Using Dependency Information to Enhance Attention Mechanism for Aspect-based Sentiment Analysis

Luwen Pu[1], Yuexian Zou[1(✉)], Jian Zhang[2], Shilei Huang[3], and Lin Yao[3]

[1] ADSPLAB/Intelligent Lab, School of ECE, Peking University, Shenzhen,China
{plw,zouyx}@pku.edu.cn

[2] Dongguan University of Technology, Dongguan, China
zjian03@gmail.com

[3] IMSL Shenzhen Key Lab, PKU-HKUST Shenzhen Hong Kong Institute
{shilei.huang, lin.yao}@imsl.org.cn

Abstract. Attention mechanism has been justified beneficial to aspect-based sentiment analysis (ABSA). In recent years there arise some research interests to implement the attention mechanism based on dependency relations. However, the disadvantages lie in that the dependency trees must be obtained beforehand and are affected by error propagation problem. Inspired by the finding that the calculation of the attention mechanism is actually a part of the graph-based dependency parsing, we design a new approach to transfer dependency knowledge to ABSA in a multi-task learning manner. We simultaneously train an attention-based LSTM model for ABSA and a graph-based model for dependency parsing. This transfer can alleviate the inadequacy of network training caused by the shortage of sufficient training data. A series of experiments on SemEval 2014 restaurant and laptop datasets indicate that our model can gain considerable benefits from dependency knowledge and obtain comparable performance with the state-of-the-art models which have complex network structures.

Keywords: Aspect-based Sentiment Analysis, Multi-task Learning, Dependency Parsing, Attention Mechanism.

1 Introduction

Text sentiment analysis or opinion mining is a computational study of opinions, sentiments, evaluations, attitudes, subjectivity, etc., expressed in reviews, blogs, discussions, news, comments or any other documents [14]. With the rapid development of the Internet, social networks, and e-commerce, sentiment analysis has become one of the most active areas of natural language processing (NLP). Whether it is for individuals, groups or countries, the analysis on numerous comments and reviews is also of great practical value.

Aspect-based sentiment analysis is a fine-grained text sentiment analysis task. Given one sentence and one aspect term, its goal is to get the sentiment polarity (usually positive, negative or neutral) towards this opinion target. For example, in the sentence *"great food but the service was dreadful"*, targets are *food* and *service*, and the corresponding sentiment polarities are positive and negative respectively.

Attention mechanism plays a very important role in ABSA task. It can enforce the model to learn the relationship between the aspect and its corresponding contexts. But when sentence become complex, especially when the contexts are far away the target, traditional attention model has limited capacity to capture long-range information [4]. In order to overcome this shortcoming, some researchers utilize dependency relation to fully capture long-range information for certain target. In these works, the dependency tree is used to extract aspect-related features for sentiment classification in traditional machine learning methods and neural network based methods [1, 9, 28], or to establish specific recursive structure used for the input in recursive neural network methods [6, 21, 28]. But these approaches highly rely on the input dependency parsing trees, which are produced by automatic dependency parsers. The trees can have errors, thus suffering from the error propagation problem.

After a deeper analysis of the computational process of the attention-based LSTM model [29], we detect that the calculation of the attention mechanism is actually a part of the graph-based dependency parsing. The attention mechanism is to calculate the relationship between the target and any word in the sentence, while the graph-based dependency parsing will calculate the relationship between any two words in the sentence. So knowledge from graph-based dependency parsing can assist the training of attention networks. In this paper, we combine an attention-based LSTM model and a graph-based dependency parsing model in a multi-task learning manner. One embedding matrix and one LSTM based feature extractor are shared by these two models. We demonstrate our approach's effectiveness through a series of experiments and the visualization of improvement on attention mechanism.

The major contributions of this work are as follows:

- As far as we know, we are the first to detect that the calculation of attention layer is a part of graph-based dependency parsing. Therefore, joint learning with graph-based dependency parsing can help the training of attention layer.
- We propose a general approach for aspect based sentiment analysis, which transfers dependency knowledge to get better aspect-related representation. This architecture is effective for all LSTM based ABSA models.
- We propose an effective method to enhance attention mechanism. It transfers dependency knowledge without using extra dependency parser. In the prediction stage, it can save a lot of computing resources.

