Recurrent Neural CRF for Aspect Term Extraction with Dependency Transmission

Lindong Guo[1], Shengyi Jiang[1,2(✉)], Wenjing Du[1], and Suifu Gan[1]

[1] School of Information Science and Technology,
Guangdong University of Foreign Studies, Guangzhou, China
`l216920263@qq.com, 515056384@qq.com, 657742829@qq.com,`
`jiangshengyi@163.com`
[2] Engineering Research Center for Cyberspace Content Security of Guangdong
Province, Guangzhou, China

Abstract. This paper presents a novel neural architecture for aspect term extraction in fine-grained sentiment computing area. In addition to amalgamating sequential features (character embedding, word embedding and POS tagging information), we train an end-to-end Recurrent Neural Networks (RNNs) with meticulously designed dependency transmission between recurrent units, thereby making it possible to learn structural syntactic phenomena. The experimental results show that incorporating these shallow semantic features improves aspect term extraction performance compared to a system that uses no linguistic information, demonstrating the utility of morphological information and syntactic structures for capturing the affinity between aspect words and their contexts.

Keywords: Aspect term extraction · Dependency transmission
Recurrent neural networks · CRF

1 Introduction

Aspect term extraction [1, 2], also generally named opinion target extraction [3, 4] in some literature, is aimed to identify objects commented in subjective texts. For instance, in a product review of a phone "an average phone with a great screen, but poor battery life", the review is targeted at the phone's "screen" and "battery life", which are aspect terms expected to be extracted. Aspect term extraction is an important prerequisite for fine-grained sentiment analysis. However, sentiment analysis has been based on sentence or paragraph level for years [5, 6], where much accurate information and different opinions towards distinct targets could be missed. To overcome such limitation, aspect-based sentiment analysis has become the focus of a growing number of research.

There are two types of aspects defined in aspect-based sentiment analysis task: explicit aspects and implicit aspects [7]. Explicit aspects are words which appear explicitly in opinioned sentences. In the above example, the opinion targets, "screen" and "battery life" explicitly mentioned in the text are the explicit aspects. On the contrary, implicit aspects are the targets that are not explicitly mentioned in the text but

can be inferred from the context or opinion words. We can see from the following sentence "My phone is shine but expensive" the appearance and price of the phone are implicit aspects which can be deduced from the opinion words "shine" (corresponds to the appearance of the phone) and "expensive" (corresponds to the price of the phone).

In this paper, we focus on the explicit aspect term extraction task. We propose BiLSTM-DT (Bidirectional LSTM with Dependency Transmission), a novel neural network architecture combining the ability of learning long-term sequential dependencies of LSTM [8] and the guidance of syntactical structural priors provided by dependency transmission. The network takes as input a variable length of embeddings, which consist of character-level embedding, word-level embedding and POS embedding. Specifically, the character-level embedding is the concatenation of the two final state outputs of a bidirectional Recurrent Neural Networks running over a character stream. The word-level embedding transforms word tokens to word vectors given a pre-trained or randomly initialized word embedding lookup table. And the POS embedding is like the word-level embedding, but it is always initialized randomly at the beginning of the training phrase and it provides beneficial lexicological information which is lacking in word-level embedding. The three types of embeddings are then fed into a bidirectional LSTM network with well-designed dependency transmission between recurrent units. To ensure that the network learns label dependencies, a CRF layer is added as the final output to restrict the decoding process. Experimental results on publicly available datasets show that our proposed model achieves new state-of-the-art performance.

2 Related Work

Various approaches have been proposed to tackle aspect term extraction. These approaches can be roughly classified into unsupervised and supervised ones. For unsupervised approaches, most of them are based on statistics or linguistic rules. With the assumption that aspects of products are mostly nouns or noun phrases, Hu and Liu [7] firstly proposed an approach where explicit aspects, with high frequency in corpus, were extracted by association mining and implicit aspects were also detected by minimum distance from opinion words. Though easy to implement, the approach probably obtains low precision because it is vulnerable to frequent noises. For instance, daily expressions, usually with high frequency, could be mistakenly recognized as explicit aspects. Later, point-wise mutual information (PMI) was introduced by Popescu and Etzioni [9]. The precision was improved by computing PMI value between candidate aspect and some entity-related meronymy discriminators. However, the algorithm needs to collect product category expressions in advance since the category indicators are used to compute PMI scores, and at the same time, the computing of PMI score needs the help of search engine, which will be a time-consuming process. Scaffidi et al. [10] proposed an approach based on language model under the assumption that product aspects are more frequently mentioned in product reviews than in general English texts. However, this approach based on statistical model resulted in favorable performance for frequent aspects but instability for infrequent aspects. Different from the above frequency-based methods, syntactic relations between aspects

