Bidirectional Independently Long Short-Term Memory and Conditional Random Field integrated model for Aspect Extraction in Sentiment Analysis

Trang Uyen Tran [1], Ha Thanh Hoang [2] and Hiep Xuan Huynh [3]

[1] Faculty of Informatics, University of Education, The Danang University, Danang, Vietnam
[2] Faculty of Statistics and Informatics, Da Nang University of Economic, The Danang Univesity, Danang, Vietnam
[3] College of Information and Communication Technology, Can Tho University, Cantho, Vietnam
`trang.tranuyen@gmail.com, ha.htt@due.edu.vn, hxhiep@ctu.edu.vn`

Abstract. Aspect extraction or feature extraction is a crucial and challenging task of opinion mining that aims to identify opinion targets from opinion text. Especially, how to explore these aspects or features from unstructured comments is a matter of concern. In this paper, we propose a novel supervised learning approach using deep learning technique for the above mentioned aspect extraction task. Our model combines a Bidirectional Independently Long Short-Term Memory (Bi-IndyLSTM) with a Conditional Random Field (CRF). This integrated model is trained on labeled data to extract feature sets in opinion text. We employ a Bi-IndyLSTM with word embeddings achieved by training GloVe on the SemEval 2014 dataset. There are 6,086 training reviews and 1,600 testing reviews on two domains, Laptop and Restaurant of the SemEval 2014 dataset. Experimental results showed that our proposed Bi-IndyLSTM-CRF aspect extraction model in sentiment analysis obtained considerably better accuracy than the state-of-the-art methods.

Keywords: Aspect extraction, Bi-IndyLSTM, CRF.

1   Introduction

In the recent years, the growth of social media like user reviews, forums, blogs or social network has enabled many people using these public opinions on their decision-making for a particular product or service. Nevertheless, extracting opinion is a challenging task because of the huge amount of information available and limited readability which make it difficult for users to precisely identify information. Consequently, constructing the automated opinion mining system that allows receiving specific opinion polarity on their interested issues easily is extremely requisite.

Sentiment analysis (SA) or opinion mining (OM) [4] is a field of research on the ability to extract and categorize opinions about entities and attributes of products or services. There are many various areas using sentiment analysis: businesses and

organizations observe the public views about their products and commercial services; individual is interested in the views of the community have experienced about the products or services that he care about; sentiments of the community about a political figure before an election. For these reasons, it is necessary to exploit a system to understand and explore opinion automatically in research. How to extract feature or aspect from reviews in order to determine the opinions expressed on it is as the main issue in the specific tasks of opinion mining. The proposed novel aspect extraction models and the improvement of existing aspect extraction techniques are always the core research directions of opinion mining.

In this paper, we propose to apply the novel deep learning approach for supervised aspect extraction task in sentiment analysis using an integrated model of Bi-IndyLSTM and CRF. Our model uses Bi-IndyLSTM with word embeddings obtained by training GloVe [11] on 7,686 reviews of the SemEval 2014 dataset. We choose this IndyLSTM model for the following reasons: (i) can control the gradient backpropagation through time to solve the gradient vanishing and exploding issue; (ii) can well process long sequences based on the ability to keep long-term memory of IndyLSTM; (iii) can raise the model's depth by a great number of IndyLSTM layers without being under gradient-decay's influence over layers like LSTM and GRU; (iv) based on the independence of neurons in each layer, neurons' operation is quite clear and obvious . The experimental results show that our Bi-IndyLSTM and CRF system used for feature extraction proved to be more effective than previous state-of-the-art approaches.

In the rest of this paper, we organize as follows. Section 2 presents the current works and research related to aspect extraction in opinion mining. Our proposed approach for aspect extraction with Bi-IndyLSTM and CRF architecture is described in Section 3. Section 4 identifies our experiments using a standard dataset and received results. Finally, we conclude on the outcomes of this experimentation and the future work in Section 5.