The remainder of this paper is structured as follows. Section 2 presents a review of the literature about aspect-based sentiment analysis. The overall design of the proposed approach is described in Section 3. Section 4 presents the experimental settings and analysis. Finally, conclusions and future works are presented in Section 5.

2 Related Work

2.1 Aspect-based Sentiment Analysis

Aspect-based sentiment analysis is typically considered as a classification problem in the literature. As we mentioned above, it can be further treated as a fine-grained classification problem. Traditional methods are based on a series of manually defined features [10,25]. But it is clear that the final result will be heavily dependent on the quality of the features. Moreover, feature engineering is labor intensive.

In later work, methods are turned into neural network-based approaches, like many other NLP tasks. In short, its development can be roughly divided into three stages. Initially, the task is modeled as a sentence classification problem. Assuming that a product has N aspects, the ABSA task is actually a $3N$-classification problem, since every aspect is related to three sentiment polarities: positive, negative, and neutral. The second stage is dominated by recursive neural networks. A lot of recursive neural network based tree structure models [6, 21, 28] are proposed. In the recent stage, most of the works are based on the idea of aspect-based sentential representations [15, 26] which generates a representation of the sentence toward specific aspect. Wang et al. [29] adopt this idea and take advantage of the attention mechanism to generate such representations. Dehong et al. [16] design an interactive attention network(IAN) which uses two attention networks to model the target and context interactively. Tang et al. [30] propose a model named Gated Convolutional network with Aspect Embedding (GCAE), which used the aspect information to control the flow of sentence's sentiment features with CNN and gating mechanisms. Similarly, Huang et al. [8] treat the pooling result of the target as an extra convolution kernel applied on the sentence. There are also researchers who treat ABSA task as a question-answering problem where memory based networks have played a major role [3, 13, 27].

2.2 Isolated Dependency Parsing

Dependency analysis is also widely used in sentiment analysis. Most methods obtain direct or brief dependency features from dependency trees, capturing the relationship between words in the sentence. Xinbo et al. [28] add dependency embedding as additional input when calculate attention weights to capture long-range information for certain target. Tetsuji et al. [28] treat the sentiment polarity of each dependency subtree in a sentence as a hidden variable. The polarity of the whole sentence is calculated in consideration of interactions between the hidden variables. By allowing sentiments to flow from concept to concept based on the dependency relation of the input sentence, Soujanya et al. [24] achieve a better understanding of the contextual role of each concept within the sentence. But they all need additional dependency parsers, usually Stanford Dependency Parser, and are affected by error propagation problem. Moreover, its parsing process also consumes a lot of computing resources.

3 Model

We propose a multitask learning approach to transfer dependency knowledge to the aspect based sentiment analysis model. We will detail the attention-based LSTM model in Section 3.1, the graph-based dependency parsing model in Section 3.2, and the final multitask learning model in Section 3.3.

3.1 Attention-based LSTM

For aspect-based sentiment analysis task, the attention based LSTM model have proven to be useful. It builds a directional LSTM layer to extract the context representation of every word in the input text. After that an attention layer is applied to compute every word's contribution to the aspect and get the final aspect-related representation. The sentiment polarity is computed by a softmax layer finally.

Given n-words input sentence s with words $w_{s1}, w_{s2}, ..., w_{sn}$ and m-words phrase $aspect$ with words $w_{a1}, w_{a2}, ..., w_{am}$, we associate each word w_i with embedding vector $e(w_i)$ from an embedding matrix $E \in R^{V_w \times d_w}$, where V_w is the word vocabulary size and d_w is the word embedding dimension. The aspect representation e_{aspect} is computed as the average of word embeddings of the target words.

$$e_{aspect} = \frac{1}{m} \sum_{i=1}^{m} e(w_{ai}) \tag{1}$$

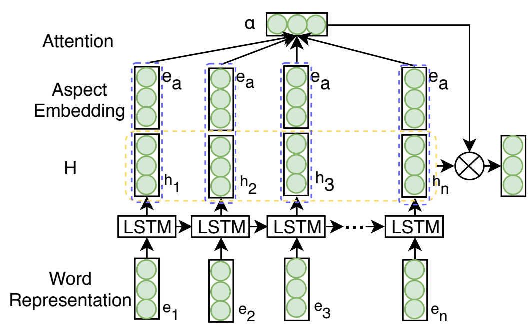


Fig. 1. Illustration of the attention-based ABSA model.

