Journal of Computers Vol. 30 No. 4, 2019, pp. 176-184
doi:10.3966/199115992019083004017

# An Extended Attention-based LSTM with Knowledge Embedding for Aspect-level Sentiment Analysis

Bo Hu[1,2], Yun Liu[1,2*], Ming-dong Sun[3], Kun Mi[3]

[1] School of Electronic and Information, Beijing Jiaotong University, Beijing 100044, China
{18111018, liuyun}@bjtu.edu.cn

[2] Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education
Beijing 100044, China

[3] Hua Yu Information Technology Co., Ltd. Beijing, China
{sunmd, mik}@thunisoft.com

Received 13 January 2019; Revised 7 March 2019; Accepted 12 March 2019

Abstract. Aspect-level sentiment analysis is a highly concerned content of sentiment analysis in recent years. For sentiment analysis, however, some of the necessary external knowledge such as commonsense is very important, and these models can't acquire this knowledge through training data. To this end, in this paper we propose an extension of the LSTM based on attention mechanism to combine explicit knowledge with tacit knowledge to improve the accuracy of the aspect-level sentiment analysis model.

Keywords: sentiment analysis, attention model, commonsense knowledge

## 1 Introduction

Sentiment analysis [1], also known as opinion mining [2], is one of the key tasks of NLP. We need to analyze the sentiment tendency of groups to an entity under multi-source heterogeneous big data, that is, classify content such as text, audio or video as *positive, negative or neutral* [3]. Early sentiment analysis focused on analyzing the overall emotional propensity of each part of the text [4], that is, based on a unified hypothesis that the whole expressed the same emotional polarity. However, researchers have found that sentiment analysis requires finer granularity to improve accuracy, which is to refine the analysis object to specific aspects of the entity. For example, the sentiment polarity in the comment "*He looks good, but his acting is poor*" is positive when the aspect is *appearance* but will be negative if the aspect is *acting skill.*

Neural networks have made remarkable progress in NLP fields such as machine translation [5] and intelligent answering. The LSTM forgetting mechanism mimics the function of the human brain. However, in the direction of fine-grained sentiment analysis, the effect still needs to be improved. Some papers propose Target-Dependent LSTM (TD-LSTM) and Target-Connection LSTM (TC-LSTM) [6], which refines sentiment analysis to the level of the target. However, these models cannot analyze the sentiment polarity of text on different aspects of the target.

In order to improve the accuracy of the sentiment analysis model, some papers have proposed Attention-based LSTM with Aspect Embedding (ATAE-LSTM) [7] to extend the LSTM model to focus on important parts of specific aspects of the sentence. The model uses an attention mechanism to determine different parts of a sentence when it covers different aspects of the entity.

However, when there is no clear common sense in the dataset, it is difficult for the model to learn explicit knowledge such as common sense knowledge. Therefore, we are concerned about the unresolved issues of the most advanced methods. In order to learn the external knowledge not included in the training data, many applications have appeared to process text [8-10] and images [11-12]. Since the

* Corresponding Author

commonsense is closely related to polarity in sentiment analysis, we propose an improved aspect-based sentiment analysis model based on the attention mechanism, by introducing external emotional common sense.

Our main contributions are:

1. We propose an improved attention-based LSTM model for aspect-level sentiment analysis. When a comment relates to different aspects of the entity, the model can determine the key parts of the sentence, and combine external knowledge to make emotional polarity judgments.

2. Because of the need to efficiently use external knowledge, we combine the aspect embedding methods and external components to integrate external knowledge, enabling the extended ATAE-LSTM model to handle common sense knowledge and output emotional polarity.

3. The experimental results show that our method improves performance compared with the baseline model. Further examples show that the attention mechanism that introduces the knowledge of external common sense plays a good role in aspect-level sentiment analysis.

## 2 Related Work

In this section, we briefly review some of the work done by sentiment analysis neural networks and sentiment analysis models based on attention mechanisms.

### 2.1 Sentiment Analysis with Neural Networks

Neural networks [13] improve the accuracy of sentiment analysis by effectively learning distributed knowledge representation. These include classical models such as recurrent neural network [14-15], recurrent neural tensor network, and LSTM. These models have achieved effective results in solving many NLP tasks. However, RNN has syntax errors due to gradient disappearance and gradient explosion problems.

