect. Our model is mainly constructed with a few neural layers, including an input layer, an aspect-specific contextualized representation layer, a position decay layer, a multi-CRF structured attention layer, and a sentiment classification layer. Figure 1 presents the architecture of our MCRF-SA model.

2.1 Input Layer

The input of our model consists of word embedding $\mathbf{w}_t^{word}$ and aspect indicator embedding $\mathbf{w}_t^{as}$. The aspect indicator embedding is to differentiate aspect words and context words and is randomly initialized. The input representation $\mathbf{x}_t$ is as follows:

$$\mathbf{x}_t = [\mathbf{w}_t^{word}; \, \mathbf{w}_t^{as}] \qquad (1)$$

2.2 Aspect-Specific Contextualized Representation

We employ a bi-directional GRU (Cho et al., 2014) to generate the contextualized representation. Since the input representation has already contained the aspect information, the aspect-specific contextualized representation is obtained by concatenating the hidden states from both directions:

$$\mathbf{h}_t = [\overrightarrow{\mathbf{h}_t}; \, \overleftarrow{\mathbf{h}_t}] \qquad (2)$$

where $\overrightarrow{\mathbf{h}_t}$ is the hidden state from the forward GRU and $\overleftarrow{\mathbf{h}_t}$ is from the backward.

2.3 Position Decay

Following the previous work (Li et al., 2018a; Zhang et al., 2019; Tang et al., 2019), we also use a position decay function to reduce the influence of
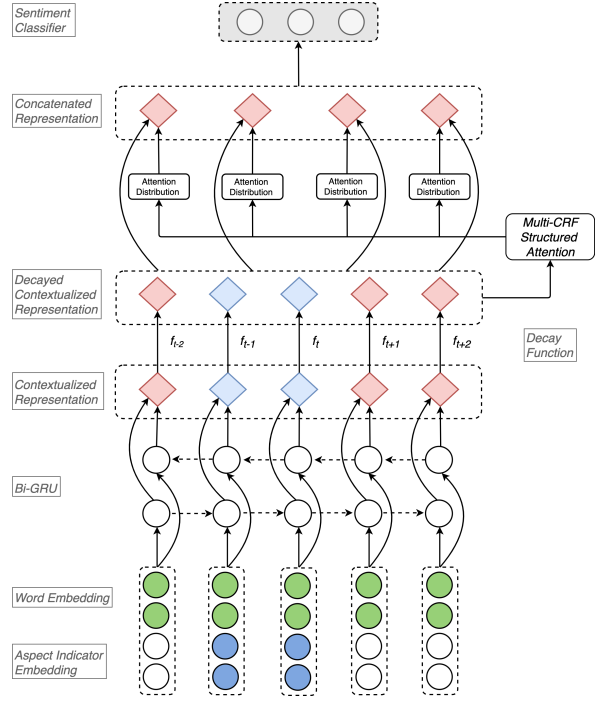

Figure 1: MCRF-SA Architecture.

the context words on the aspect as it goes further away from the aspect. We propose a higher-order decay function, which is more sensitive to distance, and the sensitivity can be tuned by γ on different datasets.

$$f(t) = \begin{cases} (\frac{L-i+t}{L})^\gamma & t < i \\ 1 & i \leq t \leq j \\ (\frac{L-t+j}{L})^\gamma & j < t \end{cases} \qquad (3)$$

where i and j are the starting and ending position of an aspect, L is the maximum length of sentences across all datasets, γ is a hyper-parameter and a larger value enables more influence from the context words that are close to the aspect. Then, the decayed contextual word representation is as follows:

$$\mathbf{r}_t = f(t) \, \mathbf{h}_t \qquad (4)$$

2.4 Multi-CRF Structured Attention

We use multiple linear-chain CRFs to intensively incorporate structure dependencies to capture the corresponding opinion spans of an aspect. In particular, we create a latent label (Wang and Lu, 2018) $z \in \{Yes, \, No\}$ to indicate whether each context word belongs to part of opinion spans. Similar to (Lample et al., 2016), given the sentence representation $\mathbf{x}$, the CRF is defined as:

$$P(\mathbf{z}|\mathbf{x}) = \frac{\exp(score(\mathbf{z}, \mathbf{x}))}{\sum_{\mathbf{z}'} \exp(score(\mathbf{z}', \mathbf{x}))} \qquad (5)$$