The LSTM layer is used to extract the context representation of every word. The outputs of every time step are the final representation.

$$[v_1, ..., v_n] = LSTM[e(w_{s1}), ..., e(w_{sn})] \tag{2}$$

After that, an attention layer is utilized to compute the weight α_i of each word w_{si} in sentence s to current $aspect$. Its output is the weighted sum of all text features.

$$z = \sum_{i=1}^{n} \alpha_i v_i \tag{3}$$

The α is computed by:

$$t = MLP(e_{aspect}) = W_\alpha e_{aspect} \tag{4}$$

$$\beta_i = f_{score}(v_i, t) = tanh(v_i^T t) \tag{5}$$

$$\alpha_i = Softmax(\beta_i) = \frac{exp(\beta_i)}{\sum_{j=1}^{n} exp(\beta_j)} \tag{6}$$

Before calculating the β_i, we multiply e_{aspect} by W_α. The reason is that one aspect with certain meaning can have several expressions in real scenarios. Taking the laptop as example, the screen can also be expressed as display, resolution and look. Therefore, similar aspect phrases should be grouped into one aspect. Here we use a simple fully connected neural network to achieve aspect phrase grouping. f_{score} is a content-based function that calculates every word's contribution to the target opinion.

At last, a softmax layer is created to predict the probability distribution p over sentiment categories based on the final representation z.

$$p = softmax(W_o z + b_o) \tag{7}$$

The loss function of this model is the cross entropy:

$$loss_{sa} = -\log p_i(c_i) \tag{8}$$

where c_i denotes the true label of current sample and $p_i(c_i)$ denotes the probability of the true label in p.

3.2 Graph-based Dependency Parsing

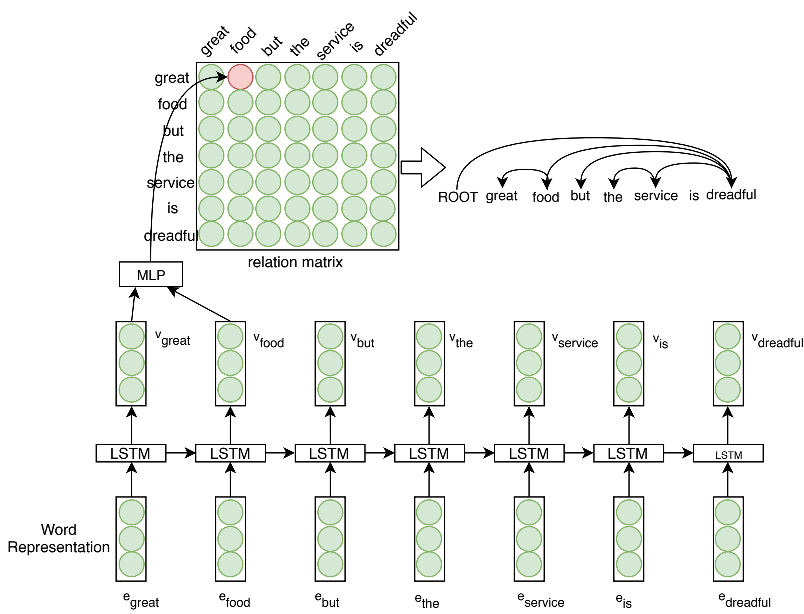


Fig. 2. Illustration of the neural model architecture of the graph-based dependency parser. All of the MLPs share the same parameters. After getting the scores of all possible $n(n-1)$ arcs, the highest scoring tree can be found by using a dynamic-programming algorithm.

We follow arc-factored graph-based dependency parser [18] where the score of a tree is the sum of its all head-modifier arcs (h, m). Considering that the datasets of ABSA task are only annotated with sentiment polarity, we only use the word embedding as the input even though Chen et al. [2] have confirmed that part-of-speech (POS) is more instructive for dependency parsing. The network structure is illustrated in Figure 2.

