Aspect-based Sentiment Analysis with New Target Representation and Dependency Attention

Tao Yang, Qing Yin, Lei Yang, and Ou Wu

Abstract—Aspect-based sentiment analysis (ABSA) is crucial for exploring user feedbacks and preferences on produces or services. Although numerous classical deep learning-based methods have been proposed in previous literature, several useful cues (e.g., contextual, lexical, and syntactic) are still not fully considered and utilized. In this study, a new approach for ABSA is proposed through the guidance of contextual, lexical, and syntactic cues. First, a novel sub-network is introduced to represent a target in a sentence in ABSA by considering the whole context. Second, lexicon embedding is applied to incorporate additional lexical cues. Third, a new attention module, namely, dependency attention, is proposed to elaborate syntactic dependency cues between words in attention inference. Experimental results on four benchmark data sets demonstrate the effectiveness of our proposed approach to aspect-based sentiment analysis.

Index Terms—ABSA, target representation, GRU, lexicon embedding, CRF, dependency attention

1 Introduction

E-COMMERCE websites contain a large amount of user reviews on products or services. Mining the user preferences or suggestions from these reviews is significant for improving the products or services. Therefore, aspect-based sentiment analysis (ABSA) has received considerable attention recently [1], [2]. ABSA aims to infer users' positive/negative attitudes on certain aspects, including aspect terms (or targets) and categories [1]. This study focuses on a target-based method considering a sentence and its contained target (aspect term); the goal is to predict the sentiment polarity about the target.

Current ABSA methods mainly rely on deep learning [3], [4], [5], [6], [7], [8], [9]. Deep learning often adopts an end-to-end framework and target embedding is a basic yet relatively important step because targets can be unseen or a phrase. Existing deep learning methods still lack an effective target representation strategy as they rarely consider the contextual cues about the target. Furthermore, lexical and semantic cues, which are proven to be relatively useful in rule-based methods [10], are not fully utilized in current deep learning-based opinion mining methods[1]. The primary lexical cues include the polarity of words. Words with negative or positive polarities play important roles in representing the sentiment polarity of texts. The primary syntactic cues involve the dependency grammar of words [3], [11]. However, to our knowledge, the grammar dependency graph of texts is not well explored in ABSA.

In this paper, new methodologies are proposed to explore contextual, lexical, and syntactic cues through further effective ways. First, a new target representation sub-network, which characterizes a target by combining K latent concepts, is proposed on the basis of the whole sentence context. Second, five types of lexical cues (i.e., polar, part of speech (POS), interrogative words, negation words, and suppositive words) are considered and used in the embedding layer. Third, syntactic dependencies among targets and other words are utilized to infer accurate attention values. Extensive experimental results verify the effectiveness of the above mentioned methodologies.

The main contributions are summarized as follows:

- Contextual cues are explicitly considered and a new target representation sub-network is used to capture the semantic and contextual information of targets further with a set of basic embeddings.
- A new dependence attention mechanism is utilized to model the syntactic dependency cues between targets and other words. The dependencies are used as regularizers in the attention inference.
- Five types of lexical cues, a total of thirteen-dimensional features, are considered to utilize the information of words further. In most previous studies, only polar cues are used.

2 Related Work

2.1 Text sentiment classification

Text sentiment classification is the most related study to ABSA. We first give a brief review for it. Text senti-

- *Ou Wu is corresponding author.*
- *The authors are with the Center for Applied Mathematics (CAM), Tianjin University, Tianjin, 300350. E-mail: yangtao087@gmail.com and {qingyin, yl7268, wuou}@tju.edu.cn*
- *This work was done while Tao Yang was at CAM as a research intern. Tao Yang proposed the target representation part and implemented the whole approach. Qing Yin performed the experimental analysis. Lei Yang participated in the discussion and gave suggestions. Ou Wu wrote the initial paper and proposed dependency attention and lexicon embedding parts. All resources are available at https://github.com/absa-nlp/CueNet.*

Manuscript received April 17, 2019; revised September 19, 2019.

1. Some methods are introduced in the related work section.

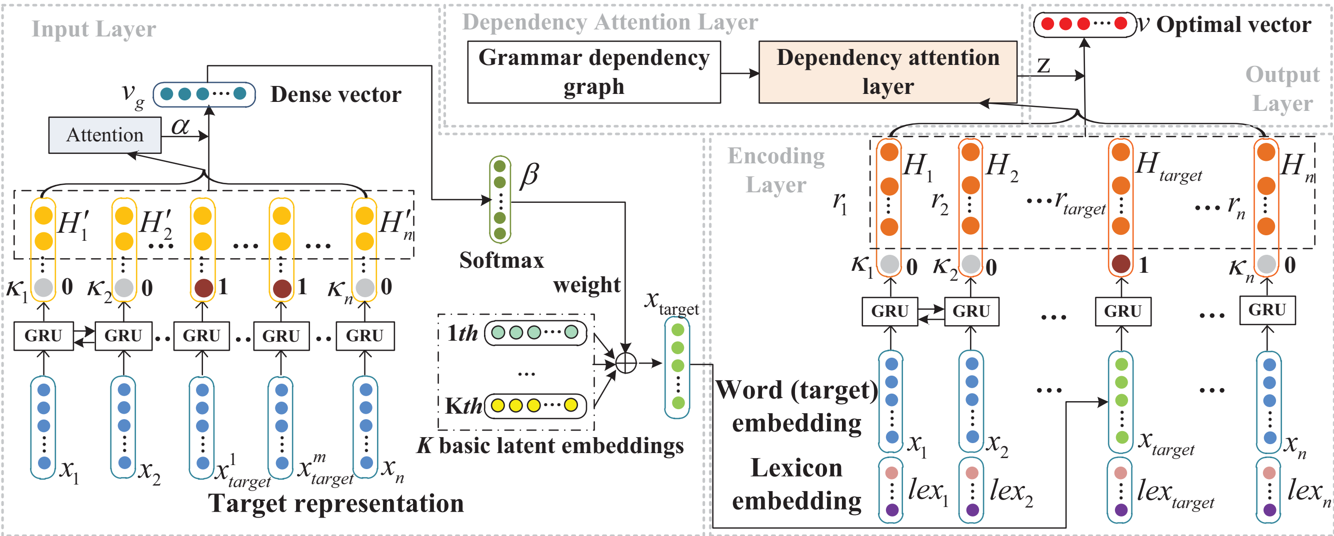


Fig. 1. Overview of the proposed model with three main components: target representation, lexicon embedding, and dependency attention.

ment classification aims to predict the sentiment polarity of an input text sample (without given aspects). Three-level sentiment polarities contain "positive", "neutral", and "negative" in numerous applications. Existing sentiment classification methods can be divided into two categories, namely, lexicon and machine learning-based methods [12], [13]. Lexicon-based methods [2], [14] construct polar and negation word dictionaries. A set of rules for polar and negation words is compiled to assess the sentiment orientation of a text document. However, this method cannot effectively predict implicit orientations. Machine learning-based methods [15] utilize a standard binary or multi-category classification approach. Different feature extraction algorithms, including bag-of-words (BOW) [15] and POS [16] are used. Deep neural networks have recently been applied to sentiment analysis [17]. There are two popular deep neural networks: Convolutional Neural Network (CNN) [18] and Recurrent Neural Network (RNN) [19], [20], which are used in sentiment analysis [21]. LSTM [22] is the most popular RNN network utilized for sentiment analysis. A bidirectional LSTM [23] with an attention mechanism is demonstrated to be effective in sentiment classification.

