A Gated Hierarchical LSTMs for Target-based Sentiment Analysis

Xiaofang Zhang[†*], Bin Liang[*], Qian Zhou[*], Hao Wang[*], Baowen Xu[†]
[†]State Key Laboratory for Novel Software Technology, Nanjing University Nanjing, China
[*]School of Computer Science and Technology, Soochow University Suzhou, China
Email: xfzhang@suda.edu.cn

Abstract—The deep neural model combining attention mechanism has achieved remarkable success in the task of target-based sentiment analysis. In current research, the attention mechanism is more broadly combined with LSTM(Long Short-Term Memory) networks, however, such neural network-based architectures generally rely on complex computation and only focus on the single target. We propose a gated hierarchical LSTMs(GH-LSTM) model of combining regional LSTM and sentence-level LSTM via a gated operation for the task of target-based sentiment analysis. This approach can distinguish different polarities of sentiment of different targets in the same sentence through a regional LSTM, and is able to concentrate on the long dependency of target in the whole sentence via a sentence-level LSTM. The experimental results on multi-domain datasets of two languages from SemEval2016 indicate that, our approach yields better performance than SVM(Support Vector Machine) and several typical neural network models.

I. INTRODUCTION

As a more fundamental task in the field of sentiment analysis, target-based sentiment analysis is capable of digging for subtler descriptions of polarity in terms of different targets in the sentence. It has become one of the hot areas of NLP (Natural Language Processing) during the recent years [1, 2]. Different from the conventional sentiment analysis, target-based sentiment analysis needs to identify the sentiment polarities of different targets in the sentence, which depends on the sentiment information of different targets in the text rather than the context [3]. For example, in sentence "Good food but dreadful service at that restaurant", the sentiment polarity of target "food" is positive while the sentiment polarity of target "service" is negative. This means that even in the same sentence different targets may have completely opposite sentiment polarities.

In the past research, most of the methods based on traditional machine learning had to rely on complicated artificial rules and feature engineering in spite of their remarkable success in tasks of traditional sentiment analysis [4]. Additionally, it is difficult for this kind of methods to do feature extraction and learning for different target words efficiently, so it tended to predict that different targets from the same sentence have the same sentiment polarity.

In recent years, deep learning methods have achieved great progress in many fields. More and more researchers started to apply deep learning methods in NLP tasks [5, 6]. Deep neural network models combined with LSTM get better results than the traditional machine learning methods in NLP tasks with target attached such as target-based relation classification [7], target-based modeling of sentence pairs [8], target-based machine translation [9] and target-based sentiment analysis [10]. This kind of deep network models combined with attention mechanism can highly focus on specific feature information of target and can adjust the parameters of the network for different targets to dig for more hidden feature information. In the task of attention-based sentiment analysis, the network model combined with LSTM can effectively save the dependency relationships between different words by receiving the input of the text sequentially. Meanwhile, through using different target attention matrixes as the input of the network, we could enable the network to pay high attention to the sentiment information of different targets during the training process.

However, LSTM network requires a sequential input of the text. The computation of every word needs to be connected with the information of previous words and the attention matrix and the input word vector need the support from weight matrix, which leads to the complexity of the model's parameter computation [11]. On the other hand, when a sentence contains more targets with complicated dependency relationships, it is hard for the models combined with only the specific attention mechanism for target to identify the sentimental polarities of different targets.