and opinion words for aspect term extraction are exploited in follow-up studies [11–13]. Given a small number of seed opinion words, Qiu et al. [12] proposed an algorithm called Double Propagation (DP) that iteratively expands opinion words and extracts aspects simultaneously with pre-defined syntactic rules from dependency parse trees. However, the rules are often targeted at a specific domain, and often encounter problems such as matching order conflict.

On the other hand, aspect term extraction can be regarded as a sequence labeling problem and many supervised methods such as HMMs-based [14] and CRFs-based [15–17] approaches are developed to solve this problem. Jakob et al. [15] conducted experiments on four different domains (movies, web-services, cars and cameras), in which the CRF model was trained with features defined as token, POS tag, short dependency path, word distance and opinion sentence. Conditional Random Fields based methods, have no problems as the above rules based method, but they also rely on a large number of manual process of feature engineering. These features are decisive for the extraction performance.

Recent studies have found that deep neural networks have the ability to automatically learn feature representations. Therefore, the deep learning based aspect term extraction has become an important research direction in this field. Liu et al. [18] proposed to employ different types of Recurrent Neural Networks to extract aspects and showed that fine-tuning RNNs outperform feature-rich CRF models without any task-specific manual features. However, the proposed method simply employed RNNs in conjunction with word embeddings, and hence many linguistic constraints cannot be learned to deliver beneficial information. Wang et al. [19] proposed an approach combining Dependency-Tree Recursive Neural Networks with CRF for aspect terms and opinion words co-extraction, in which syntactic dependencies and semantic robustness are considered. Our method is inspired by this one, however, we differ in the way we incorporate syntactic relations into neural networks. Instead of simply taking recursive neural networks, we adopt Recurrent Neural Networks with carefully-designed dependency transmission, which are able to take into account the syntactic structural priors while simultaneously reserve the naturally sequential context. Yin et al. [20] used RNNs to learn distributed representation of dependency paths and then fed the learned dependency path embeddings as one of features into CRF model to extract aspect terms. Essentially, we view this method as a CRF-based model since the training of embeddings and the extraction model are segregated, and the modeling capacity of neural networks is not made use of in the supervised phrase.

3   Method

3.1   Overview

In this paper, we investigate the problem of aspect term extraction in opinion mining, as a sequence labeling task. Our model consists of three components: an embedding component including character-level embeddings, word embeddings, and POS tagging embeddings, which capture lexicological and morphological features; a bidirectional LSTM layer with dependency transmission capturing contextual and syntactic

correlations among words; a CRF layer leveraging label information to make valid predictions. The main architecture of the full model is shown in Fig. 1.
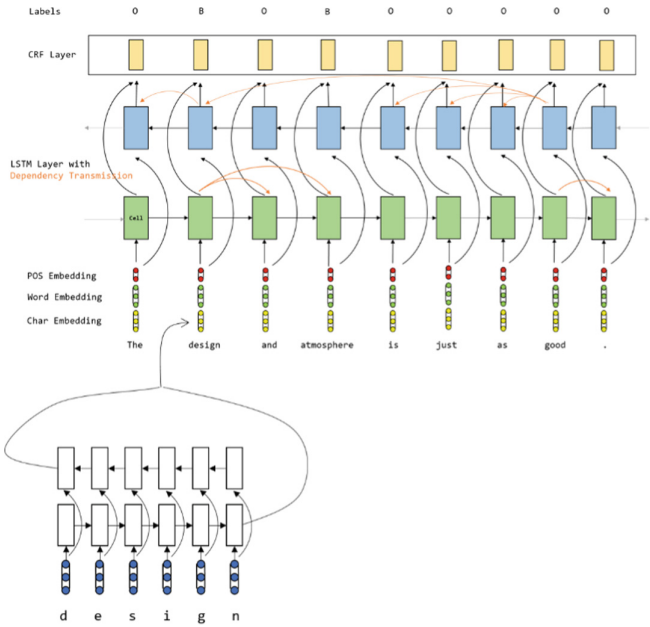

Fig. 1. Architecture of the full model. The left and right output of the character-level bidirectional LSTM are concatenated together to form char embedding. Then char embedding, word embedding, and POS embedding are fed into a sentence-level bidirectional LSTM with dependency transmission. The CRF layer is employed on top of the LSTM to predict BIO labels.