For graph-based dependency parsing, the process of obtaining LSTM layer's output is the same as that of ABSA model illustrated in Section 3.1. Suppose we have obtained the output of LSTM layer in Formula 2, the score of a head-modifier arc $score(h, m, s)$ is calculated by a simple MLP layer.

$$score(h, m, s) = MLP(v_h \circ v_m) \tag{9}$$

After all the scores of $n(n-1)$ possible arcs are got, finding the highest-scoring dependency tree becomes a problem of maximizing spanning tree in the tree space $Y(s)$. This can be solved efficiently with Eisner's decoding algorithm(1996).

The final model is:

$$\begin{aligned} parse(s) &= \arg \max_{y \in Y(s)} score_{global}(s, y) \\ &= \arg \max_{y \in Y(s)} \sum_{(h,m) \in y} score(h, m, s) \\ &= \arg \max_{y \in Y(s)} \sum_{(h,m) \in y} MLP(v_h \circ v_m) \end{aligned} \tag{10}$$

When training this model, not like [11] we only use the structure loss without using the loss produced from arc label error for that will make the model more complex and harder to be trained. In other words, we only predict the structure of parse trees while ignoring the specific categories of the arcs. The structure loss is a margin-based objective [12, 19], aiming to maximize the margin between the score of the gold tree y and the highest score of predicted parsing tree y':

$$loss_{dp} = \max(0, 1 - \max_{y` \neq y} \sum_{(h,m) \in y'} MLP(v_h \circ v_m) + \sum_{(h,m) \in y} MLP(v_h \circ v_m)) \tag{11}$$

3.3 Multi-task Learning

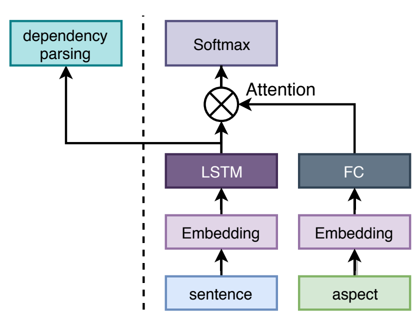


Fig. 3. Illustration of the multi-task learning model. The right part of the dashed line is the baseline model. FC represents fully connected network.

From Formula 5 for attention computing and Formula 10 for dependency parsing, we can see that the calculation of attention mechanism is only a part of graph-based dependency parsing computing. Using sentence *"great food but the service was dreadful"* as an example again, graph-based dependency parsing calculates the relationship between any two words in the sentence, while attention mechanism only calculates the relationship between the target word *"food"* and any other words in the sentence. So the information from the dependency syntax can help the training of attention layer.

We propose to joint learning with graph-based dependency parsing model, inspired by the successful application of deep learning in dependency parsing [2, 11]. Two models share one word embedding layer and one LSTM layer, while other layers are task-specific. The structure of the final model is illustrated in Figure 3. Sentence and aspect share the same word embedding matrix. The overall loss is calculated by:

$$L = loss_{sa} + \lambda loss_{dp} \tag{12}$$

where λ is a hyper-parameter that affects the direction of network optimization.

4 Experiments

4.1 Datasets

For ABSA task, we use two public aspect level annotated datasets: SemEval 2014 Task 4 restaurant and laptop review datasets [23]. The training and test sets have also been provided. Full statistics of these two datasets are presented in Table 1.

Table 1. Statistics of the ABSA task datasets.

Dataset	Train/Test	positive	negative	neutral
Restaurant	Train	2164	807	637
	Test	728	196	196
Laptop	Train	994	870	464
	Test	341	128	169

For graph-based dependency parsing task, we use the Stanford Dependency [5] conversion of the Penn TreeBank (PTB) [17] dataset, with the same train/test splits as [11]. These data are collected from the 1989 *Wall Street Journal*.