where $score(\mathbf{z}, \mathbf{x})$ is a score function that is defined as the summation of transition scores and emission scores from the Bi-GRU:

$$score(\mathbf{z}, \mathbf{x}) = \sum_{t=0}^{n} T_{z_t, z_{t+1}} + \sum_{t=1}^{n} E_{\mathbf{t}, z_t} \quad (6)$$

where T is a transition matrix and $T_{z_t, z_{t+1}}$ denotes the transition score from label z_t to z_{t+1}. $E_{\mathbf{t}, z_t}$ denotes the emission score of label z_t at the t-th position, and the score is obtained from a linear layer, which takes $\mathbf{r}_t$ as input and returns a vector whose length is label size.

2.4.1 Marginal Inference

The latent labels introduced in the CRF layer show whether the word influences the given aspect's sentiment. Intuitively, we can understand that the marginal probabilities on the Yes label indicate the influence of the current context word on the aspect word's sentiment. By using the forward-backward algorithm, we calculate the marginal distribution of the latent label. With the marginal distribution, the sentence representation $\mathbf{s}$ is obtained:

$$\mathbf{s} = \sum_{t=1}^{n} P(z_t = Yes | \mathbf{x}) \mathbf{r}_t \quad (7)$$

The final representation for classification is obtained by concatenating the sentence representations from all CRFs:

$$\mathbf{q} = [\mathbf{s}_1; \mathbf{s}_2; ...; \mathbf{s}_a] \quad (8)$$

where a is the number of CRFs.

2.5 Sentiment Classification

The sentence representation $\mathbf{q}$ is passed to a sentiment classier to obtain the distribution of sentiment polarities:

$$P(\mathbf{y}|\mathbf{q}) = \text{Softmax}(W\mathbf{q} + \mathbf{b}) \quad (9)$$

where W and $\mathbf{b}$ are learnable parameters for the sentiment classifier layer. We learn model parameters by minimizing the negative log-likelihood.

3 Experiments

3.1 Experimental Setup

Our proposed MCRF-SA model is evaluated on four benchmark datasets: SemEval 2014 Task4 (Pontiki et al., 2014), SemEval 2015 Task12 (Pontiki et al., 2015) and SemEval 2016 Task 5 (Pontiki et al., 2016). Following the previous works (Tang et al., 2016; Chen et al., 2017; Wang and Lu, 2018;

Dataset	Train			Dev			Test		
	#Pos.	#Neu.	#Neg.	#Pos.	#Neu.	#Neg.	#Pos.	#Neu.	#Neg.
14Rest	1796	539	666	368	94	139	728	196	196
14Lap	824	383	717	161	72	149	340	167	128
15Rest	808	29	228	147	5	44	340	28	195
16Rest	1106	54	406	191	9	60	474	29	127

Table 1: Statistics of datasets.

He et al., 2018), we remove a few examples that have conflicting labels. Detailed statistics of the datasets can be found in Table 1.

We use the 300d GloVe (Pennington et al., 2014) to initialize our word embeddings. One-sixth of instances are randomly selected from the original training dataset as the development dataset, and the model is only trained with the remaining data. With the development set, we tune our model hyper-parameters using an open-source black-box tuner (Alberto and Giacomo, 2018). We set the hidden size of GRU to 32 or 64. The batch size is set to 64 or 96. The dropout rate is selected from 0.3 to 0.8, with a step size of 0.1. The dimension of the aspect indicator is selected from $\{50, 70, 90\}$. The value of γ in the position decay function is selected from $\{1,2,3\}$. The number of layer of GRU is selected from $\{1,2,3\}$. We adopt Adam (Kingma and Ba, 2014) to optimize our model with a learning rate of 0.008. All hyper-parameters are selected based on the best performance on the development set.

3.2 Baselines

Our MCRF-SA model is compared with the following methods[2]. SVM (Kiritchenko et al., 2014) is a support vector machine based method that integrates surface, lexicon, and parse features. ATAE-LSTM (Wang et al., 2016) is an LSTM (Hochreiter and Schmidhuber, 1997) based model, which has an extra attention to perform soft-selection over the context words. MemNet (Tang et al., 2016) introduces a deep memory network to implement attention mechanisms to learn the relatedness of context words towards the aspect. IAN (Ma et al., 2017) utilizes two LSTM based attention models to learn both context and aspect representations interactively. SA-LSTM-P (Wang and Lu, 2018) employs structured attention networks with multiple regularizers to capture the opinion spans for ASC. TNets (Li et al., 2018a) implements a context-preserving mechanism to get the aspect-specific word representations and uses a Convolutional Neu-

---

[2]Note that our focus is not on exploring the power of pre-trained models (e.g., BERT and ELMo) for ASC.