3.2 Embedding Layer

The embedding layer consists of word embedding, character embedding, and POS tag embedding. Word embedding reflecting resemblance between words has been shown effective in a wide range of NLP tasks. Therefore we use it as the basic input of our model. Noting that words themselves of different languages contain rich morphological (e.g. English) or hieroglyphic (e.g. Chinese) information, we can further exploit a neural model to encode words from their own characters in order to make full use of these character-level knowledge. We use Bi-LSTM here since we are tackling English and Recurrent Neural Networks are capable of capturing position-dependent features. We also take advantage of the POS tagging information which guarantees providing strong indicative knowledge for targeted words. Finally, each word in the sentence is associated with a word embedding, a final state output of the forward pass of the character Bi-LSTM, a final state output of the backward pass of the character Bi-LSTM and a POS tagging embedding. These features will be concatenated as a vector and fed to the next layer.

3.3   RNNs Incorporating Dependency Transmission

Given a review sentence $s = \langle w_1, w_2, \ldots, w_T \rangle$ consisting of T words, each represented as an n-dimensional word embedding x_t learned by unsupervised neural nets, a recurrent unit, at time step t, receives the current word embedding x_t and the previous hidden state h_{t-1}, and returns a output representation h_t and a new hidden state.

We first produce a dependency parse for each review sentence using an off-the-shelf dependency parser. In the dependency parse, all words except one are assumed to have a syntactic governor and there is a pre-defined genre of dependency relationship represented as an arc between the governor word and the dependent word. The arcs begin from the previous word and end to the current word will be used in the forward pass of the Recurrent Neural Networks. Similarly, the arcs arise at the rear word and drop at the current word will be used in the backward pass of the network, as depicted in Fig. 2. Each arc is represented as a vector $\mathbf{r} \in \mathbb{R}^d$ and an affine function $(f(\mathbf{r}) = W_r \mathbf{r} + \mathbf{b}_r)$ is introduced to transform the dependency embedding $\mathbf{r}$ to a vector $\mathbf{d}_r$ with the same dimension as the hidden state of the recurrent unit.
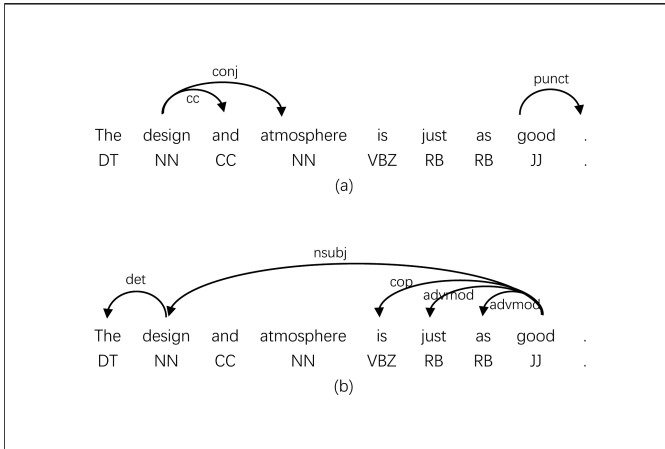

Fig. 2. An illustration of the dependency transmission. (a) represents the dependencies used in the forward pass, and (b) represents the dependencies used in the backward pass.