Deep learning-based methods rarely utilize lexical cues. Shin et al. [24] combined the lexicon embeddings of polar words with word embeddings for sentiment classification. A recent survey can be referred to Zhang et al. [25].

2.2 Aspect-based Sentiment Analysis

ABSA refers to three key issues: target extraction [26], [27], sentiment analysis (the targets are provided), and joint aspect extraction and sentiment analysis [28]. This study focuses on the second issue, which also consists of two main technical lines, namely, rule-based [10] and machine learning-based [29]. As deep learning has emerged as a powerful technique for almost all natural language processing (NLP) tasks, classical deep neural networks (e.g., Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN)) are adopted for this fine-grained task.

Wang et al. [6] added the targets in the input embedding and attention layers. Tang et al. [5] applied two long short-term memory networks (LSTM) to capture the left and right contextual information of a target. Chen et al. [7] utilized a memory module to synthesize the word sequence features. A multiple-attention mechanism is then run on the memory to identify important information to predict the final sentiment. Wang and Lu [9] proposed a segmentation attention-based LSTM model, which can effectively capture the structural dependencies between the target and sentiment expressions with a linear-chain conditional random field (CRF) layer [30]. Wang et al. [31] adopted a hierarchical aspect-specific attention network which firstly segments a sentence into several clauses and then utilizes a word-level and a clause-level attention layer to capture the importance degrees of all words and clauses, respectively.

As previously stated, although numerous achievements have been made, target representation and several lexical and semantic cues are not well explored. Thus, they are the focus of this study.

3 Methodology

ABSA can be formulated as follows. By considering a target–sentence pair $\{g, s\}$, where g is a sub-sequence of s, we aim to predict the sentiment polarity of the given target g in s. Let $\{x^1_{target}, x^2_{target}, ..., x^m_{target}\}$ and $\{x_1, x_2, ..., x_n\}$ be the corresponding word embeddings of g and s, respectively.

3.1 Whole Model

Figure 1 shows the whole architecture of our model. The major layers include input, encoding, dependency attention, and output.

In the input layer, an independent sub-network is used to pursue target representation, which output is x_{target}. It will be detailedly introduced in following subsections.

In the encoding layer, five types of lexical cues are embedded in accordance with one-hot encoding. In addition, a bidirectional gated recurrent unit (bi-GRU) [32] is used to infer the hidden representation of each word in a sentence due to the small amount of data and simplification of parameters. In our model, the input and output of bi-GRU at time t or position t are calculated as follows:

$$GRU_{in} = [x_t, lex_t] \tag{1}$$

$$GRU_{out} = H_t = [\vec{h}_t, \overleftarrow{h}_t] \tag{2}$$

where x_t and lex_t are the embeddings of word (target) and lexicon, respectively; $\vec{h}_t$ and $\overleftarrow{h}_t$ are the hidden vectors from the forward and backward GRUs. The backward GRU is similar to the forward one except that the input sequence is fed in a reversed way.

The output of the bi-GRU and a binary indicator vector (κ_t) [9], [33] are concentrated as follows:

$$r_t = [H_t, \kappa_t] \tag{3}$$

where κ_t is utilized to indicate whether the word is (or contained in) a target or not, and κ_t is selected from two learnable vectors. Compared with fixed vectors, the performance is much better when κ_t adopts learnable vectors according to our experiments. The concentrated vector r_t is then used to infer the t-th attention score and produce the final optimal vector.

In the dependency attention layer, the proposed new attention mechanism is used. Our attention mechanism utilizes the dependency cues between the target and other words in the involved sentence. The grammar dependency graph is used to analyze the accurate dependency relationships between targets and other words in a sentence compared with the structural dependency utilized by Wang and Lu [9].

In the output layer, the outputs of the encoding layer are summed with the associated attention scores to produce the final optimal feature vector:

$$v = \sum_t p(z_t = 1) \cdot H_t \tag{4}$$

where $p(z_t = 1)$ is the dependency attention score of the t-th word. Then, a softmax function is used to predict the final category:

$$y = softmax(W_v^T v + b_v) \tag{5}$$

The three main components (target representation, lexicon embedding, and dependency attention) are introduced as follows.

3.2 Target Representation

Target representation or embedding is a basic step in ABSA. Two classical target representation strategies are presented in previous studies. The first strategy uses the average embeddings of the target words as the target representation [6], [34]. This strategy may be inappropriate in some cases because different target words often do not contribute equally [35]. The second strategy uses a LSTM network with attention on the (sequential) target words and the weighted pooling of the hidden vectors is selected as the target representation [8], [36]. Although this strategy seems more effective than the first one, it still possesses two disadvantages. First, in real applications, targets often contain no more than two words[2]. For example, the targets can be a company name (e.g., Google and Apple), products (e.g., TV, car, and book), or places (e.g., Beijing and California). In these cases, the LSTM, which is suitable for sequential data may be inappropriate. Second, the target representation depends on not only its contained words but also the context. For example, when "apple" is the target, its representation is subject to the context (mobile phone or fruit).

Our initial analysis on practical data suggests that a target can be further represented by a set of basic attribute-level concepts. For example, if the opinion concerns a person, then the words in concrete names (e.g., "Tom" and "Luce") are considerably less important than the underlying concepts, such as "age group" and "gender"; if the opinion relates to a product, then the words in concrete names (e.g., "iPhone" and "Benz") are still considerably less important than the underlying concepts, such as "industry category".

To this end, considering that the application of opinion mining often focuses on a specific domain, we assume that target representation in a specific domain can be one of, or the weighted combination of, the K latent embedding vectors. The K latent embedding vectors can also be seen as K basic attribute-level concepts.

The left part of Figure 1 shows the sub-network of target representation. A bi-GRU is also utilized on the whole sentence. In the attention layer, the binary indicator vector is reused to indicate the target word. The dense feature vector v_g is obtained on the basis of the following calculation:

$$\alpha_t = softmax(W_\alpha^T \begin{bmatrix} H'_t \\ \kappa_t \end{bmatrix} + b_\alpha) \tag{6}$$

$$v_g = \sum_t \alpha_t \cdot H'_t \tag{7}$$

where H'_t is the outputs of the bi-GRU (without lexical embedding) used in the right sub-network. The prediction for the target category (in $[1, 2, ..., K]$) is as follows:

$$\beta = \mathrm{softmax}(W_\beta^T v_\mathrm{g} + b_\beta) \tag{8}$$

Let $E = [e_1, ..., e_K]$ be the K latent embeddings (vectors), and the dimension of e is consistent with word embedding. In this study, the representation of a concrete target is the weighted combination the K basic latent concepts:

$$x_{target} = E\beta^T \tag{9}$$

E is learned in the experiments when K is fixed. We have used a relatively simple network to capture potential concepts, and further efficient networks will be designed in future works.

3.3 Lexicon Embedding

Most lexicon-based sentiment analysis methods rely on four types of words, namely, positive, negative, neutral and negation. These words are useful cues for predicting the sentiment labels of input texts. The incorporation of these words into our neural network model shall also be useful. The sentiment expressed in a conditional or interrogative

2. In the four experimental data sets, the proportions for targets with one word are 61.42%, 74.74%, 29.99%, and 54.72%, respectively.

sentence is often difficult to assess due to the semantic or uncertain condition. In addition, POS is commonly used as a key cue in sentiment analysis [37].