In order to address these problems, in this paper, a gated hierarchical LSTMs model, namely GH-LSTM, for target-based sentiment analysis is proposed. This model could be able to extract both targets(pre-defined by the dataset) and sentence information in the given text. The sole gated hierarchical LSTMs aims to discriminate different polarities of different targets in the same sentence and capture the feature information of the whole sentence. The regional LSTMs receive a regional sequential input including a target, aiming to concentrate on the specific target in the sentence we are considering. The sentence-level LSTMs receive the sequential input of the whole sentence, which can explicitly reveal the important relations between the specific target and the whole sentence. After that regional LSTMs will receive the feature extracted from sentence-level LSTMs modulated by a gated operation.
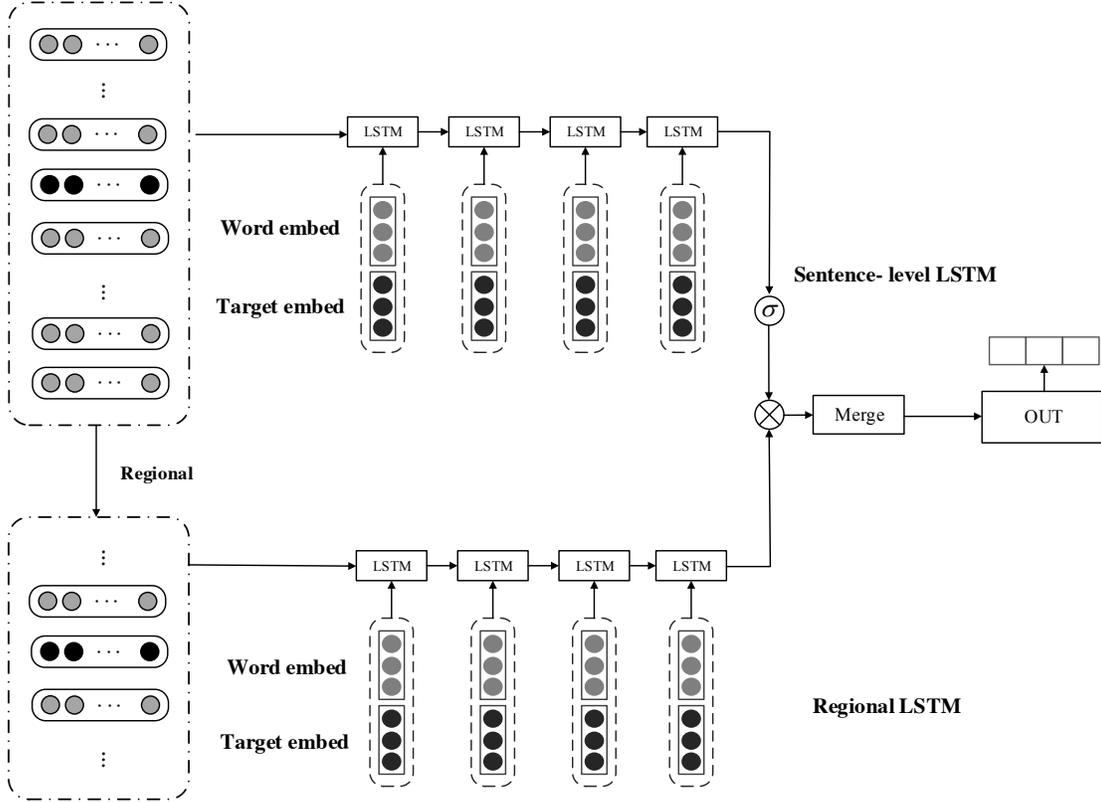
Fig. 1. The illustration of our gated hierarchical LSTMs (GH-LSTM). Regional LSTM receive a region of the sentence divided based on specific target. Sentence-level LSTM receive the whole sentence sequential input. The output of sentence-level LSTM that modulated by a gated operation and the final output of regional LSTM are fed into a merge layer to output a probability distribution of sentiment.

II. RELATED WORK

In this section, we will introduce some related works including target-based sentiment analysis and LSTMs.

A. Target-based sentiment analysis

Target-based sentiment analysis is sentiment analysis on a deeper level and a fine-grained text classification task, which has achieved much attention from researchers since its emergence [12]. Qiu et al. [13] propose a bidirectional back propagation algorithm to determine the sentiment polarity of target by the specific relationship between the sentiment word and the target word. Through the extension of this dictionary, the approach uses the updated dictionary and extracts the relationship between the target words to predict the dependency between the target and the sentiment words. Finally, the sentiment polarity of the target can be obtained. Kiritchenko et al. [14] use a SVM classifier which is combined with multiple features for aspect-level sentiment analysis. This method adds unigram, bigram, sentiment dictionary and other features to the libSVM model to exploit multiple types of sentiment information in the text, making the classifier capable of identifying sentiment polarities in different aspects.

As for the deep learning-based methods, Nguyen and Shi-rai [15] proposed a target-based sentiment analysis model based on RNN(Recursive Neural Network) and dependency tree. This model works on a binary phrase dependency tree containing the element structure and dependency relationship tree of sentence, increasing the correct rate of target-based sentiment analysis considerably and reducing a large amount of feature projects during the task. Dong et al. [16] used an AdaRNN(Adaptive Recursive Neural Network) model to handle target-based sentiment analysis. This model makes use of an adjustable neural network model to learn the connection between the target and the words as well as the syntacticstructure of the sentence. Then the model extends the sentiment information by the relation between the target and the other words to identify the sentiment polarity of the target effectively.