In many NLP missions, LSTM has achieved significant results. However, averaging only the embedded words of the target phrase, while ignoring different aspects of the target entity that may appear in the sentence, is not enough to indicate the emotional polarity of the target entity in the sentence, resulting in loss of accuracy.

### 2.2 Aspect-Based Sentiment Analysis

Although the above method has achieved good results, it is still a task to be completed to analyze the emotional polarity of the target entity in a finer granularity. Therefore, some papers propose to classify emotional polarity at the aspect level to achieve better performance.

As we mentioned earlier, the biggest problem in aspect-level emotion classification that needs to be solved is to effectively represent the emotional information in the sentence about the specific aspects of the entity description. The use of attention mechanisms to embed word embedding and aspects into the join input can enhance the representation. For a word in a sentence, the attention vector quantifies its sentiment conspicuousness and its relevance to a given aspect.

Some work combines external knowledge with deep neural networks and is closely related to sentiment analysis. An external knowledge base can be used as a source for determining the characteristics of an entity.

## 3 Extended ATAE-LSTM with Knowledge Embedding

### 3.1 Long Short-term Memory (LSTM)

Some papers have proposed long-short-term memory networks to protect and control information through forgetting gates, input gates, and output gates. Fig. 1 shows the structure of the standard LSTM.

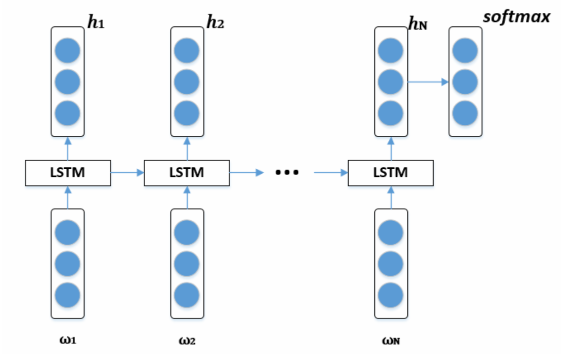


**Fig. 1.** The structure of the standard LSTM

The given sentence is of length N, and $\{w_1, w_2, \dots, w_N\}$ is the word vector of this sentence. $\{h_1, h_2, \dots, h_N\}$ represents the hidden vector. Specifically, each cell in the standard model will be calculated as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * tanh(C_t) \tag{6}$$

In these representations, $w_f, w_i, w_o$ are the weighting matrix, and $b_f, b_i, b_o$ are deviation of the LSTM learning during training, as parameters to represent the conversion process of the input gate, the forgetting gate and the output gate. $C_t$ represents the cell state at time t. The sigmoid function is denoted by $\sigma$ and elements multiplication by *. $h_{t-1}$ represents the vector output by hidden layer of the previous cell, And $x_t$ represents the input of the current cell. $x_t$ also represents the word embedding vector in Fig. 1.

The last vector $h_N$ of the hidden layer output is used to represent the sentence, and the soft max function will normalize the vector to a three-dimensional vector. We classify sentiment polarity labels into three categories, namely *{positive, negative, neutral}*.

### 3.2 Attention-based LSTM with Aspect Embedding (ATAE-LSTM)

When performing sentiment analysis on a given entity's comments, different aspects may have different polarities. Therefore, the standard LSTM model needs to be improved by learning aspect embedding. But when analyzing a sentence, the standard LSTM model itself cannot determine which part of the sentence is a critical part related to a given aspect.

The ATAE-LSTM model [7] solves the above problem. The model combines the learning of aspect embedding vector and attention mechanism to apply aspect-level information to the training process while focusing on the key parts of the sentence for a given aspect. Fig. 2 shows the structure of the ATAE-LSTM.