4.2 Experimental Settings

Our model is implemented in python, using the DyNet toolkit [20] for neural network training. In all of our experiments, we use 300-dimension GloVe vectors[4] [22] pre-trained on unlabeled data of 840 billion tokens to initialize the embedding layer while all of the other parameters are randomly initialized. Only the top 10,000 words in word frequency are included in the word embedding matrix E while the remaining low-frequency words are replaced by $< unk >$. All parameters are updated with network training. For overall loss in formula 12, λ is set to 0.05 after some attempts. For

[4] https://nlp.stanford.edu/projects/glove/

optimization, a RMSProp optimizer with decay rate and the base learning rate set to 0.001 is used. Other parameters not mentioned above are set to default values provided by DyNet.

4.3 Comparison with Existing Methods

To authoritatively demonstrate the performance of the model, we compare it against the following models:

LSTM+ATT uses the attention mechanism to extract context representation toward the current aspect and then applies a softmax layer to classify.

TD-LSTM [26] integrates the connections between target words and context words when building a learning model. It uses two LSTM networks to capture the connection between target words and their context to generate the target-dependent representation.

ATAE-LSTM [29] utilizes the concatenation of aspect embedding and word embedding as the LSTM layer's input, and then adds a common attention layer to get aspect-related representation. Wang et al. reveal that the sentiment polarity of a sentence is also related to the connected aspect.

MemNet [27] is a memory network based method for ABSA task. It stacks a multi-layer attention model to get the contribution of each context word to the judgment of the sentiment polarity toward current aspect. This model not only greatly exceeds LSTM based models in speed, but also achieves comparable performance with the state-of-the-art feature based SVM systems.

DOC:MULT [7] transfers knowledge from document-level sentiment classification in a multi-task learning manner. It is also based on LSTM+ATT. Document-level labeled data are relatively easily accessible online such as Amazon reviews.

GCAE [30] is a model based on convolutional neural networks and gating mechanisms. It has an additional convolutional layer on aspect terms. And then, the pooling result of the target as an extra convolutional filter applied on the sentence.

We use the commonly used accuracy and macro-f1 as the evaluation metrics. The results are shown in Table 2. Based on them, we have the following observations:

- When compared with LSTM+ATT model, we observe that the dependency knowledge is quite helpful. It brings tremendous improvement to our proposed model in both metrics across all datasets.
- DOC:MULT is another multi-task learning method. When compared with the model without multi-task learning, it has also made great progress. However, when the sentiment polarity of the whole sentence is inconsistent with that of the aspects, this sentence level sentiment information will interfere with the prediction of aspect level sentiment polarities. In that case, knowledge from dependency arcs can still help us find the sentiment words corresponding to the aspects. Therefore, multi-task learning with graph-based dependency parsing can achieve better performance.
- As a multi-task learning method, both DOC:MULT and DP:MULT have greatly improved the performance of the LSTM+ATT model, which also reflects a fact that data is scarce. Current data are not enough to train a very effective neural network based model.

- When we analyze the confusion matrix of test results illustrated in Table 3, we find that the imbalance of samples between classes also bring difficulties to network training. The recall rate of the neutral category is much lower than that of the other two categories. On the one hand, it is due to the ambiguity of the neutral sample itself. On the other hand, too few neutral samples make it difficult for the model to learn neutral-related patterns.

Table 2. The accuracy and macro-f1 results over 5runs of all models on corresponding test set. The results with '*' are retrieved from the papers of compared methods, while the results of other models are retrieved from [7]'s recurrences.

Methods	Restaurant		Laptop	
	Acc	Macro-F1	Acc	Macro-F1
LSTM+ATT	0.7683	0.6648	0.6807	0.6482
TD-LSTM (2016)	0.7537	0.6751	0.6825	**0.6596**
ATAE-LSTM (2016)	0.7860	0.6702	0.6888	0.6393
MemNet (2016)	0.7687	0.6640	0.6891	0.6279
DOC:MULT (2018)	0.7741	0.6668	0.6865	0.6457
GCAE (2018)	0.7728*	-	0.6914*	-
DP:MULT(OURS)	**0.7929**	**0.7036**	**0.7053**	0.6539

Table 3. The confusion matrix of classification results on SemEval 2014 restaurant test dataset.