For each time step t, the input hidden state is now computed from the direct previous hidden state, the vector $\mathbf{d}_r$, and the previous output vector which has a dependency relation with the current token. As an illustration, consider the relation "conj" between words "design" and "atmosphere" in Fig. 2(a), we first calculate the hidden representation of the dependency relation *conj* as follows:

$$\mathbf{d}_{\text{conj}} = f\left(\mathbf{W}_r \mathbf{r}_{\text{conj}} + \mathbf{b}_r\right)$$

where $\mathbf{W}_r$ and $\mathbf{b}_r$ denote the weight matrix and the bias vector, respectively, and f is a non-linear activation function. In our experiments, we choose the hyperbolic tangent

$\tanh(\cdot)$ to be the activation function. After this, the new hidden state is now become the summation of the dependently connected output vector $\boldsymbol{o}_r$, the hidden representation vector $\mathbf{d}_{\text{conj}}$ and the directly previous hidden state vector $\mathbf{c}_{t-1}$, i.e.:

$$\mathbf{c}_{t-1} \leftarrow \mathbf{c}_{t-1} + \boldsymbol{o}_r + \mathbf{d}_{\text{conj}}$$

The updates for LSTM units now become:

$$i_t = \sigma(\boldsymbol{W}_{xi}\boldsymbol{x}_t + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{ci}(\boldsymbol{c}_{t-1} + \boldsymbol{o}_r + \boldsymbol{d}) + \boldsymbol{b}_i)$$

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_{xf}\boldsymbol{x}_t + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{cf}(\boldsymbol{c}_{t-1} + \boldsymbol{o}_r + \boldsymbol{d}) + \boldsymbol{b}_f)$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t \cdot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \cdot \tanh(\boldsymbol{W}_{xt}\boldsymbol{x}_t + \boldsymbol{W}_{hc}\boldsymbol{h}_{t-1} + \boldsymbol{b}_c)$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_{xo}\boldsymbol{x}_t + \boldsymbol{W}_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{W}_{co}(\boldsymbol{c}_t + \boldsymbol{o}_r + \boldsymbol{d}) + \boldsymbol{b}_o)$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t \cdot \tanh(\boldsymbol{c}_t)$$

The motivation for this dependency transmission is that the syntactic priors offer beneficial clues about the high-level abstract concepts of the sentence that may help the extraction of the aspect words.

3.4   CRF Layer and Objective Function

The CRF layer is superior to conventional cross entropy loss and has been proved effective in modeling sequence labeling decisions with strong dependencies. In aspect term extraction scenario, for example, it is impossible that a label I is directly followed by a label O. Accordingly, we feed the outputs of the previous bidirectional LSTM layer into a CRF layer as the unary potentials.

Given an input sentence $\mathbf{x} = \langle x_1, x_2, \ldots, x_T \rangle$ and a label sequence $\mathbf{y} = \langle y_1, y_2, \ldots, y_T \rangle$, the score of the label predictions for $\mathbf{x}$ is calculated as:

$$\text{score}(\mathbf{x}, \mathbf{y}) = \sum_{t=0}^{T} T_{y_t, y_{t+1}} + \sum_{t=1}^{T} H_{t, y_t}$$

where T is the transition matrix denoting the probabilities of one tag transiting to another, and H is the matrix stacked by the Bi-LSTM outputs.

The probability for the label sequence $\mathbf{y}$ given $\mathbf{x}$ is then computed from $\text{score}(\mathbf{x}, \mathbf{y})$, using a softmax transformation:

$$p(\mathbf{y}|\mathbf{x}) = \frac{e^{s(\mathbf{x}, \mathbf{y})}}{\sum_{\hat{\mathbf{y}} \in \mathbf{Y}_{\mathbf{x}}} e^{s(\mathbf{x}, \hat{\mathbf{y}})}}$$

where $\mathbf{Y}_{\mathbf{x}}$ is the set containing all conceivable assignments of sequence labels for $\mathbf{x}$.

The network parameters are chose to minimize the negative log-likelihood of the gold tag sequence for an input $\mathbf{x}$:

$$L(\theta) = -\sum_{\mathbf{x},\mathbf{y}} \log(p(\mathbf{y}|\mathbf{x}))$$

While inferencing, we find the best label sequence $\mathbf{y}^*$ that gives a maximum probability, which can be calculated efficiently by Viterbi algorithm.

$$\mathbf{y}^* = argmax_{\mathbf{y}' \in \mathbf{Y_X}} p(\mathbf{y}'|\mathbf{x})$$

4 Experiments

4.1 Datasets and Evaluation

The experiments are conducted on two publicly available datasets provided by SemEval-2014 Task 4: Aspect Based Sentiment Analysis. Table 1 presents some basic corpus statistics and feature statistics we used in the experiments.

Table 1. Corpus statistics: words are case-insensitive and an off-the-shelf NLP tool was used to tokenize the review sentence and generate the POS and dependencies.