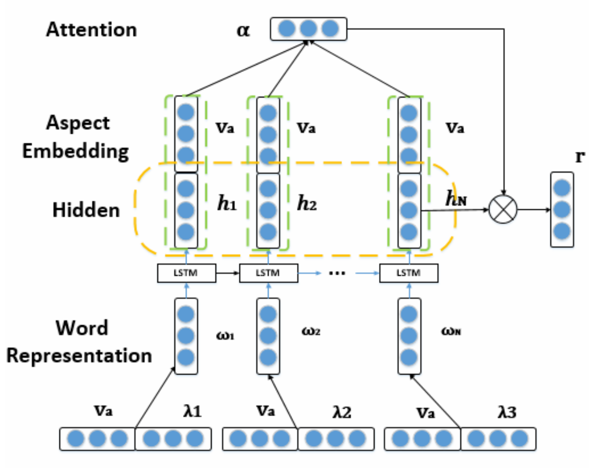


**Fig. 2.** The structure of the ATAE-LSTM

First, when determining the weight of attention, the ATAE model introduces aspect embedding, which participates in the decision participation with the sentence representation of the LSTM hidden layer output. The given sentence is of length N, and $\{w_1, w_2, \ldots, w_N\}$ is the word vector of this sentence. $\{h_1, h_2, \ldots, h_N\}$ represents the hidden vector.

Furthermore, aspect embedding vector $v_a$ constitutes matrix A. The attention weight vector α is generated by the attention mechanism, and the process is performed by the hidden layer output of the LSTM model interacting with the aspect vector.

The interaction mode and the weighted hidden representation r are generated as follows:

$$M = \tanh(\begin{bmatrix} W_h H \\ w_v v_a * e_N \end{bmatrix}) \tag{7}$$

$$\alpha = softmax(\omega^T M) \tag{8}$$

$$r = H\alpha^T \tag{9}$$

Where matrix H consists of hidden vectors $\{h_1, h_2, \ldots, h_N\}$, and $e_N$ is a vector of 1s.

$$h^* = \tanh(W_p r + W_x h_N) \tag{20}$$

Where $h^*$ represents the final sentence representation. And $W_p$ and $W_x$ are parameters that need to be learned during training. Finally, the *softmax* layer will convert the final sentence representation into a conditional probability distribution ($W_s$ and $b_s$ are the parameters of this layer):

$$y = softmax(W_s h^* + b_s) \tag{31}$$

Furthermore, in order for the model to make fuller use of the aspect information knowledge, when the vector of each word is input, the aspect is embedded and appended with the word vector. So aspect embedding can participate in the generation of the output of the hidden layer together with the word embedding of the sentence. That is to say, in the process of calculating $\alpha$, the relationship between words and aspects is also taken into account.

### 3.3 Extended ATAE-LSTM with Knowledge Embedding

Existing work has proven that ATAE-LSTM effectively improves the performance of aspect-level sentiment analysis relative to other models. However, in the face of the task of accurate sentiment analysis, the ATAT-LSTM model still needs some improvement. In this paper, we consider embedding emotional common sense knowledge as an external source of knowledge into ATAE-LSTM, since it is difficult for the model to learn explicit knowledge itself. The purpose of introducing common sense knowledge is to connect concepts to the aspects or sentiment polarity that are closest to it. For example, the word *"battery life"* obviously has the *'electronic product'* attribute, which can point to *"laptop"*, *"mobile phone"* or *"electronic product performance"*. And some common expressions of feelings in common sense knowledge can help the sentiment analysis system to better calculate emotional polarity.

Based on the above, we propose an extended ATAE-LSTM model: more effective level-level sentiment analysis by introducing external common sense knowledge. Fig. 3 shows the structure of the extended ATAE-LSTM, where $\{\lambda_1, \lambda_2, \ldots, \lambda_{1N}\}$ represents the knowledge embedding vector.

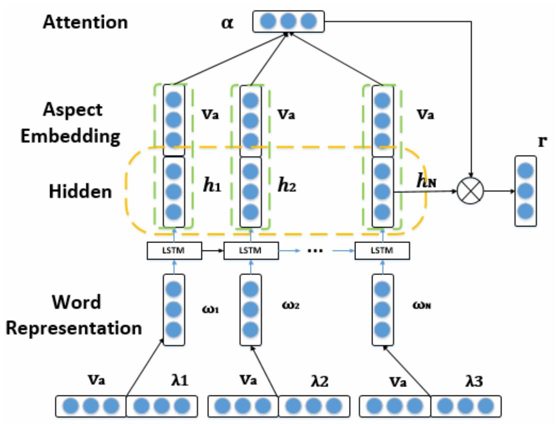


**Fig. 3.** The structure of the extended ATAE-LSTM

As shown in Table 1, the AffectNet dataset contains a wide range of common sense concepts related to emotional attributes. These data not only provide common sense concept features, but also provide semantic connections between these concepts and aspects, emotional polarity. If there is a strong semantic connection between the concept and the attribute or emotional polarity, the weight under the attribute will be larger. If the connection is weaker, the weight is lower. If there is no connection, the weight is zero.