	Models	14Rest		14Lap		15Rest		16Rest	
		Acc.	F_1	*Acc.*	F_1	*Acc.*	F_1	*Acc.*	F_1
Baselines	SVM (Kiritchenko et al., 2014)	80.16♮	-	70.49♮	-	-	-	-	-
	ATAE-LSTM (Wang et al., 2016)	77.20♮	-	68.70♮	-	-	-	-	-
	MemNet (Tang et al., 2016)	79.61*	69.64*	70.64*	65.17*	77.31*	58.28*	85.44*	65.99*
	IAN (Ma et al., 2017)	79.26*	70.09*	72.05*	67.38*	78.54*	52.65*	84.74*	55.21*
	SA-LSTM-P (Wang and Lu, 2018)	81.60♮	-	75.10♮	-	-	-	88.70♮	-
	TNet-LF (Li et al., 2018a)	80.42*	71.03*	74.61*	70.14*	78.47*	59.47*	89.07*	70.43*
	TNet-ATT (Tang et al., 2019)	81.53♮	72.90♮	77.62♮	73.84♮	-	-	-	-
	ASCNN (Zhang et al., 2019)	81.73*	73.10*	72.62*	66.72*	78.48*	58.90*	87.39*	64.56*
	ASGCN (Zhang et al., 2019)	80.86*	72.19*	74.14*	69.24*	79.34*	60.78*	88.69*	66.64*
Reproduce[3]	TNet-ATT (Tang et al., 2019)	79.38	69.44	76.22	71.51	-	-	-	-
	ASGCN (Zhang et al., 2019)	79.73	70.48	72.91	68.06	78.74	57.67	87.71	70.29
Ours	MCRF-SA	**82.86†**	**73.78†**	**77.64†**	**74.23†**	**80.82†**	**61.59†**	**89.51†**	**75.92†**

Table 2: Experimental results (%). The results with symbol "♮" are retrieved from the original papers, and those with ∗ are retrieved from Zhang et al. (2019). The marker † refers to p-value < 0.01 when comparing with ASGCN.

ral Network (CNN) (Lecun et al., 1998) layer to obtain the sentence representation. TNet-ATT (Tang et al., 2019) is an extension of TNet-LF, and it provides an attention supervision mining mechanism to improve the previous model. ASCNN and ASGCN (Zhang et al., 2019) use CNN and Graph Convolutional Network (GCN) (Kipf and Welling, 2017) to capture the long-range dependencies and syntactic information.

3.3 Experimental Results

Our proposed model shows significant improvements on the four datasets, Table 2 shows the performance comparisons. Our method outperforms SVM (Kiritchenko et al., 2014) by 2.7 and 7.15 *Acc.* score on 14Rest and 14Lap, respectively. This indicates that our neural approach extracts more effective features than hard-coded feature engineering. Compared to the attention-based methods – ATAE (Wang et al., 2016), MemNet (Tang et al., 2016), IAN (Ma et al., 2017), and TNet-ATT (Tang et al., 2019), our MCRF-SA model pays more attention to the aspect-specific opinion spans, which bring significant performance improvement on the four datasets.

We also compare our model with methods that focus on word segmentations for sentiment classification. Our method outperforms the previous regularizers guided structured attention model SA-LSTM-P (Wang and Lu, 2018) by more than 1.2 *Acc.* score on 14Rest and 14Lap. TNet-LF (Li et al., 2018a) and ASCNN (Zhang et al., 2019) em-

---

[3]We train their models with the default parameters and their released training data, and report the average results on our test sets from 3 runs. Note that these works did not release development sets.

ploy CNN to evaluate word spans regarding how much it contributed to the sentiment, but the kernel size limits the length of the span. ASGCN (Zhang et al., 2019) employs GCN over the dependency tree to capture syntactic and dependency information. However, the performance heavily relies on the accuracy of the dependency trees. Our proposed multi-CRF structured attention along with the position decay function allows MCRF-SA to perform soft-selection of multiple aspect-specific opinion spans that influence the aspect's sentiment. The large performance gaps between our model and baseline models confirm the effectiveness of our proposed architecture. Such results also demonstrate that sentiment classification can benefit greatly from aspect-specific opinion spans.