	Laptop		Restaurant	
	Training set	Test set	Training set	Test set
# Sentences	3045	800	3041	800
# Aspect terms	2358	654	2786	818
# Words	4217	1913	4558	2251
# Chars	85	87	80	82
# POS	44	44	44	43
# Dependency	33	33	33	33

Evaluation. We choose the same evaluation metrics suggested in ABSA task.

$$F_1 = \frac{2TP}{2TP + FP + FN}$$

True positives (TP) are defined as the set of aspect terms in gold standard for which there exists a predicted aspect term that matches exactly. In our experiments, t-test is used to evaluate the statistical significance between two models and the corresponding p-value is reported in the table.

4.2   Experimental Settings and Compared Models

Pre-trained Word Embeddings. We used two domain-specific corpus of Amazon Product Data and Yelp Open Dataset for word embedding pre-training. Due to the similarity consideration, we chose the Electronics category provided by the Amazon corpus, which consists of 7,824,482 user reviews. The Yelp Open Dataset consists of 4,736,898 user reviews of various restaurants. All the unlabeled review texts were tokenized by the MBSP system. We trained the word embeddings of 300-dimensions using CBOW architecture with negative sampling.

Experimental Settings. For the two labeled review datasets, we perform tokenization, part-of-speech tagging, and dependency parsing all by Stanford CoreNLP [21]. We build the char vocabulary and word vocabulary from training set and embedding raw corpus by removing low frequency words. This resulted in a vocabulary of approximate 20 K/13 K words for Laptop/Restaurant dataset. In addition, we replace normal number strings, ordinal number, and time expression with NUM, ORD, and $TIM $, respectively. In test phrase, when encounter an unknown word, we replace it with UNK. For char level, we just ignore that character. All sentences will be padded to the maximum length with PAD.

The dimension of the word embedding, character embedding, POS tagging embedding and dependency embedding is 300, 100, 100, and 100, respectively. The size of the hidden state of the character Bi-LSTM is set to 100, while the sentence-level Bi-LSTM is set to 300. We adopt the Adam optimizer default parameters (lr: 0.001, beta1: 0.9, beta2: 0.999) and batch size 20. All hyper-parameters are chosen via cross validation. To further eliminate the influence of random error, we train 10 models with the same hyper-parameters and an average score is calculated on the test set, instead of only 5 times as [22].

Baseline and Comparable Models. To evaluate the effectiveness of our method with dependency transmission, we conduct comparison experiments with the following state-of-the-art models:

ISH_RD_Belarus: The top system for Laptop domain in SemEval 2014 Challenge Task 4. A linear-chain CRF model with a variety of hand-engineered feature sets, including token, part-of-speech, named entity, semantic category, semantic orientation, frequency of token occurrence, opinion target, noun phrase, semantic label and SAO features. The model was trained on a blend of both two domain training sets and used to predict all test sets with the same settings.

DLIREC: The top system for Restaurant domain in SemEval 2014 Challenge Task 4. The system also used a CRF-based model with rich handcrafted features. In addition to general features commonly used in NER systems, voluminous extrinsic resources are exploited to generate word cluster and name list as features in the system.

RNCRF + F (Wang et al. [19]): A Recursive Neural Network with CRF as the output layer. The results reported here are produced by the best setting incorporating hand-crafted features such as name list and sentiment lexicon.

WDEmb_CRF (W + L + D + B Yin et al. [20]): A CRF-based model with embedding features as input, in which the word embedding, the linear context embedding, and the dependency context embedding are trained unsupervisedly.

MIN (Li et al. [22]): A LSTM-based model with memory interactions. The full model is trained with multi-task learning setting.

Giannakopoulos et al. [23]: A regular bidirectional LSTM based model with CRF layer as final output. Additionally, the author conducted experiments on automatically labelled datasets.

4.3   Results and Analysis

In Table 2 we present the extraction performances evaluated by F1 score, compared with those of previous state-of-the-art models. We see that our model significantly outperforms the best systems in SemEval 2014 challenge, by 5.67% and 1.96% absolute gains on Laptop and Restaurant domains respectively, suggesting that deep neural network is capable of memorizing pivotal patterns for aspect term extraction while the latter systems rely on extensive hand-crafted feature engineering and template rules. With consideration of POS tagging information and dependency transmission, our model have surmounted the results of previous published works on each dataset. It clearly demonstrates the effectiveness of leveraging linguistic knowledge and the carefully designed dependency transmissions between recurrent units.