2 Related Works

Most of recent extracting aspects work focus on four main approaches [4]: frequency of nouns and noun phrases; relationship between features and opinions; topic modeling and supervised learning.

 Hu and Liu [15] focused on occurrence frequencies of nouns and noun phrases that have been recognized by a POS tagger and only the frequent ones were considered as aspects. Popescu and Etzioni [1] improved the above algorithm by detecting whether noun or noun phrase is an aspect or not using PMI measure. Scaffidi et al. [8] considered the extracted nouns and noun phrases frequency from sentiment text with the appearance ratio of them in a public English corpus to identify aspects. Long, Zhang and Zhu [7] extracted aspects by using the frequency-based method. They applied the information distance to find the other words related to aspects to choose which reviews discussing these aspects.

The relationships between aspects and opinions are also one of the main approaches that attract scientific researches. Kobayashi, Inui and Matsumoto [17] used a

dependency parser to recognize such dependency relations of individual words for aspect extraction task. Wu et al. [26] used a phrase dependency parser for noun phrases and verb phrases extraction that maybe aspects. Qiu et al. [10] proposed aspect extraction approach using domain dependent corpus. In this method, the syntactic relationship between opinion and aspect with set of seed words is used to extract concurrent opinions and aspects.

In recent year, using topic model for extracting aspect is one of the most popular approaches. Lu et al. [27] applied a blended model of aspect and opinion based on a topic model for opinion mining by using probabilistic Latent Semantic Analysis (pLSA). Lin and He used a method allow identifying both topic and sentiment from document by Latent Dirichlet Allocation (LDA) [6]. However, extracted aspect and opinion words from their model were not clearly separated. In [20] authors applied topic model to identify aspect and then realize opinion words that express on this aspect by only examining adjectives.

Extracting aspect based on supervised learning approach is also recently interested. The most decisive methods are suitable for sequential learning using machine learning techniques such as Hidden Markov Models (HMM) and Conditional Random Fields (CRF). Jakob and Gurevych [16] applied CRF for feature extraction tasks such as tokens, POS tags, and syntactic dependencies. Li et al. [9] also used the same approach to extract both features and opinions by combining two variations of CRF such as Skip-CRF and Tree-CRF. Zhang [14] extracted both relevant aspects and opinions in expanded CRF model using neural network technique. A method of continuousword embedding was used to replace for discrete aspects in the CRF and a neural layer was added in this model. In [24], authors applied an integrated model of recursive neural network and CRF to extract aspects and opinions.

In addition to using traditional machine learning techniques, aspect extraction with supervised learning also applies deep learning methods to increase model accuracy. Katiyar and Cardie [2] used the two-dimensional long short-term memory (LSTM) for extracting opinion entities and identifying the IS-FROM and IS-ABOUT relations between an opinion expression and its holder and target. In [18] authors also used the same number of parameters to explored the application of the two-dimensional deep recurrent neural network (RNN) for the aspect extraction, which outperformed traditional RNNs. Particularly, [19] proposed a common class of the discriminant models using RNN and word embedding. Word embedding in this model was trained from three different sources in various types of RNN including Elman-type, Jordan-type, LSTM and the variants of them for opinion target identification task. A deep convolutional neural network (CNN) was applied in [22]. Authors used a 7-layer deep CNN to tag each word in sentiment text as aspect or not and combine with the set of linguistic patterns. B-LSTM and CRF [3] were also used to extract aspect terms and automatically constructed dataset for this task. Our novel proposed model also belongs to deep learning approach and obtains better accuracy in F1-score than the previous approaches.