TABLE 1
Example Words of Lexicon Cues.

Positive	Neutral	Negative	Negation	Suppositive	Interrogative
amazing	air	betrayal	never	unless	what
excellent	claim	repressed	rarely	if	how
wonderful	water	turbulent	neither	assume	when
elegance	house	resentment	barely	supposing	where

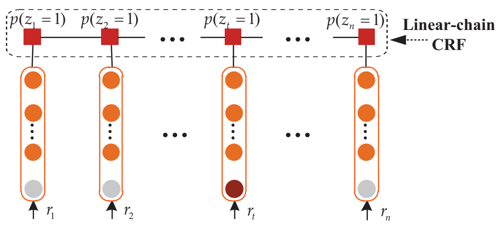
These lexical cues are not well considered in existing deep learning-based ABSA. In this study, we have sorted out some lexical features. Words are divided into six categories: “positive,” “neutral”, “negative”, “negation”, “suppositive”, and “interrogative”. The last two types of words are regarded as markers of conditional and interrogative sentences. These lexicon cues are obtained from Wu et al.[3] [38]. Some example words are shown in Table 1. These types of information are characterized by a six-dimensional one-hot vector.

Based on our initial case studies, the seven most common POS types are considered: “noun”, “adjective”, “verb”, “pronoun”, “adverb”, “preposition”, and “accessory”. This type of information is represented by a seven-dimensional one-hot vector.

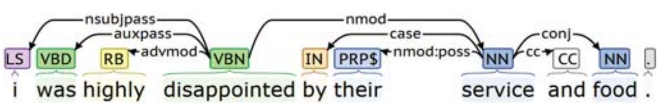
These two vectors are concentrated, and finally, a thirteen-dimensional vector is used to represent the lexical cues.

The model in Figure 1 shows that the lexicon embedding is used as a part of the input for the bi-GRU in the network.

3.4 Dependency Attention



(a) Linear-chain CRF-based attention module.



(b) Grammar dependency graph of a sentence (“service” or “food” is the target).

Fig. 2. CRF-based attention and grammar dependency graph.

Attention is important for deep neural network. Our proposed dependency attention is based the segmentation attention proposed by Wang and Lu [9]. The underlying motivation of segmentation attention is that the opinion expression associated with a target may be in the form of a chunk or linear span structure. Therefore, if one word is important, its adjacent words may also be important. The linear-chain CRF is used to specify the segmentation attention scores for each involved word.

Let a binary latent variable z in $\{0,1\}$ be the state of a word. Figure 2(a) shows the linear-chain CRF model of the attention scores. On the basis of the linear-chain CRF theory, the distribution of a possible sequence of $\mathbf{z}=\{z_1,z_2,...,z_n\}$ can be defined as follows:

$$p(\mathbf{z}|R)=\frac{1}{Z(R)}\prod_{c\in C}\psi(\mathbf{z}_c|R) \tag{10}$$

where $\mathbf{z}$ is formed by $\mathbf{z}_c$; $R=\{r_1,r_2,...,r_n\}$ refers to concentrated vectors defined in Eq. (3) for the sequential words. $Z(R)$ is a normalization function, and $\psi(\mathbf{z}_c|R)$ is the potential function of the clique c:

$$\prod_{c\in C}\psi(\mathbf{z}_c|R)=\prod_{i=1}^{n}\phi_1(z_i|R)\prod_{i=1}^{n-1}\phi_2(z_i,z_{i+1}|R) \tag{11}$$

where ϕ_1 and ϕ_2 are defined as follows:

$$\phi_1(z_i|R)=\exp(W^{crf}_{z_i}\cdot r_i+b_{crf}) \tag{12}$$

$$\phi_2(z_i,z_{i+1}|R)=\exp(W^{e}_{z_i z_{i+1}}) \tag{13}$$

where W^{crf} and b_{crf} are learnable parameters that aim to map each concentrated vector r to the attention score and W^e is a transition matrix defined for each pair of latent variable. Eqs. (11)-(13) indicates that the attention score of the word depends on itself and the scores of its adjacent words.

Two regularizers proposed by Wang and Lu [9] are used to implement the segmentation attention, and denoted as follows:

$$\Omega_1=\sum_i\sum_{j\neq i}\max(0,W^e_{ij}-W^e_{ii}) \tag{14}$$

$$\Omega_2=\sum_{i=1}^{n}p(z_i=1) \tag{15}$$

where $p\,(z_i=1)$ is the attention score of the i-th word. The first regularizer Ω_1 is utilized to encourage the state of z_{i+1} to remain the same as z_i. The second regularizer Ω_2 is used to enforce the attention weight of a meaningful word since only a few opinion words have an impact on the target. By introducing these regularizers, their model can concentrate on short, yet meaningful, opinion spans.

The segmentation attention utilizes the structure relationships between adjacent words in a sentence. However, syntactic dependency can guide the algorithm to capture the target-related opinions directly. Thus the dependencies between words can be further investigated and utilized. Indeed, dependency analysis [39] between words is an active research area in NLP. A grammar dependency graph is a directed diagram for a sentence. The graph vertex is a word in the sentence and the edge represents a grammar relationship between words. Figure 2(b) shows an example, the word *“disappointed”* is the adjacent node of the targets *“service”* and *“food”*, although the former is not adjacent to the latter with regard to position.

It is quite challenging for existing models to learn complex syntactic dependence by themselves due to the

3. https://github.com/Tju-AI/two-stage-labeling-for-the-sentiment-orientations

small amount of training data in ABSA. Therefore, we impose syntactic dependency cues as explicit regularizers to supervise model learning. An additional advantage of using regularizers is that no additional syntax parsing is required during the inference stage. In this work, the word relationships from the grammar dependency graph are used as an additional regularizer in the model. Specifically, the distance $d_{i,t}$ between i-th word and the target (t-th) word on the grammar dependency graph is calculated. Then, a more reasonable regularizer than Ω_1 is defined as follows:

$$\Omega_{gd} = \sum_{i}^{n} [1 - p(z_i = 1)](d_{\max} - d_{i,t})/d_{\max} \tag{16}$$

where $d_{\max}$ is the maximum distance on a graph. Eq. (16) penalties the word with low attention scores yet small distance from the target.

In addition, some words are close to the target with less contributions to sentimental polarities. Words, such as "*disappointed*", "*by*", "*their*", "*and*", and "*food*", are close to the target "*service*" in the dependency graph; however, "*disappointed*" is the only word that can express the exact sentiment. Then, Ω_2 remains valid for our dependency attention:

$$\Omega_{bias} = \Omega_2 = \sum_{i}^{n} p(z_i = 1) \tag{17}$$

3.5 Objective Function

The objective function of the proposed model presented in Figure 1 is denoted as follows:

$$L = \frac{1}{N}[\sum_{i=1}^{N} -y_i \log p(y_i) + \lambda_1 \Omega_{gd} + \lambda_2 \Omega_{bias}] \tag{18}$$

where N is the number of training samples, y_i indicates the true label, $p(y_i)$ represents the predicted output by using the softmax function of the model, and Ω_{gd} and Ω_{bias} signify the regularizers of Eqs. (16) and (17), respectively. λ_1 and λ_2 refer to the weights of related regularizers, which are searched in the experiment.