**Table 1.** A snippet of the AffectNet matrix

| AffectNet | IsA-pet | KindOf-food | Arises-joy | …… |
|---|---|---|---|---|
| Win lottery | 0 | 0 | 0.991 | …… |
| Songbird | 0.672 | 0 | 0.862 | …… |
| Gift | 0 | 0 | 0.899 | …… |
| Bunny | 0.611 | 0.892 | 0.594 | …… |
| rattlesnake | 0.432 | 0.235 | 0 | |

However, because of the wide range of concepts and attributes, the AffectNet dataset has a wide range of dimensions and there is a large problem of sparse data. The dimension needs to be reduced before being applied to the neural network.

The vector space model, AffectiveSpace2, is built by random projection, which allows analogy to reason about natural language concepts. By reducing the dimensions of emotional common-sense knowledge, the model allows for the abstraction of semantic features associated with concepts, allowing concepts to be intuitively clustered according to their semantic and emotional relationships. So we use AffectiveSpace2 to map the concept of AffectNet to continuous low-dimensional embedding. This maximizes the various semantic associations in the original data. Based on this new vector space model, we embed the concept level information into the ATAE-LSTM model to better classify the aspects and emotions of the sentence.

First, we calculate the candidate embedding into a single vector. Where K represents the number of elements in the concept set $\{\lambda_{i,1}, \lambda_{i,2}, \lambda_{i,3}, \ldots, \lambda_{i,k}\}$.

$$\lambda_i = \frac{1}{K}\Sigma_j \lambda_{i,j} \tag{42}$$

We assume that at each time step *t,* a set of knowledge concept candidates can be triggered and mapped to a low dimensional space. These vectors are concatenated in the input layer with the aspect embedding vector to provide additional information to the memory cells. Our model is illustrated by the following formula:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t, \lambda_t] + b_f) \quad \textbf{(53)}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t, \lambda_t] + b_i) \quad \textbf{(14)}$$

$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad \textbf{(15)}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t) \quad \textbf{(16)}$$

$$o_t^c = \sigma(W_{co} \cdot [h_{t-1}, x_t, \lambda_t] + b_{oc}) \quad \textbf{(17)}$$

$$h_t = o_t * tanh(C_t) + o_t^c * tanh(W_c \lambda_t) \quad \textbf{(18)}$$

Through these steps, our work adds knowledge concepts to the ATAT-LSTM forgotten, input and output gates to help filter extraneous information to a greater extent. The role of the sentiment concept existing in the input gate is to prevent the memory unit from being affected by pre-existing knowledge conflicts. Further, the output gate uses this knowledge to filter out irrelevant information stored in memory. Therefore, this extended ATAE-LSTM model can help improve the emotional polarity calculation of the aspect-level sentiment analysis model through a wealth of external common sense knowledge.

### 3.4 Model Training

The extended ATAE-LSTM model can be end-to-end trained through a feedback mechanism with an objective function (loss function) as cross entropy loss. The symbol $y$ is the aspect distribution of the sentence and $\tilde{y}$ is the predicted emotional distribution. The goal of the training classifier is to minimize the sum of the predicted cross entropy losses between $y$ and $\tilde{y}$ of all sentences.

$$loss = -\Sigma_i \Sigma_j y_i^j \log \hat{y}_i^j + \lambda \| \theta \|^2 \quad \textbf{(19)}$$

Where $i$ and $j$ are the index of the input sentence and the index of the emotional polarity category, respectively. Emotions are classified into three categories: *{positive, negative, neutral}*. λ is the L2 - regularization term. θ is the parameter set. And the parameter sets for each baseline model and the extended ATAE-LSTM model are set as follows:

**Standard LSTM.** The parameter set is $\{W_i, b_1, W_f, b_f, W_o, b_o, W_c, b_c, W_s, b_s\}$ .It should be noted that word embedding is also a system parameter. Also, if aspects are embedded in the input of the LSTM unit, the dimensions of $W_i, W_f, W_o, W_c$ will increase accordingly.

