An Integrated Semantic-Syntactic SBLSTM Model for Aspect Specific Opinion Extraction

Zhongming Han[1,2(✉)], Xin Jiang[1], Mengqi Li[1], Mengmei Zhang[1], and Dagao Duan[1]

[1] School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China
webir@l63.com
[2] Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing, China

Abstract. Opinion Mining (OM) of Internet reviews is one of the key issues in Natural Language Processing (NLP) field. This paper proposes a stacked Bi-LSTM aspect opinion extraction model in which semantic and syntactic features are both integrated. The model takes embedded vector which is composed by word embedding, POS tags and dependency relations as its input while taking label sequence as its output. The experimental results show the effectiveness of this structural features embedded stacked Bi-LSTM model on cross-domain and cross-language datasets, and indicate that this model outperforms the state-of-the-art methods.

Keywords: Aspect · Opinion extraction · Dependency tree
Stacked Bi-LSTM

1 Introduction

With the evolution of the Internet, OM has become one of the most vigorous research areas in NLP field. An aspect is a concept in which the opinion is expressed in the given text [1]. Aspect specific OM task can be divided into four main subtasks: aspect extraction, opinion extraction, sentiment analysis and opinion summarization [2]. This paper focuses on the second subtasks: opinion extraction. In this paper, we propose a hierarchical model based on stacked Bi-LSTM using both semantic information and syntactic information as input to extract aspect specific opinions.

Internet reviews OM can be carried out from three directions: document-level OM [3], sentence-level OM and aspect-level OM. Aspect-level OM is to extract both the aspects and the corresponding opinion expressions in sentences [4]. The extraction of opinion towards its corresponding aspect is a core task in Aspect-level OM. In recent years, the neural network has reached remarkable effect in NLP. Pang et al. [5] committed a survey of the current deep models used to handle text sequence issues. Socher et al. [6] proposed the recursive neural tensor network and represent phrases by distributed vectors. RNN [7] and its variants such as LSTM [8] and GRU [9] stood out from various deep learning methods. Huang et al. [10] proposed a bidirectional LSTM-CRF model for sequence labeling, and on this basis, Ma et al. [11] joined the CNNs in

X. Meng et al. (Eds.): WISA 2018, LNCS 11242, pp. 191–199, 2018.
https://doi.org/10.1007/978-3-030-02934-0_18

the model to encode character-level information of a word into its character-level representation. Du et al. [12] proposed an attention mechanism based RNN model which contains two bidirectional LSTM layers to label sequences so that to extract opinion phrases. Nevertheless, the neural networks' performance drops rapidly when the models solely depend on neural embedding as input [11].

2 SBLSTM Model

2.1 SBLSTM Model Structure

We model aspect opinion extraction as a sequence labeling. The input of the model includes embedded vector, POS tags and dependency relations. The output is the corresponding label sequence of the input text sequence. We use a stacked Bi-LSTM between the input layer and output layer. Opinion expressions extraction has often been treated as a sequence labeling task. This kind of method usually uses the conventional B-I-O tagging scheme.

The basic idea of LSTM is to present each sequence forwards and backwards to two separate hidden states to capture past and future information. Then the two hidden states are concatenated to form the final output. The bidirectional variant of one unit's hidden state's update at time step t is as following.

$$\overrightarrow{h_t} = \vec{g}\left(\overrightarrow{h_{t-1}}, x_t\right)\left(\overrightarrow{h_0} = 0\right) \tag{1}$$

$$\overleftarrow{h_t} = \overleftarrow{g}\left(\overleftarrow{h_{t+1}}, x_t\right)\left(\overleftarrow{h_T} = 0\right) \tag{2}$$

$h_t = \left[\overrightarrow{h_t}, \overleftarrow{h_t}\right]$ can be regarded as an intermediate representation containing the information from both directions to predict the label of the current input x_t.