4 Experiments

4.1 Data Sets

Data from SemEval are used in most previous studies to evaluate the proposed methods [6], [7], [9], [34], [40]. Four data sets adopted in most recent studies are utilized for a fair comparison. Thus, the results of several state-of-the-art methods proposed in these studies can be directly used in accordance with their publications. The first two data sets are from SemEval-2014 Task 4 [41]. These data sets consist of user review data from two domains, namely, laptop and restaurant. The preprocessing in previous works [7], [9], [34] was followed to discard the sentences that contain "conflict" labels. The third data set was collected from twitter comments by Dong et al. [3]. The last and latest data set is English comments about restaurant, which comes from SemEval-2016 Task 5 [42]. Following previous work [43], we remove a total of 127 conflict samples if a target is associated with different sentiment polarities. In this work, accuracy is mainly utilized for evaluation. Table 2 presents the details of the four data sets.

TABLE 2
Details of the Experimental Data Sets.

Data	Laptop	Rest14	Twitter	Rest16
Train	2313	3602	6257	2417
Test	638	1120	694	825

4.2 Comparative Methods

In order to comprehensively evaluate the performance of our proposed model, we adopted some baseline approaches for comparison, which are introduced as follows.

- **SVM** [40]: This method adds lexicon, surface (e.g., unigrams (single words) and bigrams (two-word sequences)), and parsing features into the classical SVM model.
- **AdaRNN** [3]: This technique adaptively propagates the sentiments of words to the target depending on the context and syntactic relationships between words and the target derived from the dependency tree.
- **AT-LSTM** [6]: In the network, the representation of the involved target is added in the input and attention layers.
- **MemNet** [34]: This process utilizes multiple neural attention layers over an external memory to capture the importance of each context word explicitly when inferring the sentiment polarity of an aspect.
- **RAM** [7]: This method proposes a new recurrent layer that adopts the multiple-attention mechanism to capture sentiment features separated by a long distance.
- **IAN** [8]: This method utilizes two LSTMs and proposes an interactive attention mechanism to learn the relationships of the targets and the contexts.
- **A-LSTM** [9]: This method is similar to AT-LSTM. The only difference is that A-LSTM use a binary-based target representation.
- **SA-LSTM** and **SA-LSTM-P** [9]: The SA-LSTM method applies the segmentation attention layer on top of the LSTM. When the penalty terms on long spans are added, SA-LSTM becomes SA-LSTM-P.

We also compare our method with some of the latest approaches published in 2018 and 2019, which are listed as follows.

- **Cabasc** [44]: This method adopts two attention enhancing mechanisms, namely, sentence-level content attention mechanism and context attention mechanism to deal with multi-aspect sentences and the syntactically complex sentence structures.
- **TransCap** [45]: This method transfers sentence-level semantic knowledge from document-level sentiment classification to aspect-based sentiment analysis via capsule network.
- **MGAN** [46]: In this framework, Coarse2Fine attention can help the coarse-grained aspect category task modeling at the same fine-grained level with aspect term task.
- **TNet** [47]: This architecture introduces multiple context-preserving transformation (CPT) layers to

dynamically compute the importance of target words based on each sentence word rather than the whole sentence.
- **TNet-ATT (+AS)** [48]: This method utilizes a progressive self-supervised attention learning approach that automatically mines useful attention supervision information from a training corpus to refine attention mechanisms in TNet-ATT.
- **BERT** [49]: This model leverages the vanilla BERT pre-trained weights and fine-tunes on different data sets.
- **BERT-PT** [50]: This method post-trains BERT's weights on laptop or restaurant domain data set using the joint post-training algorithm and fine-tunes on ***Laptop*** or ***Rest14*** data set.

Our proposed method is consists of several new modules. To investigate whether the two major components: target representation and dependency attention also work for other models, we added these two components in some typical models and compared their performance. In addition, we also utilize the BERT pre-trained model in our method to make a fair comparison with some BERT-based methods [49], [50]. The variants of our method are listed as follows:

- **AT-LSTM+TR**, **IAN+TR**, and **SA-LSTM+TR**: We replaced the target embedding with our target representation module in AT-LSTM, IAN, and SA-LSTM.
- **SA-LSTM+DA**: SA-LSTM with our dependency attention mechanism.
- **CueNet**: Our cue-guided method adds target representation, the CRF-based attention structure into the conventional bi-GRU. In CueNet, the lexicon embedding, Ω_{gd}, and Ω_{bias} are not incorporated.
- **CueNet-L**: The lexicon embedding is utilized in CueNet.
- **CueNet-DL**: Two regularizers, namely, Ω_{gd}, and Ω_{bias} are added compared with the CueNet-L model to realize the dependency attention. Therefore, this method includes all the proposed modules in this study.
- **CueNet-DL (BERT)**: BERT pre-trained model is used. The output of the last layer of BERT is used to replace the word embeddings of the CueNet-DL model.
- **CueNet-DL (BERT-PT)**: This method is similar to CueNet-DL (BERT) except for the pre-trained model involved. In this method, we utilize BERT-PT as the pre-trained model.

4.3 Training Settings

The training settings used by Wang and Lu [9] are followed, specifically:

- GloVe [51] is used to generate a 300-dimensional word embedding.
- One-sixth of the training data are retained as the validation sets.
- 30-dimensional embedding for target's binary indicator.

The model is trained by applying Keras[4], which is backed with Tensorflow[5]. We utilize the Stanford Parser[6] [52] to calculate the dependency distance and NLTK[7] to tag POS.

To facilitate experiment repeating, some hyper-parameters are listed in Table 3, which are the best setting in our search ranges.

TABLE 3
Best Setting for Hyper-parameters.

Hyper-params	Laptop	Rest14	Twitter	Rest16
hidden units	150	100	100	100
$\boldsymbol{K}$	5	10	15	11
$\boldsymbol{\lambda_1}$	1.2	0.9	1.5	0.3
$\boldsymbol{\lambda_2}$	0.0001	0.001	0.001	0.001
batch size	16	16	25	16
dropout	0.5	0.5	0.3	0.5

4.4 Overall Competing Results

Table 4 presents the main results (classification accuracies) that our models compared with 8 baseline methods on four benchmark data sets. The proposed model, namely, CueNet-DL, achieves the highest accuracies on all data sets. Compared with the state-of-the-art model, namely, SA-LSTM-P, the results increased by 2.6%, 0.3%, 5.5%, and 1.1% respectively on four data sets. The simplified model, namely, CueNet, which only contains the new target representation and the CRF-based attention structure, also obtains better results than SA-LSTM-P on two data sets. The CueNet-L model has added the lexicon features compared with CueNet. Therefore, the former is better than the latter. CueNet-L is inferior to CueNet-DL, indicating that dependency attention module benefits the whole model.

TABLE 4
Results on Four Benchmark Data Sets Regarding Accuracy (%). Best Scores Are in Bold.