3   Methodology

Recurrent neural networks (RNNs) [13], one of deep learning techniques, have been recently used for sequential learning tasks. They have ability to model sequences of arbitrary length. This capacity is due to repetition of a recurrent unit along tokens in the sequence. However, RNNs are limited because of the issues involved in vanishing and exploding gradients [25]. As a result, RNNs are not sufficient for learning long-term dependencies. Based on these disadvantages of RNNs, long short-term memory (LSTM) and gated recurrent unit (GRU) were applied to solve these weak points, but the hyperbolic tangent and sigmoid functions that were used in LSTM and GRU as the activation function can cause gradient decay over layers. As a result, modeling and training a deep LSTM or GRU is really unfeasible. Conversely, if we can use relu as the non-saturated activation function, the layers of neural network can be stacked into an effective deep network aimed for more specific learning purposes. For this reason, we propose using independently long short-term memory (IndyLSTM), a variant of IndRNN [21], in this paper, as the solution for the above mentioned extraction task based on its advantages over RNN, LSTM and GRU.

3.1   Independently Long Short-Term Memory (IndyLSTM)

IndyLSTM is a variant of IndRNN. IndyLSTM was designed to solve vanishing and exploding gradient problem of RNN and gradient decay when using hyperbolic tangent and sigmoid functions as activation functions in multiple LSTM and GRU in a deep network. IndyLSTM was constructed relied on the principle of independent neurons in each layer. Based on IndRNNs and LSTMs, IndyLSTM can be depicted as follows:

$$f_t = \sigma_g\big(W_f x_t + u_f \circ h_{t-1} + b_f\big)$$

$$i_t = \sigma_g\big(W_i x_t + u_i \circ h_{t-1} + b_i\big) \qquad (1)$$

$$o_t = \sigma_g\big(W_o x_t + u_o \circ h_{t-1} + b_o\big)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c\big(W_c x_t + u_c \circ h_{t-1} + b_c\big)$$

where $x_t \in \mathbb{R}^M$ is the input state and $h_t \in \mathbb{R}^N$ is the hidden state at time step t, $W \in \mathbb{R}^{N \times M}$ is the weight for the current input, $U \in \mathbb{R}^{N \times N}$ is the weight for the recurrent input, $b \in \mathbb{R}^N$ is the bias, σ is an activation function of the neurons. Using $\circ$ in these above formulas denotes the Hadamard operator and means that the independence of neurons in an IndyLSTM layer is obvious. Consider in the layer, each neuron will be able to see its own state h and c not all states. The connection between neurons only appears when stack multiple layers of IndyLSTM. In detail, outputs of all neurons in an arbitrary layer are processed by each neuron in the next layer.

3.2   Bidirectional Independently Long Short-Term Memory (Bi-IndyLSTM)

A main issue of unidirectional IndyLSTM is that it allows learning representations from previous time steps. Thus, it only preserves information of the past because the only inputs it has seen are from previous time steps. In some cases, we want to learn

representations from future time steps to better understand the context and eliminate the ambiguity incurred by learning one way. Bidirectional IndyLSTM (Bi-IndyLSTM) can solve this issue.

Figure 1 presents the illustration of the basic Bi-IndyLSTM structure with two components. Bi-IndyLSTM will run inputs in two ways, one from the past to the future and one from the future to the past. Thanks to this advantage, it can preserve information from both past and future and maybe show better results as it can understand context better.
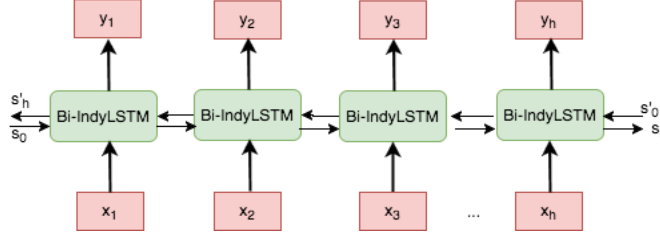

Fig. 1. Bidirectional IndyLSTM model

3.3 Proposed Bi-IndyLSTM-CRF model for extracting aspect

Our network contain three layers: one first word embedding layer; one Bi-IndyLSTM layer composed of two components: forward independently LSTM and backward independently LSTM; and the final CRF layer.