Furthermore, we observe that the performance on 15Rest is not as good as the other three datasets. Such behavior is caused by the different distribution of positive, neutral, and negative sentiment between training and test set, shown in Table 1.

4 Analysis

4.1 Effect of Number of CRFs

To fully investigate the effect of the number of CRFs, we conduct additional experiments on 14Rest and 14Lap with the number of CRFs $\in \{1, 2, 3, ..., 16\}$. Figure 2 shows the experimental results. The model achieves the best performance when the number of CRFs equals to 4. Particularly, the performance becomes relatively plateau when a large number of CRFs is adopted. We believe this is because the sizes of the four benchmark datasets are relatively small, and an excessively large number of parameters may not be able to further extract
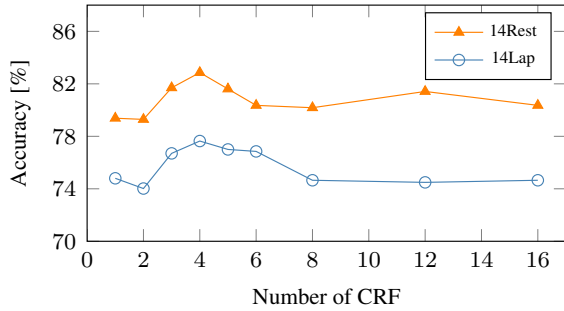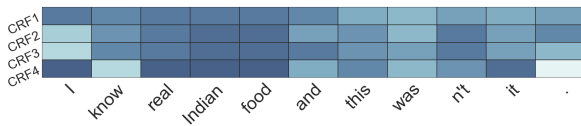
Figure 2: Effect of number of CRFs.

(a) SA-LSTM-P.

(b) MCRF-SA with 4 CRFs.

Figure 3: Marginal distributions of "Yes" label.

effective features.

4.2 Case Study and Error Analysis

Figure 3 shows the marginal distributions (Equation 5) of SA-LSTM-P (Wang and Lu, 2018) and our MCRF-SA model. The aspect for the given example is *"Indian food"* with negative sentiment, and only our model predicts correct sentiment. From Figure 3b heat map, the different marginal distributions on the four CRFs indicate that our model indeed captures different opinion features. It can be observed that MCRF-SA is able to attend to the two major opinion spans: *"real"* and *"n't"*. The SA-LSTM-P model returns positive sentiment as it focuses too much on wrong opinion words.

We also analyze some common errors from our MCRF-SA model, ASGCN, and TNet-ATT on the Lap14 dataset. We observe two major types of errors, and Table 3 shows the examples for error analysis. The first two sentences belong to the type 1 error and the last one presents a type 2 error. The first type of errors appear frequently in neutral cases. In general, the neural models cannot well differentiate if the negative expressions (e.g. *"cost"*, *"shouldn't"*, etc.) is associated with the target/aspect. The second type typically involve complicated sentence structures with non-trivial semantics, which requires advanced language understanding capability.

Case Study	MCRF-SA	ASGCN	TNet-ATT
1. When considering a Mac, look at the total cost of ownership and not just the initial **price tag**$_{NEU}$.	NEU	NEG✗	NEG✗
2. It shouldn't happen like that, I don't have any **design app**$_{NEU}$ open or anything .	NEG✗	NEU	NEU
3. The smaller **size**$_{POS}$ was a bonus because of space restrictions.	NEG✗	NEG✗	NEG✗

Table 3: The words highlighted in blue denote the given aspects, and gold sentiment labels are marked as subscripts. ✗ indicates incorrect prediction.

Models	14Rest		14Lap	
	Acc.	F_1	Acc.	F_1
MCRF-SA	82.86	73.78	77.64	74.23
– aspect indicator	79.02	66.96	72.76	67.56
– decay function	81.52	70.94	76.69	73.12
– structured attention	80.00	68.89	69.61	63.74

Table 4: Ablation Study.