Method	Laptop	Rest14	Twitter	Rest16
SVM	70.5*	80.2*	63.4*	-
AdaRNN	-	-	66.3*	-
AT-LSTM	68.9*	77.2*	-	83.8*
MemNet	70.3*	78.2*	68.5*	83.1*
RAM	74.5*	80.2*	69.4*	83.9*
IAN	72.1*	78.6*	69.1	-
SA-LSTM	74.5*	79.8*	69.9*	87.2
SA-LSTM-P	75.1*	81.6*	69.0*	87.5
CueNet	76.6	81.1	73.9	87.5
CueNet-L	76.8	81.6	74.1	87.7
CueNet-DL	**77.7**	**81.9**	**74.5**	**88.6**

* This symbol means that the results are retrieved from Wang et al. [9], Ma et al. [8], and He et al. [43].

The comparison in Table 4 shows that the proposed models outperform state-of-the-art methods in previous studies. The introduced lexical features also improve the performance.

4. https://github.com/keras-team/keras/
5. https://www.tensorflow.org
6. https://nlp.stanford.edu/
7. http://www.nltk.org/

TABLE 5
Comparison of Classical Methods With or Without Our Target Representation (Termed 'TR') Strategy.

Method	Laptop	Rest14	Twitter	Rest16
AT-LSTM	68.9*	77.2*	-	83.8*
IAN	72.1*	78.6*	69.1	-
SA-LSTM	74.5*	79.8*	69.9*	87.2
A-LSTM	72.7*	78.4*	68.2*	85.9
AT-LSTM+TR	74.9	80.0	72.5	86.7
IAN+TR	73.6	79.0	72.7	-
SA-LSTM+TR	**75.1**	**80.4**	**74.0**	**87.5**

TABLE 6
Comparison of Different Attention Mechanisms.

Method	Laptop	Rest14	Twitter	Rest16
AT-LSTM	68.9*	77.2*	-	83.8*
MemNet	70.3*	78.2*	68.5*	83.1*
RAM	74.5*	80.2*	69.4*	83.9*
IAN	72.1*	78.6*	69.1	-
SA-LSTM	74.5*	79.8*	69.9*	87.2
SA-LSTM-P	75.1*	81.6*	69.0*	87.5
SA-LSTM+DA	**75.4**	**81.7**	**73.6**	**88.1**

Table 5 shows the performance comparison of the model which replaced target embedding with our target representation module in original model. We compared different target presentation methods. Among them, AT-LSTM utilizes the target representation method of averaging word vectors; IAN use an LSTM network to capture target words and the weighted pooling of the hidden vectors is selected as the target representation; SA-LSTM and A-LSTM indicates the location of target words by a learnable binary vector. From the second group in Table 5, we can observe that, all the replaced models perform better than the original ones on all datasets, which verifies the efficacy of our target representation module.

Attention mechanism is crucial for calculating the correlation between targets and contexts. As Table 6 shows, we also compared five different attention mechanisms: AT-LSTM adopts a standard attention layer to weight the hidden vectors; MemNet and RAM utilizes memory network which is consist of multiple stacks of standard attention layers; IAN proposes an interactive attention mechanism to interactively learn attentions in the contexts and targets; SA-LSTM and SA-LSTM-P applies a CRF-based segmentation attention which focused on the dependencies of adjacent structures; SA-LSTM+DA adopts our dependency attention in SA-LSTM by using grammar dependency graph as an additional supervisor.

TABLE 7
Compared with the Latest Approaches on the Three Most Commonly Reported Data Sets.

Methods	Laptop		Rest14		Twitter	
	Acc.	F1	Acc.	F1	Acc.	F1
Cabasc	75.07†	-	80.89†	-	71.53†	-
TransCap	73.87†	70.1†	79.55†	71.41†	-	-
MGAN	76.21†	71.42†	81.49†	71.48†	74.62†	73.53†
TNet	76.54†	71.75†	80.69†	71.27†	77.60†	76.82†
TNet-ATT (+AS)	77.62†	73.84†	81.53†	72.90†	78.61†	77.72†
CueNet-DL	77.77	74.11	81.90	72.98	74.57 78.18‡	72.96 77.42‡
BERT	75.29†	71.91†	81.54†	71.94†	78.47‡	77.76‡
CueNet-DL (BERT)	77.62	73.87	82.86	73.56	**78.90**‡	**78.18**‡
BERT-PT	78.07†	75.08†	84.95†	76.96†	-	-
CueNet-DL (BERT-PT)	**79.50**	**76.43**	**85.63**	**77.68**	-	-

† This symbol means that the results are cited from the papers of compared methods.
‡ We noticed that TNet-ATT (+AS) and TNet have a different preprocessing method compared with ours on ***Twitter*** data set. Therefore, CueNet-DL, BERT, and CueNet-DL (BERT) are following their preprocessing method to make a fair comparison.

As shown in Table 6, SA-LSTM+DA achieves the best performance on all data sets compared to other four attention mechanisms, which verifies the efficacy of our dependency attention mechanism. Compared with AT-LSTM, MemNet and RAM, which both based on standard attention mechanism, the CRF-based models get better performance. We noticed that SA-LSTM+DA outperforms SA-LSTM and SA-LSTM-P on ***Twitter*** by a substantial margin, obtaining 4.15% average accuracy improvement. The crucial reason is that ***Twitter*** data is relatively confused and loosely structured, which makes segmentation attention ineffective. Dependent attention which focused on grammatical relevance between words and targets is beneficial for this task.

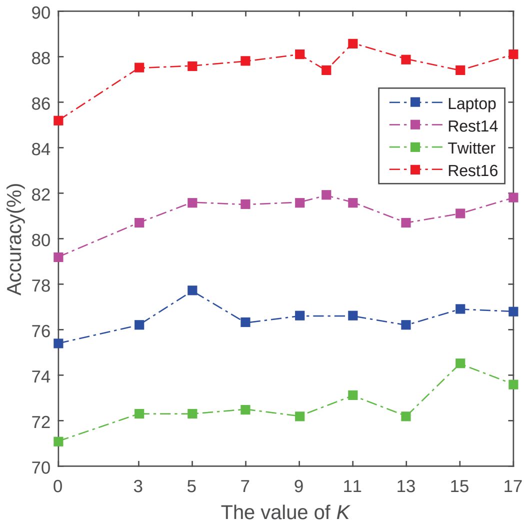
In addition, we compare our proposed CueNet-DL with the seven latest approaches which are published in 2018 and 2019 on the three most commonly reported benchmark data sets, namely, ***Laptop***, ***Rest14***, and ***Twitter***. Table 7 presents the comparison results. The accuracy (Acc.) and Marco-F1 (F1) are reported in this comparison. As the first group shows, CueNet-DL obtains the highest classification accuracies and Marco-F1 scores on ***Laptop*** and ***Rest14*** data sets compared to the latest methods with GloVe as pre-trained word vectors. The second group presents a comparison between BERT-based models. We observed that BERT and BERT-PT (BERT post-trained on domain knowledge) can effectively enhance the performance of classification. Our proposed approaches combined with BERT or BERT-PT can further improve the performance of ABSA.

Our proposed models consist of three main modules and include additional parameters. The successive experiments investigate how the parameters affect the final performances.