B. Hierarchical model

In recent years, hierarchical models of neural networks have obtained much attention in the field of NLP. Lin et al. [17] propose a hierarchical recurrent neural network language model (HRNNLM) for document modeling by capturing relations between sentences. Li et al. [18] use a hierarchical LSTM auto-encoder to preserve and reconstruct multi-sentence paragraphs and a hierarchical LSTM is used for learning representations of text spans based on attention mechanism [19].

For the task of aspect-based sentiment classification, Ruder et al. [20] introduce a hierarchical bidirectional LSTM (HP-LSTM), which is able to leverage both intra-sentence and inter-sentence relations. This modal inspires us to use a gated hierarchical LSTMs model to leverage both the target and the sentence-level sentiment information.

III. GATED HIERARCHICAL LSTMs MODEL

As is shown in Figure 1, our GH-LSTM model contains following components:

- Regional LSTM: receive a regional sequential input of sentence divided based on specific target, each LSTM unit receives a target embedding amalgamated with word embedding to focus on the specific target information in the process of training.
- Sentence-level LSTM: receive sequential input of sentence including word and target embeddings to extract long dependency of the specific target in the whole sentence.
- Gated merge layer: combine the output of regional LSTM and sentence-level LSTM through a gated operation.
- Fully connected layer: get the sentiment distribution of the specific target in the given sentence.

The targets in the sentence are pre-defined by the dataset. In this section, we will introduce the above components in detail.

A. Task Definition and Aspect Representation

Given a sentence $s = \{w_1, w_2, \ldots, t_i, \ldots, t_j, \ldots, w_n\}$, where t_i and t_j are two different targets in the sentence, the task of target-based sentiment analysis is to discriminate the sentiment polarities of different targets in the same sentence. For example, the sentence "Good food but dreadful service at that restaurant", the sentiment polarity of the target "food" is positive, while for the target "service" it is negative, so even in the same sentence, the opposite polarities can still appear since they belong to different targets. For each word in the sentence, we generate a m-dimensional embedding $\mathbf{x} \in \mathbb{R}^m$ to represent the word and target. Let $\mathbf{x}_i \in \mathbb{R}^m$ be the m-dimensional word vector corresponding to the i-th word, a sentence of length n concatenated with an aspect embedding can be represented as:

$$\mathbf{E}_s = \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \ldots \oplus \mathbf{x}_n \tag{1}$$

where $\oplus$ is the concatenation operation.

B. Region division

Regional LSTM is an improvement of LSTM for handing the tasks with a long text input. We divide each sentence into one or more regions based on targets.[1] The regional LSTM in our work is primarily inspired by a regional CNN(Convolutional Neural Network)-LSTM model for long

---

[1]Unlike using an individual sentence as a region, dividing sentences based on the targets can focus on the most important words of different targets and distinguish the sentence with multiple targets effectively in prediction process.

text sentiment analysis [21]. We divide sentences into specific regions based on specific target, which not only preserves the important feature of specific target in the sentence, but also distinguishes the sentiment information of different targets in sentences availably. Given a sentence $s = \{w_1, w_2, \ldots, t_i, \ldots, t_j, \ldots, w_n\}$, we focus on the targets and the words around the target in the sentence, and divide the sentence into different regions based on different targets. For example, in sentence $s = \{w_1, w_2, \ldots, t_i, \ldots, t_j, \ldots, w_n\}$, where t_i and t_j are two different targets in the sentence, we divide the sentence into two independent regions $r_1 = \{w_{i-l/2}, w_{i-l/2+1}, \ldots, t_i, \ldots, w_{i+l/2}\}$ and $r_2 = \{w_{j-l/2}, w_{j-l/2+1}, \ldots, t_j, \ldots, w_{j+l/2}\}$, where l is the length of a region.

C. Regional LSTM

In this part, we introduce the novel regional LSTM in our approach. As Figure 1 illustrates, the regional LSTM receives a region sequential input matrix of a sentence to focus on the most important information of the specific target. The input of each LSTM unit is composed of the previous output of hidden layer and the current word and target embeddings, defined as:

$$\mathbf{E}_s = \mathbf{W}_h \cdot \mathbf{h}_{i-1} \oplus \mathbf{W}_t \cdot \mathbf{t}_i \oplus \mathbf{W}_x \cdot \mathbf{x}_i \tag{2}$$

where $\mathbf{h}_{i-1}$ is the last hidden output, $\mathbf{W}_h$ is the weight matrix of hidden output. $\mathbf{t}_i$ is the i-th target embedding, $\mathbf{W}_t$ is the weight matrix of target embedding. $\mathbf{x}_i$ is the i-th word embedding, $\mathbf{W}_x$ is the weight matrix of word embedding.