Stacked RNNs is stacked by k($k \geq 2$) RNN networks. The first RNN receives the word embedding sequences as its input and the last RNN forms the abstract vector representation of the input sequence which is used to predict the final labels. Suppose the output of $\mathbf{j^{th}}$ RNN on time-step t is $\mathbf{h_t^j}$, the stacked RNNs can be formulated as following.

$$h_t^j = \begin{cases} g\left(h_{t-1}^j, x_t\right) & j = 0 \\ g\left(h_{t-1}^j, h_t^{j-1}\right) & otherwise \end{cases} \tag{3}$$

The function g in (3) can be replaced by any RNN transition functions. We expect to capture the important opinion elements. Therefore, we choose the 2 layer Stacked-BiLSTM network as the basic model, and adds attention mechanism to it. In this attention model, the second BLSTM's input i_t^2 at time t can be expressed as:

$$i_t^2 = \sum_{s=1}^{T} \alpha_{ts} h_s^1 \tag{4}$$

where h_s^1 is the output vector of the first BLSTM at time s, α_{ts} is the weight of the output vector sequence $[h_1^1, h_2^1, h_3^1, \ldots, h_T^1]$, the product of which is the input of the second BLSTM at the time t. The weight α_{ts} is calculated as follows:

$$e_{ts} = tanh\left(W^1{}_s^1 + W^2{}_{t-1}^2 + b\right) \tag{5}$$

$$\alpha_{ts} = \frac{exp\left(e_{ts}^T e\right)}{\sum_{k=1}^{T} exp\left(e_{tk}^T e\right)} \tag{6}$$

where W^1 and W^2 are the parameter matrixs that update in the model training process. b is the bias vector. e and e_{ts} has the same dimension and also update with the above adjustable parameters in the model training process.

Figure 1 demonstrates a stacked Bi-LSTM model consisting two Bi-LSTMs with an attention layer. The input is distributed word vectors of texts while the output is a series of B-I-O tags predicted from the network. In order to make the stacked RNNs to be extended easily, we use stacked bidirectional LSTMs with depth of 2 as our basic model in this paper.

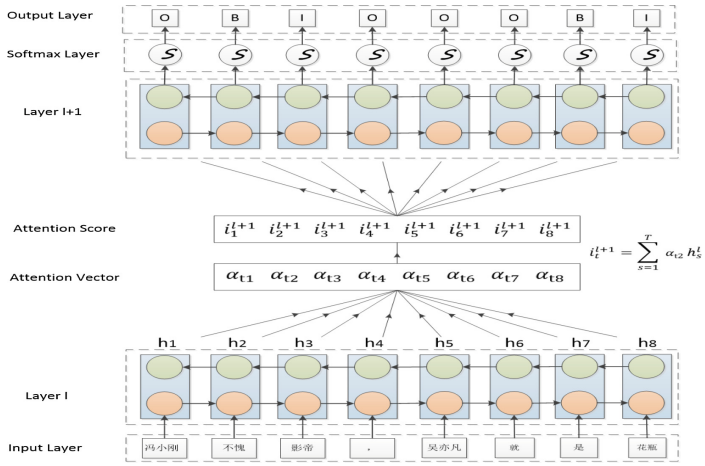

Fig. 1. A stacked bidirectional LSTM network

2.2 Features in SBLSTM Model

In SBLSTM model, the features used is as following

- *Word embeddings*. The word embedding is a kind of distributed vector which contains the semantic information.

- *POS tags*. We use Stanford Tagger to obtain the POS tags.
- *Syntactic tree*. Here we particularly apply the syntactic information, dependency tree in the model. Figure 2 displays the dependency tree for a movie review.

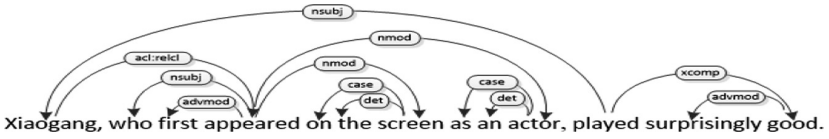

Fig. 2. Dependency tree of an example context

The syntactic representation of one word is defined as its m ($m \geq 0$) children in a dependency tree, where m denotes the window size to limit the amount of the dependency relations of one word for the learning models. Introducing the window size could prevent excessive usage on VRAM.

Finally, the three type's features will be concatenated as the input vector and fed to the SBLSTM model. Figure 3 shows the final features composition of one word.