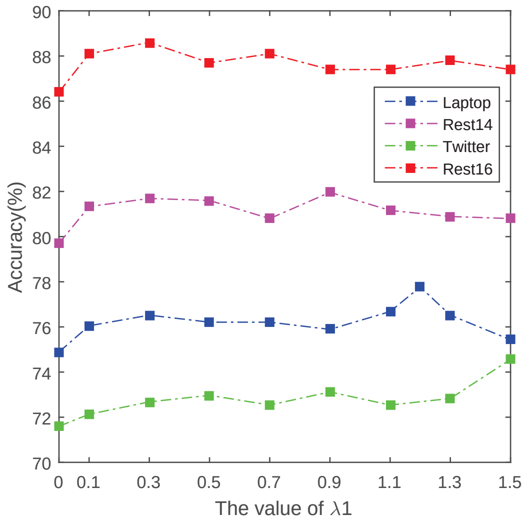
4.5 Results on Different Parameters

In this section, we investigate the effects of three main parameters, namely, K, λ_1, and λ_2, which are denoted in the methodology section. All experiments are conducted on the basis of the four data sets. Figure 3 shows all the results.

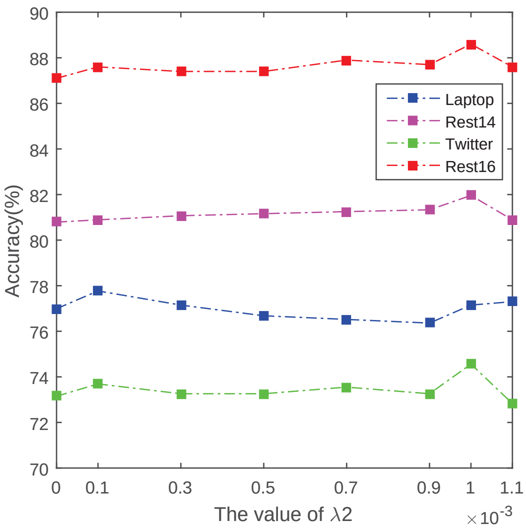
Figure 3(a) illustrates that the parameter K in the target representation is set from 0 to 17. On the basis of our experiment, data sets with large amounts of data are advised to use large K values. Figure 3(b) presents the result of parameter λ_1, which is selected from 0 to 1.5, for which the relatively high dependency regularizer can achieve improved results. Figure 3(c) displays the influence of λ_2. This hyper-parameter is valued from 0 to 0.0011, because the magnitude of losses caused by λ_2 is relatively large. For ***Rest14***, ***Twitter*** and ***Rest16***, our model can achieve the best



(a) Accuracy with different K



(b) Accuracy with different λ_1

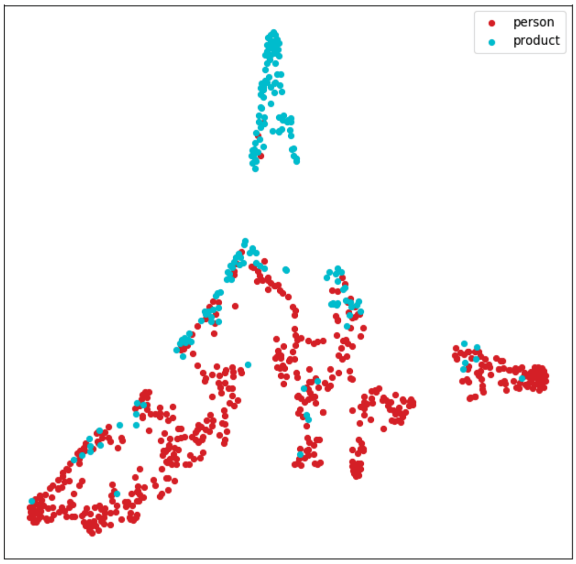


(c) Accuracy with different λ_2

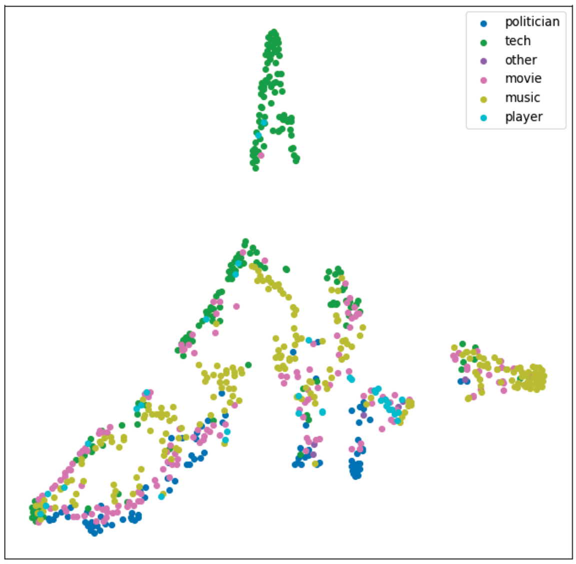
Fig. 3. Effects of different parameters in search ranges.

accuracy when $\lambda_2 = 0.001$, but for ***Laptop***, 0.0001 is the best.

Although our model has an additional parameter compared with the state-of-the-art model SA-LSTM-P, in our search ranges for the parameters, its performance under an arbitrary parameter value is comparable with that of SA-LSTM-P.



(a) Visulization of target representation vectors divided into two attributes.