4.3 Ablation Study

We examine the effectiveness of the major components of our MCRF-SA model, and Table 3 presents the ablation results on 14Rest and 14Lap datasets. Without the aspect indicator, our model becomes a sentence-level sentiment classification method which inevitably produces wrong predictions for sentences having multiple aspects with different sentiments. Removing the position decay function hurts the performance by 2.84 and 1.11 F_1 score on 14Rest and 14Lap, respectively. Lastly, without multi-CRF structured attention layer, the architecture becomes a simple Bi-GRU based model and the performance drops significantly by 4.89 and 10.49 F_1 points on 14Rest and 14Lap.

5 Conclusion

We propose a simple and effective MCRF-SA model to extract aspect-specific opinion span features. In particular, with the proposed multi-CRF structured attention layer and the effective position decay function, our model is capable of extracting various aspect-specific opinion span features from different representation sub-spaces. The experimental results demonstrate that our method effectively exploits the corresponding opinion features for sentiment classification. One future direction is to investigate how to integrate the two different attention mechanisms, namely the standard attention and structured attention for NLP applications.

Acknowledgements

References

Costa Alberto and Nannicini Giacomo. 2018. Rbfopt: an open-source library for black-box optimization with costly function evaluations. *Mathematical Programming Computation*.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of ICLR*.

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

Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. In *Proceedings of SSST*.

Hongliang Dai and Yangqiu Song. 2019. Neural aspect and opinion term extraction with mined rules as weak supervision. In *Proceedings of ACL*.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017. An unsupervised neural attention model for aspect extraction. In *Proceedings of ACL*.

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

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*.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*.

Mengting Hu, Shiwan Zhao, H. Guo, Renhong Cheng, and Z. Su. 2019. Learning to detect opinion snippet for aspect-based sentiment analysis. In *Proceedings of CoNLL*.

Yoon Kim, Carl Denton, Luong Hoang, and Alexander M. Rush. 2017. Structured attention networks. In *Proceedings of ICLR*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. In *Proc. of ICLR*.

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

Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. 2014. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of SemEval*.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of ICML*.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proceedings of NAACL*.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*.

Hao Li and Wei Lu. 2019. Learning explicit and implicit structures for targeted sentiment analysis. In *Proceedings of EMNLP*.

Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018a. Transformation networks for target-oriented sentiment classification. In *Proceedings of ACL*.

Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019a. A unified model for opinion target extraction and target sentiment prediction. In *Proceedings of AAAI*.

Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018b. Aspect term extraction with history attention and selective transformation. In *Proceedings of IJCAI*.

Zheng Li, Xin Li, Ying Wei, Lidong Bing, Yu Zhang, and Qiang Yang. 2019b. Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning. In *Proceedings of EMNLP*.

Zheng Li, Ying Wei, Yu Zhang, Xiang Zhang, Xin Li, and Qiang Yang. 2018c. Exploiting coarse-to-fine task transfer for aspect-level sentiment classification. In *Proceedings of AAAI*.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.

Jiangming Liu and Yue Zhang. 2017. Attention modeling for targeted sentiment. In *Proceedings of EACL*.

Kang Liu, Liheng Xu, and Jun Zhao. 2013. Syntactic patterns versus word alignment: Extracting opinion targets from online reviews. In *Proceedings of ACL*.

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

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*.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2019. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *Proceedings of AAAI*.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of EMNLP*, pages 1532–1543.

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 Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of SemEval*.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of SemEval*.

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 SemEval*.

Duyu Tang, Bing Qin, and Ting Liu. 2016. Aspect level sentiment classification with deep memory network. In *Proceedings of EMNLP*.

Jialong Tang, Ziyao Lu, Jinsong Su, Yubin Ge, Linfeng Song, Le Sun, and Jiebo Luo. 2019. Progressive self-supervised attention learning for aspect-level sentiment analysis. In *Proceedings of ACL*.

Bailin Wang and Wei Lu. 2018. Learning latent opinions for aspect-level sentiment classification. In *Proceedings of AAAI*.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2017. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In *Proceedings of AAAI*.

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

Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. 2018. Double embeddings and cnn-based sequence labeling for aspect extraction. In *Proceedings of ACL*.

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. In *Proceedings of EMNLP*.

Min Yang, Wenting Tu, Jingxuan Wang, Fei Xu, and Xiaojun Chen. 2017. Attention based lstm for target dependent sentiment classification. In *Proceedings of AAAI*.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of EMNLP*.