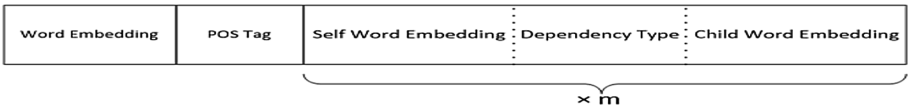

Fig. 3. Input composition of one word

3 Experiments Design and Analysis

3.1 Datasets

For now, there are no available benchmark datasets that mark phrase boundaries of aspect specific expressions. Therefore, two manually constructed datasets are used in our experiment. Mukherjee constructed an annotated corpus which considers 1, 2, and 3-star product reviews from Amazon in English. The other dataset consists of online reviews of three Chinese movies Mr. Six, The Witness and Chongqing Hotpot collected from Douban, Mtime and Sina microblog. The movie reviews dataset is manually annotated.

The statistics information of these two datasets are showed in Table 1. Figures 4 and 5 display the sentences length distributions of the two datasets.

Table 1. Statistics of dataset

Domain	Review	Aspect	Min Length	MaxLength	Language
Product	12102	Earphone, gps, keyboard, mouse, mp3player, router domains	4	292	English
Movie	50000	13 Actors of 3 movies: ***Mr. Six, The Witness*** and ***Chongqing Hotpot***	3	209	Chinese

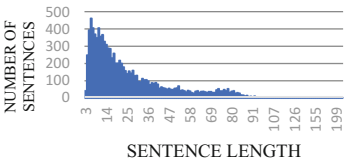

Fig. 4. Distribution of movie dataset sentences length

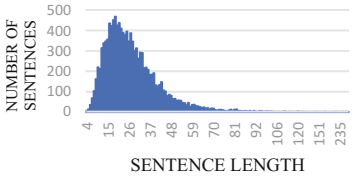

Fig. 5. Distribution of product dataset sentences length

3.2 Experimental Setting

In experiments, we use Stanford parser to obtain the syntactic information. The SBLSM models is implemented in python 2.7 and we use the Keras framework to construct the deep neural networks. The input length is limited to be 60 in LSTMs' models and the amount of LSTM input units is 60 while the number of the output units is 64. The word embeddings dimension is set to be 100. The window size in extracting dependency features is set to be 4 initially. In training process, the ratio between training set and validation set is 4:1. The activation function chosen is softmax function and the batch size to train the model is 256.

3.3 Quantitative Analysis

Evaluation Metrics. Precision, recall and F1 score are commonly used to evaluate the performance of OM models. In OM task, the boundaries of opinion expressions are hard to define. Therefore, we use *proportional overlap* as a soft measure to evaluate the performance.

Model Comparative Analysis. To illustrate the performance boost of our SBLSTM model, we firstly compare our model with some baseline methods on both two datasets. Since we use stacked bidirectional LSTM with depth of 2 as the core model, we choose the LSTM network and the bidirectional LSTM as the baselines. Furthermore, we also

compare our model with the CRF model and rules based method which depends on the dependency tree.

As shown in Table 2, we reported the accuracy, the precision, the recall and the F1-score across all single runs for each approach. We could found that the proposed SBLSTM model outperforms the baseline methods in terms of accuracy, recall and F1. Bi-LSTM outperforms all of the others in terms of precision in the movie dataset, and compared with the CRF model, which achieves the highest precision in the product dataset, our proposed model also provides a comparable precision. Another observation is for both datasets, Bi-LSTM outperforms the normal LSTM model with absolute gains of 4.73% and 4.87% in terms of F1 score.

Table 2. Results of our proposed model against baseline methods

		Accuracy (%)	Precision (%)	Recall (%)	F1-score
Product	Rules-based	78.35	50.78	51.32	0.5105
	CRF	80.66	**74.82**	47.38	0.5802
	LSTM	81.09	72.41	53.26	0.6138
	Bi-LSTM	83.22	73.02	60.39	0.6611
	SBLSTM	**85.78**	73.62	**65.21**	**0.6916**
Movie	Rules-based	83.36	61.13	71.03	0.6571
	CRF	85.64	76.97	54.43	0.6377
	LSTM	86.55	75.41	58.31	0.6576
	Bi-LSTM	90.01	**83.21**	61.35	0.7063
	SBLSTM	**90.73**	77.54	**75.11**	**0.7631**

Feature Comparative Analysis. In training process, the batch size is set to be 256 and the epoch number is set to be 30. Table 3 shows the comparison of experimental results using different feature sets.