D. Sentence-level LSTM

In order to capture in-depth sentiment information of the sentence, and extract the long-distance dependency of the target across a sentence, sequential inputs composed of word and target embedding are fed into the sentence-level LSTM, as is demonstrated in Figure 1. The input of each LSTM unit is calculated in the same way as regional LSTM, as is shown in formula (2).

E. Gated merge layer

We attempt to combine the information extracted by regional LSTM and sentence-level LSTM to leverage the specific sentiment features in the sentence. The final output of regional LSTM and sentence-level are fed as input to the merge layer controlled by a gated operation:

$$\mathbf{H} = \mathbf{h}_r \otimes \sigma(\mathbf{W}_s \mathbf{h}_s + b) \tag{3}$$

where $\mathbf{h}_r$ is the final output of regional LSTM, $\mathbf{h}_s$ is the final output of sentence-level LSTM, $\mathbf{W}_s$ and b are learned parameters. σ is the sigmoid function and $\otimes$ is the element-wise product between matrices.

F. Model training

Finally, the output $\mathbf{H}_m$ of merge layer is fed into a softmax layer to predict the probability distribution of sentiment.

$$y = soft\max(\mathbf{W}\mathbf{H}_m + b) \tag{4}$$

TABLE I
STATISTICS OF THE DATASETS. DOMAIN OF PHONES AND CAMERAS HAVE
NO NEUTRAL SENTENCE.

Dataset	#Positive	#Negative	#Neutral
REST-Train	1640	736	98
REST-Test	582	178	37
LAPT-Train	1631	1070	180
LAPT-Test	432	261	40
PHNS-Train	748	566	0
PHNS-Test	302	204	0
CAME-Train	802	442	0
CAME-Test	328	124	0

TABLE II
DETAILS OF HYPER-PARAMETER IN OUR EXPERIMENTS.

Hyper-parameter	Value
l_2 constrain	3
Mini-batch	32
Dropout	0.5
Length of region	11

Where $\mathbf{W}$ and b are the parameters for softmax layer. We use the back propagation algorithm to train the model and optimize the model by minimizing the cross entropy error of sentiment classification:

$$loss = -\sum_{i \in D} \sum_{j \in C} \hat{y}_i^j \log y_i^j + \lambda \|\theta\|^2 \qquad (5)$$

where D means all training instances, C is the number of sentiment categories, $\hat{y}$ is the correct distribution of sentiment, y is the predicted sentiment distribution, and $\lambda \|\theta\|^2$ is l_2 regularization.

IV. EXPERIMENTS

A. Datasets and training

We conduct experiments on four datasets of two languages[2] from Semeval2016 Task 5 [22]. Statistics of the datasets are shown in Table 1. We remove sentences with no target which are out of the scope of our task.

We use Glove[3] [23] to initialize word vectors for English and Leipzig Corpora Collection[4] for Chinese. We use 300-dimensional word vectors in our experiments. We train all models with a mini-batch size of 32, dropout rate of 0.5, l_2 regularization weight of 0.001, and the update rule of AdaGrad. The length of regions is 11 words, and we segment Chinese data first. The details of experiments hyper-parameters are shown in Table 2.

[2]The four domain datasets including restaurants, laptops, phones, and cameras, and the languages are English and Chinese. Sentiment classes of restaurants and laptops are positive, negative, and neutral, and domain of phones and cameras are positive and negative.
[3]http://nlp.stanford.edu/projects/glove/
[4]http://corpora2.informatik.uni-leipzig.de/download.html/

TABLE III
BINARY PREDICTION ACCURACY OF OUR RLSTM AND GH-LSTM FOR
TARGET-BASED SENTIMENT CLASSIFICATION ON DIFFERENT DOMAIN
DATASETS IN COMPARISON TO COMPARATIVE MODELS. BEST SCORES ARE
IN BOLD.