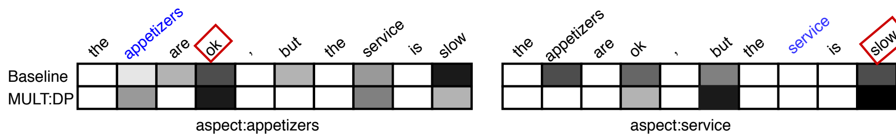
true \ pred	positive	negative	neutral
positive	672	38	18
negative	42	139	15
neutral	87	32	77

4.4 Visualization of Attention Weights

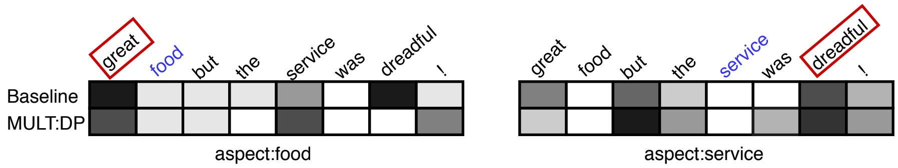
In this section, we pick some testing samples from the dataset and visualize their attention weights. By comparing with the results of the model without multi-task learning, we can confirm whether the dependency information has played its due role. The results are shown in Figure 4. The selected samples all comment on multiple aspects with opposite sentiment polarities, which cannot be analyzed with sentence-level sentiment analysis methods properly.

The key observations are as follows:

- Our model can locate aspect-related sentiment words more accurately 4(b). Even if the review comments on multiple aspects, containing multiple sentiment words, the model still can find those related sentiment words toward specific aspects.
- Higher weights will be given to the aspect-related sentiment words in our model 4(a). This can let the aspect-related representation obtained by the attention layer contain more sentiment information.



(a) comparison between DP:MULT and LSTM:MULT on sentence *"the appetizers are ok, but the service is slow"* towards aspects *"appetizers"* and *"service"*



(b) comparison between DP:MULT and LSTM:MULT on sentence *"great food but the service was dreadful !"* towards aspects *"food"* and *"service"*

Fig. 4. The visualization of the attention weights in DP:MULT model and LSTM+ATT model.

– At the same time, our model will give higher weights to the contrastive connectives, such as *"but"* in figure 4(a) and 4(b), when the sentiments toward different aspects are not the same. This phenomenon is especially obvious when the sentimental tendencies toward different aspects are completely opposite. These words contain abundant dependency information and can bring great benefits to the attention mechanism.

These phenomena all indicate that our model has better attention performance and confirm that dependency knowledge is quite helpful for attention mechanism. This knowledge helps the model get better aspect-related representation and improves the overall performance ultimately.

5 Conclusion

In this paper, we presented a novel approach for aspect-based sentiment analysis based on multi-task learning strategy. As far as we know, we are the first to detect the relationship between attention mechanism and graph-based dependency parsing. We use dependency knowledge to enhance the performance of attention layer and then improve the overall performance. We have demonstrated the effectiveness of our proposed approach and visualized the improvement on attention layer. Our method also has certain versatility. It can be applied to other LSTM-based ABSA models to further boost their performance. In the future, we will look for more effective ways to transfer dependency knowledge for ABSA task and will pay more attention to the identification of neutral comments.

Acknowledgments This work is jointly supported by National Natural Science Foundation of China under the grant No. U1613209 and Shenzhen Fundamental Research Grant under the grant No. JCYJ20180507182908274 and JCYJ20170817160058246.