Table 3. Comparison of the models performance using different features

	Accuracy (%)	Precision (%)	Recall (%)	F1-score
Word embedding	85.76	75.30	52.23	0.6165
Word embedding + POS	88.13	81.09	55.32	0.6577
POS + dependency relation	89.58	71.29	74.17	0.7270
ALL(W + P + D)	90.73	77.54	75.11	0.7631

Our proposed methods which introduces all of the three types feature performs best in term of accuracy, recall and F1 score. Refer to the third line and the fourth line of Table 3, adding word embeddings into the feature set makes the performance of the model improved in a similar way. This indicates that both word embedding and POS tags have some help in extracting the aspect specific opinion expressions. Particularly, we can observe that the recall measure and the F1-score are improved by 20% and 10%

respectively when the dependency relations have been added into features, providing the evidence that syntactic information does play an important role in extracting the opinions.

Window Size Analysis. We conduct a series of experiments with different window sizes to compare and analyze the impact of the children amount in dependency tree on model performance on movie dataset. The batch size in training process is set to be 256 and each trial was carried out in 300 epochs. Table 4 shows the comparison of the predictive performance of the proposed stacked Bi-LSTM models with both the semantic features and the syntactic features.

From Table 4, we found that F1-score increase with the growth of the window size in general, tending to be stable when the window length is greater than 4.

Table 4. Performance of different window sizes

Window Size	Accuracy (%)	Precision (%)	Recall (%)	F1-score
1	88.85	81.09	55.32	0.6577
2	90.23	69.15	71.32	0.7022
3	90.29	75.17	72.06	0.7358
4	90.73	77.54	75.11	**0.7631**
5	89.73	76.49	74.35	0.7581
6	90.38	75.61	73.56	0.7526
7	88.39	76.47	72.26	0.7431
8	88.83	69.17	78.23	0.7465
9	89.98	69.42	76.47	0.7486
10	89.43	74.42	71.02	0.7387

3.4 Qualitative Analysis

To explore the contribution of this paper, we conducted a qualitative analysis experiment on five chinese movie comments below and the aim aspect is ***FengXiaogang***.

The experiment uses rules-based model, CRF model, LSTM network and Bi-LSTM network as baseline methods. The aspect specific opinion extraction results of different methods are shown in Table 5. The green words are annotated opinion expression which we want models to extract, and the red words refer to those words haven't been extracted by the model while these blue ones are those words not in annotation.

The dependency rule based method is more effective when the sentence is short and simple. When here comes a complex sentence, it is impossible to obtain more information when the comment contains a demonstrative pronoun. Most importantly, no matter the length of the sentence, our model can extract the opinion information well.