Models	REST	LAPT	PHNS	CAME
SVM	81.97	75.90	70.36	76.33
LSTM	81.58	76.62	69.57	76.11
ATT-CNN	82.89	78.64	71.15	76.99
ATT-LSTM	84.87	**80.81**	72.53	78.98
RLSTM	81.97	75.61	70.36	75.22
GH-LSTM	**85.53**	80.66	**73.91**	**79.65**

B. Comparison models

We compare our gated hierarchical LSTMs with several typical and state-of-the-art models, including SVM [14], LSTM [24], ATT-CNN [8], ATT-LSTM [10].

RLSTM: Region LSTM model. This model is part of the layered network model proposed in this paper, RLSTM model only use the region LSTM and is capable to exploit sentiment information for different targets. But it cannot get enough sentiment information of the whole sentence.

GH-LSTM: The complete model of our work, which is able to distinguish different targets information in the same sentence and capture the long dependency of target across the review.

SVM: A feature-based SVM classification model. This model got a better result in experiment of target-based customer reviews classification than the previous researches, but it needs some extra features.

LSTM: Standard LSTM without any attention of target that cannot infer the sentiment polarity of different targets in the same sentence exactly.

ATT-CNN: An attention-based CNN model that achieves state-of-the-art performance on sentence pairs modeling. We use a similar model for receiving the word embedding and target embedding in the experiment that is able to highly focus on the target in the process of training. But the structure of the model is relatively complex and the effect is quite dependent on the attention matrix.

ATT-LSTM: An attention-based LSTM that can concentrate on different parts of a sentence for different targets. This model achieves state-of-the-art performance on aspect-level sentiment classification. But the model needs a high training time cost.

C. Comparative results

The binary prediction (positive and negative) results are shown in Table 3. Our models achieve the second best result on laptops dataset (ATT-LSTM achieve the best result) and achieve the best results on other 3 domain datasets. In addition, the results of experiment including neutral polarity category are presented in Table 4. We can find that, similar to the results of binary prediction, our models achieve better performance than other comparative models on all domain datasets, indicating that our model can discriminate the sentiment polarity

Models	REST	LAPT
SVM	77.85	70.94
LSTM	78.80	71.62
ATT-CNN	81.18	74.49
ATT-LSTM	82.69	76.26
RLSTM	79.29	71.49
GH-LSTM	**83.35**	**76.81**

Method	Time cost
LSTM	108
ATT-CNN	62
ATT-LSTM	324
RLSTM	84
GH-LSTM	167

of different targets and capture sufficient feature information in more complex sentiment categories via a gated hierarchical network. Another phenomenon is that the GH-LSTM which using hierarchical input layer and gated operation can improve the accuracy in contrast to RLSTM on all experiments, revealing that the hierarchical input layer of sentence-level LSTM and gated merge layer are valid in our approach.
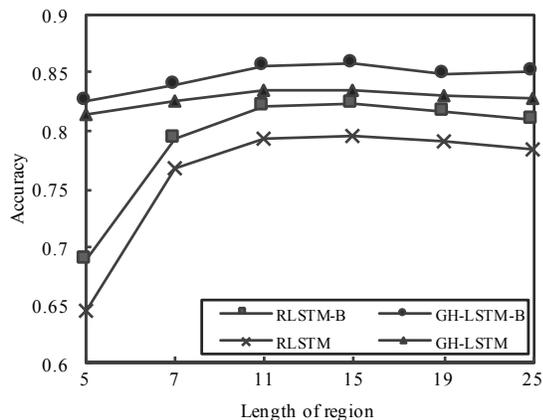
Fig. 2. Classification accuracy of our RLSTM and GH-LSTM with different length of region on restaurants dataset. "-B" represents binary prediction.

D. Effects of the length of region

In order to validate the effectiveness of dividing regions based on targets and the effects of using gated hierarchical LSTMs, we implement different lengths of region in GH-LSTM and RLSTM on restaurants dataset. As shown in Figure 2, before the length is 11, both GH-LSTM and RLSTM achieve better performance when the length of region is increased. But the results fluctuated when the length of the region is greater than 11. So the value of the length of the region is 11 in our experiment. We can also find that when the length is very small, GH-LSTM can also get a good prediction results, while RLSTM does bad when the length is 5, so that the gated hierarchical LSTM can improve the classification when the model lacks regional information.