**ATAE-LSTM.** The additional parameters are required: $\{A, W_h, W_v, W_p, W_x, \omega\}$, where A is the aspect of embedding and $W_h, W_v, W_p, W_x, \omega$ are the parameters of attention mechanism. Moreover, due to the concatenation of aspect embedding, the dimensions of $W_i, W_f, W_o, W_c$ are expanded compared to the parameters of standard LSTM.

**Extended ATAE-LSTM.** The parameter settings have changed due to the expansion of the ATAE-LSTM model. The $b_{oc}$ is the error generated during the training of the ATAE-LSTM model after adding external knowledge.

# 4 Experiment

## 4.1 Dataset and Resources

In our work, all word vectors are initialized by Glove1 [7]. The embedded vector is pre-trained on an unmarked corpus. The dimension of the word vector, the dimension embedded in the direction, the dimension of the hidden layer, and the dimension of the knowledge embedding are 300. The length of attention weight and the length of the sentence are the same. We use TensterFlow to build the neural network model.

The experimental data set for SemEval 2014 Task 4 removes a subset of non-target and empty targets, consisting of customer reviews, each containing a list of aspects and corresponding polarities.

## 4.2 Experiment Setting

We evaluated our approach on the target-specific task of aspect-based emotional analysis, given a set of predetermined aspects for the aspect classification, this task was to determine the polarity of each aspect, and we output the highest probability tags for each aspect.

For example, in the three-class setting, the labels are $\{positive, negative, neutral\}$. In SemEval 2014 Task 4 data set, only restaurant data has certain polarity. For example, the given a sentence *"this* restaurant location is too biased", contains one aspect of the location polarity that is negative. Our statistics is presented in Table 2 where $\{F_O., Pr., Se., Am., An.\}$ refer to $\{food, price, service, ambience, anecdotes\}$ *"Asp."* refers to aspect.

**Table 2.** Data size for each emotional polarity

| Asp. | Positive | | Negative | | Neural | |
|---|---|---|---|---|---|---|
| | Train | Test | Train | Test | Train | Test |
| Fo. | 867 | 302 | 209 | 69 | 90 | 31 |
| Pr. | 179 | 51 | 115 | 28 | 10 | 1 |
| Se. | 324 | 101 | 218 | 63 | 20 | 3 |
| Am. | 263 | 76 | 98 | 21 | 23 | 8 |
| An. | 546 | 127 | 199 | 41 | 357 | 51 |
| Total | 2179 | 657 | 839 | 222 | 500 | 94 |

## 4.4 Model Test Results

Table 3 gives the results of the performance comparison of each model, that is, the accuracy of the aspect-level emotional polarity classification in the comments on the restaurant. "Three-way" means three-class of polarity: "positive, negative, neutral". *Pos./Neg.* indicate that the predicted polarity is divided into two categories, where all neutral instances are ignored. The best score is in bold.

**Table 3.** Accuracy in classification of emotional polarity in respect to restaurant reviews

| Models | Three-way | Pos./Neg. |
|---|---|---|
| LSTM | 74.3 | - |
| ATAE-LSTM | 77.2 | 87.6 |
| Extended ATAE-LSTM | **77.8** | **87.8** |

## 4.5 Performance Comparison between Extended ATAE-LSTM and Baseline Models

**LSTM.** The standard LSTM model cannot determine for itself the key parts of a sentence for a given aspect. Therefore, when performing aspect-level sentiment analysis, the aspect information cannot be utilized, and when different aspects are given, the same emotional polarity may be obtained. So the performance of the standard LSTM model is the worst.

**ATAE-LSTM.** The ATAE-LSTM model introduces aspect embedding, so it can fully learn the knowledge and solve the problem of inconsistent direction of word embedding and aspect embedding. And because of the introduction of the attention mechanism, the model can determine the key parts of the sentence for a given aspect. Therefore, the performance of ATAE-LSTM is improved compared to the standard LSTM model.