Table 5. Aspect specific opinion extraction results of different methods

Comments	Director Feng 's acting is just right, and the nature of the characters of the men in Beijing is in place.	Feng is really worthy of winning the Golden Horse Award, acting just the right place.	A group of old drama kings really acted well, the old Beijing charm was showed out, making people unforgettable, and the last tears is still in my mind.	The best actor title of Feng is well deserved. Tolerance, responsibility and quality which men should have were acted incisively and vividly by him.	The movie's plot is good, but I do not understand the Feng's logic. The lines are always annoying with dirty words
Rules-based	Director Feng 's acting is just right, and the nature of the characters of the men in Beijing is in place.	Feng is really worthy of winning the Golden Horse Award, acting just the right place.	A group of old drama kings really acted well, the old Beijing charm was showed out, making people unforgettable, and the last tears is still in my mind.	The best actor title of Feng is well deserved. Tolerance, responsibility and quality which men should have were acted incisively and vividly by him.	The movie's plot is good, but I do not understand the Feng's logic. The lines are always annoying with dirty words
CRF	Director Feng 's acting is just right, and the nature of the characters of the men in Beijing is in place.	Feng is really worthy of winning the Golden Horse Award, acting just the right place.	Agroup of old drama kings really acted well, the old Beijing charm was showed out, making people unforgettable, and the last tears is still in my mind.	The best actor title of Feng is well deserved. Tolerance, responsibility and quality which men should have were acted incisively and vividly by him.	The movie's plot is good, but I do not understand the Feng's logic. The lines are always annoying with dirty words
LSTM	Director Feng 's acting is just right, and the nature of the characters of the men in Beijing is in place.	Feng is really worthy of winning the Golden Horse Award, acting just the right place.	A group of old drama kings really acted well, the old Beijing charm was showed out, making people unforgettable, and the last tears is still in my mind.	The best actor title of Feng is well deserved. Tolerance, responsibility and quality which men should have were acted incisively and vividly by him.	The movie's plot is good, but I do not understand the Feng's logic. The lines are always annoying with dirty words
Bi-LSTM	Director Feng 's acting is just right, and the nature of the characters of the men in Beijing is in place.	Feng is really worthy of winning the Golden Horse Award, acting just the right place.	A group of old drama kings really acted well, the old Beijing charm was showed out, making people unforgettable, and the last tears is still in my mind.	The best actor title of Feng is well deserved. Tolerance, responsibility and quality which men should have were acted incisively and vividly by him.	The movie's plot is good, but I do not understand the Feng's logic. The lines are always annoying with dirty words
OurModel	Director Feng 's acting is just right, and the nature of the characters of the men in Beijing is in place.	Feng is really worthy of winning the Golden Horse Award, acting just the right place.	A group of old drama kings really acted well, the old Beijing charm was showed out, making people unforgettable, and the last tears is still in my mind.	The best actor title of Feng is well deserved. Tolerance, responsibility and quality which men should have were acted incisively and vividly by him.	The movie's plot is good, but I do not understand the Feng's logic. The lines are always annoying with dirty words

4 Conclusions

In this paper, we proposed a method to embed syntactic information into the deep neural models. Experimental results on two domains and different languages data sets showed that the proposed stacked bidirectional LSTM model outperform all of the baseline methods, proofing that the syntactic information did play a significant role in correctly locating the aspect-specific opinion expressions.

References

1. Poria, S., Cambria, E., Gelbukh, A.: Aspect extraction for opinion mining with a deep convolutional neural network. Knowl.-Based Syst. **108**, 42–49 (2016)
2. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004), pp. 168–177. ACM, New York (2004)
3. Valakunde, N.D., Patwardhan, M.S.: Multi-aspect and multi-class based document sentiment analysis of educational data catering accreditation process. In: International Conference on Cloud & Ubiquitous Computing & Emerging Technologies (CUBE 2013), pp. 188–192. IEEE Computer Society, Washington (2013)
4. Singh, V.K., Piryani, R., Uddin, A., et al.: Sentiment analysis of movie reviews: a new feature-based heuristic for aspect-level sentiment classification. In: International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4 s 2013), pp. 712–717. IEEE Computer Society, Washington (2013)
5. Pang, L., Lan, Y.Y., Xu, J., et al.: A survey on deep text matching. Chin. J. Comput. **40**(04), 985–1003 (2017). (in Chinese with English abstract)
6. Socher, R., Perelygin, A., Wu, J.Y., 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 (EMNLP 2013), pp. 1631–1642. ACL, Stroudsburg, PA (2013)
7. Goller, C., Kuchler, A.: Learning task-dependent distributed representations by backpropagation through structure. In: IEEE International Conference on Neural Networks, vol. 1, pp. 347–352. IEEE (2002)
8. Hochreiter, S., Jurgen, J.: Long short-term memory. Neural Comput. **9**, 1735–1780 (1997)
9. Cho, K., Merrienboer, B.V., Bahdana, D., Bengio, Y.: On the properties of neural machine translation: encoder-decoder approaches. In: Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar, 25 October 2014, pp. 103–111 (2014)
10. Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF models for sequence tagging. Computing Research Repository, abs/1508.01991 (2015)
11. Ma, X., Hovy, E.: End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF (2016)
12. Du, J., Gui, L., Xu, R.: Extracting opinion expression with neural attention. In: Li, Y., Xiang, G., Lin, H., Wang, M. (eds.) SMP 2016. CCIS, vol. 669, pp. 151–161. Springer, Singapore (2016). https://doi.org/10.1007/978-981-10-2993-6_13