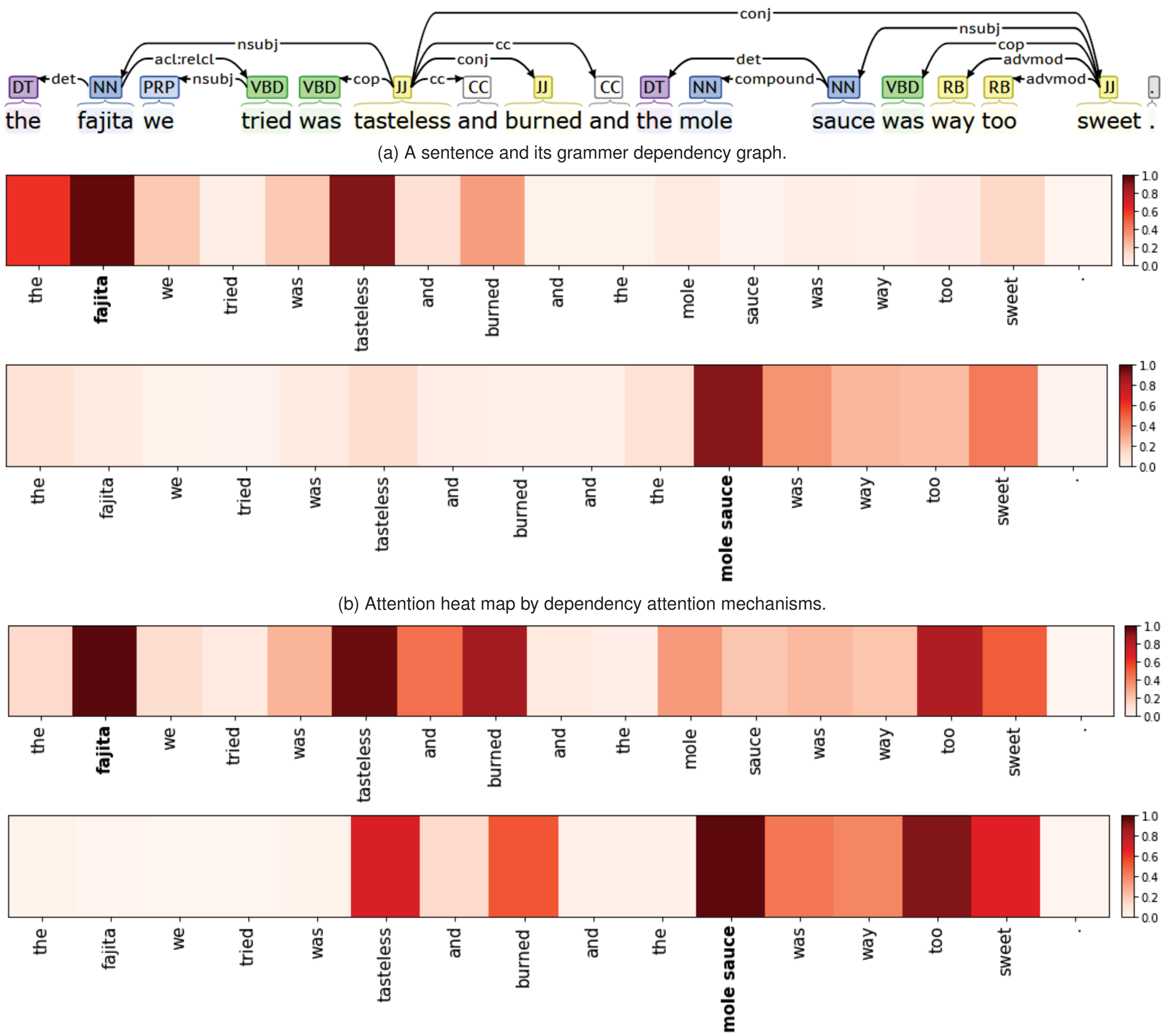


(b) Visulization of target representation vectors divided into six attributes.

Fig. 4. The visualization results of target representation vectors under different attributes partitions.

4.6 Case Studies

The previous experiments verify the effectiveness of the proposed modules, including target representation, lexicon embedding, and dependency attention with regard to the



(a) A sentence and its grammer dependency graph.

(b) Attention heat map by dependency attention mechanisms.

(c) Attention heat map by segmentation attention mechanisms.

Fig. 5. Weight visualizations of two attention mechanisms.

final classification accuracies. This subsection further illustrates the effectiveness of the proposed modules concerning case studies.

To verify whether our proposed target representation can effectively represent different types of target, we visualize the results of target representation. The specific visualization process is as follows. Firstly, we annotate 694 samples of ***Twitter*** test data according to the attributes of the target. Secondly, the target representation vectors (x_{target}) of each sample are taken out, and the t-SNE [53] is utilized to reduce the dimension of each target representation vector to two dimensions. Finally, the two-dimensional visualization results are shown in Figure 4.

As shown in Figure 4(a), we use only two attributes, person and product, to distinguish different targets considering that targets in ***Twitter*** data sets can clearly be categorized into these two categories. We can see that the distributions of these two attributes are obviously distinct. Further more, targets are categorized in more detail by their background, and each target can be assigned to one of the six categories: politician, tech, movie, music, player, and other. From Figure 4(b), we can observe that tech and politician all have relatively centralized distributions. Nevertheless, the distributions of movie, music and player are mixed and dispersed. The reason is that many persons mentioned in the targets have multiple backgrounds of singer, actor, or player. Therefore, targets in these three categories are highly related to each other. In summary, our proposed target representation module can effectively represent targets by utilizing potential concepts.

Figure 5(a) presents a sentence (“*The fajita we tried was tasteless and burned and the mole sauce was way too sweet.*”) and its grammar dependency graph. The top image in Figure 5(b) shows the dependency attention heat map for each word obtained by the proposed approach when the target is "*fajita*", and the bottom image illustrates the dependency attention heat map when the target is "*mole sauce*". We observed that the dependency attention is relatively consistent with the grammar dependency graph. The distance between the target "*fajita*" and the word "*burned*" is two based on the dependency graph, whereas it is six based on the structural distance. Actually, "*burned*" is the sentimental evaluation for the target. The distance between the target "*mole sauce*" and the word "*sweet*" is one based on the dependency graph, whereas it is four based on the structural distance. The word "*sweet*" is also the exact description for the target. Figure 5(c) displays the two structure attention [9] heat maps for the two targets. Our dependency attention can further capture the semantic relationships between targets and other words.

5 Conclusion

This study investigates several crucial issues that have not been well utilized in previous aspect-based sentiment analyses. The first issue refers to the contextual information in target (a single word or phrase) representation in a given sentence. The second issue involves the utilization of more useful lexical cues related to sentiment expressions in texts. The third issue concerns the use of word relationship cues in improving attention inference. In our work, a new strategy is used to learn K basic latent embeddings, and each target is represented by the weighted combination of the K latent embeddings via a sub-network on the whole sentence. Lexicon embeddings are used to add additional lexical cues. Further, a new attention mechanism, namely, dependency attention, is proposed in this study to explore more word relationships than those in existing studies. Experimental results on four benchmark data sets verify the effectiveness of the proposed methodologies.

In our future work, we will focus on the following aspects. (1) We will consider different types of syntactic dependencies which have been proven to improve the performance of NLP tasks in many studies [54], [55]. (2) The proposed methodologies will be considered to solve related studies, including sentiment analysis, and text classification. (3) Existing ABSA data sets are small. Thus, we aim to compile large-sized benchmark data sets for future studies.

6 Acknowledgements

This work is supported by the National Natural Science Foundation of China (61673377), the AI Key Project of Tianjin (17ZXRGGX00150).

References

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

[2] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004, pp. 168–177.

[3] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive recursive neural network for target-dependent twitter sentiment classification," in *Proc. ACL (Volume 2: Short Papers)*, vol. 2, 2014, pp. 49–54.

[4] M. Yang, W. Yin, Q. Qu, W. Tu, Y. Shen, and X. Chen, "Neural attentive network for cross-domain aspect-level sentiment classification," *IEEE Transactions on Affective Computing*, 2019.

[5] D. Tang, B. Qin, X. Feng, and T. Liu, "Target-dependent sentiment classification with long short term memory. arxiv preprint," *arXiv preprint arXiv:1512.01100*, 2015.

[6] Y. Wang, M. Huang, L. Zhao *et al.*, "Attention-based lstm for aspect-level sentiment classification," in *Proc. EMNLP*, 2016, pp. 606–615.

[7] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proc. EMNLP*, 2017, pp. 452–461.

[8] D. Ma, S. Li, X. Zhang, H. Wang, D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," in *Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017, pp. 4068–4074.

[9] B. Wang and W. Lu, "Learning latent opinions for aspect-level sentiment classification," 2018.

[10] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in *Proceedings of the 2008 international conference on web search and data mining*. ACM, 2008, pp. 231–240.

[11] K. Ganchev, J. Gillenwater, and B. Taskar, "Dependency grammar induction via bitext projection constraints," in *Joint Conference of the Meeting of the ACL and the International Joint Conference on Natural Language Processing of the Afnlp: Volume*. ACL, 2009, pp. 369–377.

[12] W. Zhao, Z. Guan, L. Chen, X. He, D. Cai, B. Wang, and Q. Wang, "Weakly-supervised deep embedding for product review sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 1, pp. 185–197, 2018.

[13] Y. Long, R. Xiang, Q. Lu, C.-R. Huang, and M. Li, "Improving attention model based on cognition grounded data for sentiment analysis," *IEEE Transactions on Affective Computing*, 2019.

[14] S. M. Jimenez-Zafra, M. T. M. Valdivia, E. M. Camara, and L. A. Urena-Lopez, "Studying the scope of negation for spanish sentiment analysis on twitter," *IEEE Transactions on Affective Computing*, 2017.