References

1. Buddhitha, P., Inkpen, D.: Dependency-based topic-oriented sentiment analysis in microposts. In: SIMBig. pp. 25–34 (2015)
2. Chen, D., Manning, C.: A fast and accurate dependency parser using neural networks. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 740–750 (2014)
3. Chen, P., Sun, Z., Bing, L., Yang, W.: Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 452–461 (2017)
4. Daniluk, M., Rocktäschel, T., Welbl, J., Riedel, S.: Frustratingly short attention spans in neural language modeling. arXiv preprint arXiv:1702.04521 (2017)
5. De Marneffe, M.C., Manning, C.D.: Stanford typed dependencies manual. Tech. rep., Technical report, Stanford University (2008)
6. Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., Xu, K.: Adaptive recursive neural network for target-dependent twitter sentiment classification. In: Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers). vol. 2, pp. 49–54 (2014)
7. He, R., Lee, W.S., Ng, H.T., Dahlmeier, D.: Exploiting document knowledge for aspect-level sentiment classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 579–585 (2018)
8. Huang, B., Carley, K.: Parameterized convolutional neural networks for aspect level sentiment classification. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. pp. 1091–1096 (2018)
9. Jiang, L., Yu, M., Zhou, M., Liu, X., Zhao, T.: Target-dependent twitter sentiment classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. pp. 151–160. Association for Computational Linguistics (2011)
10. Kaji, N., Kitsuregawa, M.: Building lexicon for sentiment analysis from massive collection of html documents. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL) (2007)
11. Kiperwasser, E., Goldberg, Y.: Simple and accurate dependency parsing using bidirectional lstm feature representations. Transactions of the Association for Computational Linguistics **4**, 313–327 (2016)
12. LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M., Huang, F.: A tutorial on energy-based learning. Predicting structured data **1**(0) (2006)
13. Li, C., Guo, X., Mei, Q.: Deep memory networks for attitude identification. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. pp. 671–680 (2017)
14. Liu, B.: Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press (2015)
15. Liu, J., Zhang, Y.: Attention modeling for targeted sentiment. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. vol. 2, pp. 572–577 (2017)
16. Ma, D., Li, S., Zhang, X., Wang, H.: Interactive attention networks for aspect-level sentiment classification. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017. pp. 4068–4074 (2017)

17. Marcus, M.P., Marcinkiewicz, M.A., Santorini, B.: Building a large annotated corpus of english: The penn treebank. Computational linguistics **19**(2), 313–330 (1993)
18. McDonald, R., Crammer, K., Pereira, F.: Online large-margin training of dependency parsers. In: Proceedings of the 43rd annual meeting on association for computational linguistics. pp. 91–98 (2005)
19. McDonald, R., Crammer, K., Pereira, F.: Online large-margin training of dependency parsers. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05). pp. 91–98. Ann Arbor, Michigan (June 2005)
20. Neubig, G., Dyer, C., Goldberg, Y., Matthews, A., Ammar, W., Anastasopoulos, A., Ballesteros, M., Chiang, D., Clothiaux, D., Cohn, T., Duh, K., Faruqui, M., Gan, C., Garrette, D., Ji, Y., Kong, L., Kuncoro, A., Kumar, G., Malaviya, C., Michel, P., Oda, Y., Richardson, M., Saphra, N., Swayamdipta, S., Yin, P.: Dynet: The dynamic neural network toolkit. arXiv preprint arXiv:1701.03980 (2017)
21. Nguyen, T.H., Shirai, K.: Phrasernn: Phrase recursive neural network for aspect-based sentiment analysis. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 2509–2514 (2015)
22. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1532–1543 (2014)
23. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: Semeval-2014 task 4: Aspect based sentiment analysis. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). pp. 27–35 (2014)
24. Poria, S., Cambria, E., Winterstein, G., Huang, G.B.: Sentic patterns: Dependency-based rules for concept-level sentiment analysis. Knowledge-Based Systems **69**, 45–63 (2014)
25. Rao, D., Ravichandran, D.: Semi-supervised polarity lexicon induction. In: Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. pp. 675–682 (2009)
26. Tang, D., Qin, B., Feng, X., Liu, T.: Effective lstms for target-dependent sentiment classification. In: COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan. pp. 3298–3307 (2016)
27. Tang, D., Qin, B., Liu, T.: Aspect level sentiment classification with deep memory network. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016. pp. 214–224 (2016)
28. Wang, X., Chen, G.: Dependency-attention-based lstm for target-dependent sentiment analysis. In: Chinese National Conference on Social Media Processing. pp. 206–217. Springer (2017)
29. Wang, Y., Huang, M., Zhao, L., et al.: Attention-based lstm for aspect-level sentiment classification. In: Proceedings of the 2016 conference on empirical methods in natural language processing. pp. 606–615 (2016)
30. Xue, W., Li, T.: Aspect based sentiment analysis with gated convolutional networks. arXiv preprint arXiv:1805.07043 (2018)