**Extended ATAE-LSTM.** The semantic connection between word attributes and emotional polarity depends on external common sense knowledge. For example, the concept of a multi-word *"rotten fish"* may mean that *"rotten"* is a modifier associated with the next word *"fish."* Therefore, the introduction of external common sense knowledge can help filter out some irrelevant information during the learning process of the neural network. Therefore, the introduction of external common sense knowledge can help filter out some irrelevant information during the learning process of the neural network. So the ATAE-LSTM model that we extended by introducing external common knowledge knowledge embedding showed the best performance in aspect-level sentiment analysis tasks.

## 5 Conclusion and Future Work

In this paper, we propose an improved approach to the AEAT-LSTM model based on common sense knowledge embedding. The key idea of this method is that through the connection between word embedding and common sense knowledge, the neural network model learns external common sense knowledge, and common sense knowledge participates in the connection of word attributes and emotional polarity. Experiments show that our extend model obtains superior performance over the baseline models.

Although the extended ATAE-LSTM model improves aspect-level sentiment analysis performance, common-sense knowledge is only involved in the generation of hidden layer output from the LSTM model. As a future work, an interesting and possible direction combines external common sense knowledge with attention mechanisms to participate in the generation of attention weights.

## Acknowledgements

This research was funded by the National Key Research and Development Program of China, grant number 2018YFC0831300, and the Fundamental Research Funds for the Central Universities, grant number 2017JBZ107.

## References

[1] T. Nasukawa, J. Yi, Sentiment analysis: capturing favorability using natural language processing, in: International Conference on Knowledge Capture, 2003.

[2] B. Liu, Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies 5(1)(2012) 1-167.

[3] E. Cambria, D. Das, S. Bandyopadhyay, A. Fraco, A Practical Guide to Sentiment Analysis Vol. 5, Springer, 2017.

[4] S. Morinaga, K. Yamanishi, K. Tateishi, T. Fukushima, Mining product reputations on the web, in: Proc. the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2002.

[5] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, C. Dyer, Neural architectures for named entity recognition, arXiv preprint arXiv:1603.01360, 2016.

[6] D. Tang, B. Qin, X.-C. Feng, T. Liu, Target-dependent sentiment classification with long short term memory, CoRR, abs/1512.01100, 2015.

[7] Y.-Q. Wang, M. Huang, X.-Y. Zhu, L. Zhao, Attention-based lstm for aspect-level sentiment classification, in: Proc. the 2016 conference on empirical methods in natural language processing. 2016.

[8] L. Ratinov, D. Roth, Design challenges and misconceptions in named entity recognition, in: Proc. the Thirteenth Conference on Computational Natural Language Learning. Association for Computational Linguistics, 2009.

[9] B.-S. Yang, T.-M. Mitchell, Leveraging knowledge bases in lstms for improving machine reading, in: Proc. the 55th Annual Meeting of the Association for Computational Linguistics, 2017.

[10] Y. Li, Q. Pan, T. Yang, S.-H. Wang, J-I. Tang, E. Cambria, Learning word representations for sentiment analysis, Cognitive Computation 9(6 )(2017) 843-851.

[11] E. Cambria, A. Hussain, Sentic album: content-, concept-, and context-based online personal photo management system, Cognitive Computation 4(4)(2012) 477-496.

[12] Q.-F. Wang, E. Cambria, C.-I. Liu, A. Hussain, Common sense knowledge for handwritten chinese text recognition, Cognitive Computation 5(2)(2013) 234-242.

[13] T. Mikolov, I. Sutskever, K. Chen, G.-S. Corrado, J. Dean, Distributed representations of words and phrases and their compositionality, in: Proc. Advances in Neural Information Processing Systems, 2013.

[14] L. Dong, F.-R. Wei, C.-Q. Tan, D.-Y. Tang, M. Zhou, K. Xu, Adaptive recursive neural network for target-dependent twitter sentiment classification, in: Proc. the 52nd Annual Meeting of the Association for Computational Linguistics, 2014.

[15] Q. Qian, B. Tian, M.-I. Huang, Y. Liu, X. Zhu, X.-Y. Zhu, Learning tag embeddings and tag-specific composition functions in recursive neural network, in: Proc. the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015.

[16] D.-Y. Tang, B. Qin, T. Liu, Document modeling with gated recurrent neural network for sentiment classification, in: Proc. the 2015 Conference on Empirical Methods in Natural Language Processing, 2015.