[15] T. Mullen and N. Collier, "Sentiment analysis using support vector machines with diverse information sources," in *Proc. EMNLP*, 2004.

[16] G. Paltoglou and M. Thelwall, "A study of information retrieval weighting schemes for sentiment analysis," in *Proc. ACL*. Association for Computational Linguistics, 2010, pp. 1386–1395.

[17] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," in *Proce. ICML)*, 2011, pp. 513–520.

[18] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[19] R. Socher, C. C. Lin, C. Manning, and A. Y. Ng, "Parsing natural scenes and natural language with recursive neural networks," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 129–136.

[20] M. Zhang, Y. Zhang, and D.-T. Vo, "Gated neural networks for targeted sentiment analysis," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[21] C. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 69–78.

[22] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," 1999.

[23] X. Zhou, X. Wan, and J. Xiao, "Attention-based lstm network for cross-lingual sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016, pp. 247–256.

[24] B. Shin, T. Lee, and J. D. Choi, "Lexicon integrated cnn models with attention for sentiment analysis," *arXiv preprint arXiv:1610.06272*, 2016.

[25] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, p. e1253, 2018.

[26] W. X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modeling aspects and opinions with a maxent-lda hybrid," in *Proceedings of the 2010 conference on empirical methods in natural language processing*. Association for Computational Linguistics, 2010, pp. 56–65.

[27] C. Wu, F. Wu, S. Wu, Z. Yuan, and Y. Huang, "A hybrid unsupervised method for aspect term and opinion target extraction," *Knowledge-Based Systems*, vol. 148, pp. 66–73, 2018.

[28] M. Mitchell, J. Aguilar, T. Wilson, and B. Van Durme, "Open domain targeted sentiment," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, pp. 1643–1654.

[29] M. Zhou, "Target-dependent twitter sentiment classification," *Proceedings of Annual Meeting of the Association for Computational Linguistics Human Language Technologies*, vol. 1, pp. 151–160, 2011.

[30] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.

[31] J. Wang, J. Li, S. Li, Y. Kang, M. Zhang, L. Si, and G. Zhou, "Aspect sentiment classification with both word-level and clause-level attention networks." in *IJCAI*, 2018, pp. 4439–4445.

[32] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.

[33] L. He, K. Lee, M. Lewis, and L. Zettlemoyer, "Deep semantic role labeling: What works and what's next," in *Proc. ACL (Volume 1: Long Papers)*, vol. 1, 2017, pp. 473–483.

[34] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," *arXiv preprint arXiv:1605.08900*, 2016.

[35] X. Li, L. Bing, P. Li, W. Lam, and Z. Yang, "Aspect term extraction with history attention and selective transformation," *arXiv preprint arXiv:1805.00760*, 2018.

[36] D. Hazarika, S. Poria, P. Vij, G. Krishnamurthy, E. Cambria, and R. Zimmermann, "Modeling inter-aspect dependencies for aspect-based sentiment analysis," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, vol. 2, 2018, pp. 266–270.

[37] E. Cambria, S. Poria, A. Gelbukh, and M. Thelwall, "Sentiment analysis is a big suitcase," *IEEE Intelligent Systems*, vol. 32, no. 6, pp. 74–80, 2017.

[38] O. Wu, T. Yang, M. Li, and M. Li, "hot lexicon embedding-based two-level lstm for sentiment analysis," 2018.

[39] T. Dozat and C. D. Manning, "Simpler but more accurate semantic dependency parsing," *arXiv preprint arXiv:1807.01396*, 2018.

[40] S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, "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.

[41] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "Semeval-2014 task 4: Aspect based sentiment analysis," *Proceedings of International Workshop on Semantic Evaluation at*, pp. 27–35, 2014.

[42] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, A.-S. Mohammad, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," in *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, 2016, pp. 19–30.

[43] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "Exploiting document knowledge for aspect-level sentiment classification," *arXiv preprint arXiv:1806.04346*, 2018.

[44] Q. Liu, H. Zhang, Y. Zeng, Z. Huang, and Z. Wu, "Content attention model for aspect based sentiment analysis," in *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, 2018, pp. 1023–1032.

[45] Z. Chen and T. Qian, "Transfer capsule network for aspect level sentiment classification," in *Proceedings of the 57th Conference of the Association for Computational Linguistics*, 2019, pp. 547–556.

[46] Z. Li, Y. Wei, Y. Zhang, X. Zhang, and X. Li, "Exploiting coarse-to-fine task transfer for aspect-level sentiment classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 4253–4260.

[47] X. Li, L. Bing, W. Lam, and B. Shi, "Transformation networks for target-oriented sentiment classification," *arXiv preprint arXiv:1805.01086*, 2018.

[48] T. Jialong, Z. Lu, J. Su, Y. Ge, L. Song, L. Sun, and J. Luo, "Progressive self-supervised attention learning for aspect-level sentiment analysis," in *Proceedings of the 57th Conference of the Association for Computational Linguistics*, 2019, pp. 557–566.

[49] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[50] H. Xu, B. Liu, L. Shu, and P. S. Yu, "Bert post-training for review reading comprehension and aspect-based sentiment analysis," *arXiv preprint arXiv:1904.02232*, 2019.

[51] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. EMNLP*, 2014, pp. 1532–1543.

[52] R. Socher, J. Bauer, C. D. Manning *et al.*, "Parsing with compositional vector grammars," in *Proc. ACL (Volume 1: Long Papers)*, vol. 1, 2013, pp. 455–465.

[53] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[54] Q. Qian, M. Huang, J. Lei, and X. Zhu, "Linguistically regularized lstms for sentiment classification," *arXiv preprint arXiv:1611.03949*, 2016.

[55] Q. Li, D. F. Wong, L. S. Chao, M. Zhu, T. Xiao, J. Zhu, and M. Zhang, "Linguistic knowledge-aware neural machine translation," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 26, no. 12, pp. 2341–2354, 2018.

Tao Yang received the B.Eng. degree in electronic information engineering from Civil Aviation University of China, Tianjin (CAUC), China, in 2016. Since Sept. 2016, he has been in M.Eng. degree program of 3 years in CAUC and majored in control engineering. Now he is a research intern in Center for Applied Mathematics (CAM), Tianjin University, China. His research interests include text mining and deep learning.

Qing Yin received the B.Eng. degree in applied mathematics from Qufu Normal University, Shandong (QNU), China, in 2017. Since Sept. 2017, she has studied for a master's degree at Center for Applied Mathematics (CAM), Tianjin University, China. The main research direction is data mining and machine learning. Now her research mainly about semantic analysis and QA system.

Lei Yang received the B.S. degree in Mathematics and Applied Mathematics from Hebei University, Baoding, China, in 2017. From September 2017, he has been in M.S. degree of 3 years in Center for Applied Mathematics (CAM) and majored in Computational Mathematics. His research interests include text mining and deep learning.

Ou Wu received the BSc degree in Electrical Engineering from Xi'an Jiaotong University, China, in 2003, and the MSc degree and the PhD degree in Computer Science from the National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, China, in 2006 and 2012. From April 2007, he joined NLPR as an Assistant Professor. From February 2017, he joined Center for Applied Mathematics (CAM), Tianjin University as a full Professor. His research interests include data mining and machine learning.