Table 2. Experimental results.

Models	D1	D2
IHS_RD (CRF-based, top system)	74.55	79.62
DLIREC (CRF-based, top system)	73.78	84.01
RNCRF + F(Wang et al., EMNLP, 2016)	78.42	84.93
W + L + D + B (Yin et al., IJCAI, 2016)	75.16	84.97
MIN (Li et al., EMNLP, 2017)	77.58	–
Giannakopoulos et al., EMNLP, 2017	77.96	84.12
Ours (full model)	**80.22**	85.96
Ours (-character embedding)	79.54	**85.97**
Ours (-POS tagging)	79.58	85.38
Ours (-dependency transmission)	79.79	85.41
Ours (only word embedding)	79.77	85.10

Ablation Experiments. To further provide insight into the contribution of the constituent parts of the overall performance, we carry out ablation experiments. Table 2 presents the ablation results in terms of F1 performance. Without character features, the extraction performance declines on Laptop domain but is roughly the same on Restaurant domain. This is because there are more OOVs in the Laptop domain, which proves that character-level embedding helps deal with unknown words. We find that using POS tagging information helps boost the performance since aspect terms usually

appear as nominal words or phrases. More importantly, we observed that dependency transmission is significantly contributive to increasing the performance, indicating that it is useful for capturing skip information.

Different Word Embeddings. Besides ablation experiments, we also carry experiments to observe the performance with different pre-trained word embeddings. The experimental results are reported in Table 3. Our domain-specific pre-trained word embedding yields the best performance on all datasets. The result indicates that the pertinence of the corpus is probably more important than its size when it is used to train the word embedding, since the size of Google News corpus is much larger than the Yelp Dataset or the Amazon Review.

Table 3. Experimental results with different pre-trained word embeddings

	D1	D2
word2vec(GoogleNews, 300d)	75.43	79.98
fastText(Wikipedia, 300d)	73.94	81.43
GloVe(Wikipedia 2014 + Gigaword 5, 300d)	71.51	81.73
Ours(Yelp Dataset/Amazon Review, 300d)	**80.22**	**85.96**

Error Analysis. Table 4 presents some examples which are not handled well by our model. Sentence (a) and sentence (b) both introduce external aspects which do not belong to its own domain. The reviewer of sentence (a) expresses his/her opinion by using the "movie" as a metaphor, while sentence (b) refers to "air flow" with strong aspect indicator "good". These linguistic phenomena are not unusual in review text and bear a non-negligible responsibility for interfering the performance of the extraction system. Applying metaphor recognition and introducing domain-specific knowledge may alleviate the interference. We left the verification of this conjecture to future research. Sentence (c) and sentence (d) are examples in which aspect words appear as verbs. It is relatively uncommon since aspect words are usually nouns or noun phrases. Other errors caused by conditions such as human error or unknown words are not discussed here since this can be solved through qualitative and quantitative improvements.

5   Conclusion

In this paper, we have presented a novel architecture leveraging both the sequential message and structural priors for aspect words extraction in opinioned reviews. In addition to fundamental features such as character-level morphology and word-level POS tagging, which have been used extensively to improve performance, we investigate the probability of incorporating structural information into Recurrent Neural Networks. To achieve this, we equip the recurrent unit with the ability to receive information from the dependently connected word, and this ensures that the Recurrent Neural Networks are able to learn features sequentially and structurally. As expected, the comparison of other state-of-the-art models with the experimental results show that

Table 4. Error analysis. For each example, the first line represents the words, and the second and the third denote the gold labels and the predicted labels, respectively.

it	's	good	but	,	like	the	movie	is	never	as	the	book	,	this	is	same	analogy	.
O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
O	O	O	O	O	O	O	B	O	O	O	O	O	O	O	O	O	O	O

(a)

be	sure	to	have	good	air	flow	where	ever	you	put	it	.
O	O	O	O	O	O	O	O	O	O	O	O	O
O	O	O	O	O	B	I	O	O	O	O	O	O

(b)

it	's	just	what	we	were	looking	for	and	it	works	great	.
O	O	O	O	O	O	O	O	O	O	B	O	O
O	O	O	O	O	O	O	O	O	O	O	O	O

(c)

starts	up	in	a	hurry	and	everything	is	ready	to	go	.
B	I	O	O	O	O	O	O	O	O	O	O
O	O	O	O	O	O	O	O	O	O	O	O

(d)

the proposed model exhibits a more favorable performance and confirm the effectiveness of the dependency transmission between relational words.