E. Runtime analysis

We study the runtime of our approach and comparative models. We implement all these approaches based on the same

neural network infrastructure, and run them on the same CPU and GPU server. As shown in Table 5, LSTM combining attention mechanism is indeed computationally expensive, the ATT-LSTM costs 324s during each training epoch. Our RLSTM is almost 4 times faster than ATT-LSTM and faster than basic LSTM, our GH-LSTM is also save half of the time contrast to ATT-LSTM.

V. CONCLUSION

We propose a novel deep gated hierarchical LSTMs model for target-based sentiment analysis. Our model not only can effectively identify the sentiment polarity of different targets, but also can obtain long distance dependencies of specific targets in the whole input sentence and extract more hidden information of specific targets. Finally, the experiment results on the 4 datasets of two languages illustrated the validity of our model by comparing with the excellent models in previous studies.

ACKNOWLEDGMENT

REFERENCES

[1] M. Pontiki, D. Galanis, J. Pavlopoulos *et al.*, "Semeval-2014 task 4: Aspect based sentiment analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Dublin, Ireland: Association for Computational Linguistics and Dublin City University, August 2014, pp. 27–35.

[2] B. Pang, L. Lee *et al.*, "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.

[3] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proceedings of the 2nd international conference on Knowledge capture*. ACM, 2003, pp. 70–77.

[4] E. Boiy and M.-F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," *Information retrieval*, vol. 12, no. 5, pp. 526–558, 2009.

[5] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, October 2014, pp. 1746–1751.

[6] P. Liu, X. Qiu, and X. Huang, "Adversarial multi-task learning for text classification," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics, July 2017, pp. 1–10.

[7] P. Zhou, W. Shi, J. Tian *et al.*, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Berlin, Germany: Association for Computational Linguistics, August 2016, pp. 207–212.

[8] W. Yin, H. Schütze, B. Xiang *et al.*, "Abcnn: Attention-based convolutional neural network for modeling sentence pairs," *Transactions of the Association of Computational Linguistics*, vol. 4, no. 1, pp. 259–272, 2016.

[9] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[10] Y. Wang, M. Huang, X. Zhu *et al.*, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, November 2016, pp. 606–615.

[11] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, November 2016, pp. 214–224.

[12] B. Liu, "Sentiment analysis and opinion mining," *Synthesis lectures on human language technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[13] G. Qiu, B. Liu, J. Bu *et al.*, "Opinion word expansion and target extraction through double propagation," *Computational linguistics*, vol. 37, no. 1, pp. 9–27, 2011.

[14] S. Kiritchenko, X. Zhu, C. Cherry *et al.*, "Nrc-canada-2014: Detecting aspects and sentiment in customer reviews," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 2014, pp. 437–442.

[15] T. H. Nguyen and K. Shirai, "Phrasernn: Phrase recursive neural network for aspect-based sentiment analysis," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, September 2015, pp. 2509–2514.

[16] L. Dong, F. Wei, C. Tan *et al.*, "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)*. Baltimore, Maryland: Association for Computational Linguistics, June 2014, pp. 49–54.

[17] R. Lin, S. Liu, M. Yang *et al.*, "Hierarchical recurrent neural network for document modeling," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, September 2015, pp. 899–907.

[18] J. Li, T. Luong, and D. Jurafsky, "A hierarchical neural autoencoder for paragraphs and documents," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Beijing, China: Association for Computational Linguistics, July 2015, pp. 1106–1115.

[19] Q. Li, T. Li, and B. Chang, "Discourse parsing with attention-based hierarchical neural networks," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, November 2016, pp. 362–371.

[20] S. Ruder, P. Ghaffari, and J. G. Breslin, "A hierarchical model of reviews for aspect-based sentiment analysis," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. Austin, Texas: Association for Computational Linguistics, November 2016, pp. 999–1005.

[21] J. Wang, L.-C. Yu, K. R. Lai *et al.*, "Dimensional sentiment analysis using a regional cnn-lstm model," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Berlin, Germany: Association for Computational Linguistics, August 2016, pp. 225–230.

[22] M. Pontiki, D. Galanis, H. Papageorgiou *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," in *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. San Diego, California: Association for Computational Linguistics, June 2016, pp. 19–30.

[23] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, October 2014, pp. 1532–1543.

[24] X. Wang, Y. Liu, C. Sun *et al.*, "Predicting polarities of tweets by composing word embeddings with long short-term memory," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Beijing, China: Association for Computational Linguistics, July 2015, pp. 1343–1353.