In addition, we have performed error analysis on a few noteworthy sentences, which may provide future directions to implement a more accurate extraction system. One of them is the metaphor recognition in reviews since the increasing number of such sentences used to express analogical feelings about the product, in which the tenor should be recognized as the aspect instead of the vehicle. Furthermore, though the proposed architecture is presented in the context of handling aspect words extraction, it will be a considerable future direction to employ this model to other NLP applications such as fine-grained sentiment classification and stance detection.

References

1. Mukherjee, A., Liu, B.: Aspect extraction through semi-supervised modeling. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 339–348 (2012)
2. Pavlopoulos, J., Androutsopoulos, I.: Aspect term extraction for sentiment analysis: new datasets, new evaluation measures and an improved unsupervised method. In: Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM), pp. 44–52 (2014)
3. Liu, K., Xu, L., Zhao, J.: Opinion target extraction using word-based translation model. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics, pp. 1346–1356 (2012)
4. Liu, K., Xu, L., Liu, Y., et al.: Opinion target extraction using partially-supervised word alignment model. In: Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, pp. 2134–2140. AAAI Press (2013)
5. Socher, R., Perelygin, A., Wu, J., et al.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1631–1642 (2013)
6. Lin, Y., Lei, H., Wu, J., et al.: An empirical study on sentiment classification of Chinese review using word embedding. In: Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation: Posters, pp. 258–266 (2015)
7. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177. ACM (2004)
8. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
9. Popescu, A.M., Etzioni, O.: Extracting product features and opinions from reviews. In: Kao, A., Poteet, S.R. (eds.) Natural Language Processing and Text Mining, pp. 9–28. Springer, London (2007). https://doi.org/10.1007/978-1-84628-754-1_2

10. Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H., Jin, C.: Red Opal: Product-feature scoring from reviews. In: Proceedings of the 8th ACM Conference on Electronic Commerce, pp. 182–191. ACM (2007)
11. Zhuang, L., Jing, F., Zhu, X.Y.: Movie review mining and summarization. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management, pp. 43–50. ACM (2006)
12. Qiu, G., Liu, B., Bu, J., et al.: Opinion word expansion and target extraction through double propagation. Comput. Linguist. 37(1), 9–27 (2011)
13. Teng-Jiao, J., Chang-Xuan, W., De-Xi, L.: Extracting target-opinion pairs based on semantic analysis. Chin. J. Comput. 40(3), 617–633 (2017)
14. Jin, W., Ho, H.H.: A novel lexicalized HMM-based learning framework for web opinion mining. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 465–472. ACM (2009)
15. Jakob, N., Gurevych, I.: Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp. 1035–1045 (2010)
16. Li, F., Han, C., Huang, M., et al.: Structure-aware review mining and summarization. In: International Conference on Computational Linguistics. Association for Computational Linguistics, pp. 653–661 (2010)
17. Yang, B., Cardie, C.: Joint inference for fine-grained opinion extraction. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 1640–1649 (2013)
18. Liu, P., Joty, S., Meng, H.: Fine-grained opinion mining with recurrent neural networks and word embeddings. In: Conference on Empirical Methods in Natural Language Processing, pp. 1433–1443 (2015)
19. Wang, W., Pan, S.J., Dahlmeier, D., et al.: Recursive neural conditional random fields for aspect-based sentiment analysis. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 616–626 (2016)
20. Yin, Y., Wei, F., Dong, L., et al.: Unsupervised word and dependency path embeddings for aspect term extraction. In: International Joint Conference on Artificial Intelligence, pp. 2979–2985. AAAI Press (2016)
21. Manning, C., Surdeanu, M., Bauer, J., et al.: The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55–60 (2014)
22. Li, X., Lam, W.: Deep multi-task learning for aspect term extraction with memory interaction. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2886–2892 (2017)
23. Giannakopoulos, A., Musat, C., Hossmann, A., et al.: Unsupervised aspect term extraction with B-LSTM & CRF using automatically labelled datasets. In: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 180–188 (2017)