Adding a CRF layer [12] on top of Bi-IndyLSTM to capture dependencies can produce state-of-the-art performance. CRF is a statistical modeling method used for pattern recognition and prediction of structure. CRF has got successful results in discriminative opinion mining tasks such as opinion and aspect term extraction [5][28]. In our proposed model, the CRF layer can add some rules of constrains to the final predicted labels, output results of Bi-IndyLSTM layer, to ensure they are valid. These constrain rules can be learned by the CRF layer automatically from the training dataset during the training process. The constrain rules can be confirmed as follows: (i) the first word label in an opinion sentence should start with "B-" or "O", not "I-"; (ii) the valid patterns should be "O B-label", "B I-label" but "O I-label" is invalid. The first label of one entity, maybe word or phrase, should start with "B-", not "I-". Thus, CRF layer with these useful constrain rules will make the invalid predicted label sequences decrease significantly.

In Figure 2, we feed sentiment sentence into the word embedding layer. The *i-th* word in the sentence is represented by $x_i \in R^k$, the *k-dimensional* word vector. Consequently, a length-*h* sentence is represented as:

$$x_{1:h} = x_1 \oplus x_2 \oplus \ldots \oplus x_h \qquad (2)$$

h is maximum length of the sentence. Each word x_i is represented by 300-dimensional embedding vectors $(w_0, w_1 \ldots w_{300})$. For word embeddings, we use pre-trained word vectors from GloVe. The training process aims for identifying similarities among words using statistical information and co-occurrence matrix. These pre-trained word vectors trained on the words of 7,686 reviews from the SemEval 2014 dataset. In other words, GloVe is as a tool for encoding semantic and syntactic properties of words from

reviews. We follow the *IOB* format for sequential labelling. According to this format, *B* is labelled for tokens that express aspects and *O* is labelled for tokens that do not express aspects of the sentence. If aspect is a phrase included two or more than two tokens, *B* is labelled for the first token and *I* is labelled for the rest token. For example:

Camera/B of/O the/O Iphone/O is/O excellent/O
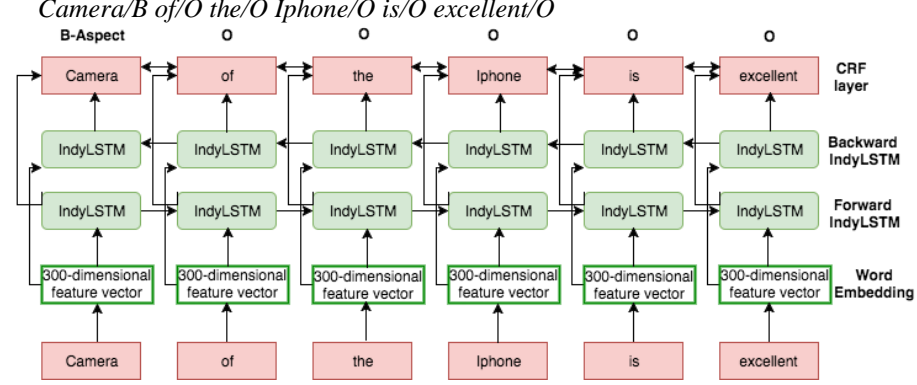
Fig. 2. Bi-IndyLSTM-CRF model for extracting aspect

For each token of an opinion sentence, through GloVe word embedding layer we create the 300-dimensional feature vector and fed to the Bi-IndyLSTM layer. Taking full advantage of the word morphology and the structure of the opinion sentence, Bi-IndyLSTM will extract feature for each token in the sentence. Finally, CRF uses these above extracted features vectors to implement sequential labelling that serve aspect/non aspect classification task. The forward and backward IndyLSTM make themselves responsible for extracting features from the opinion sentence using the previous and the next tokens of each word. This combination is intended to improve the accuracy of the model.

4   Experiments and Results

4.1   Dataset

We use the SemEval 2014 dataset[1] presented in Table 1 for training and evaluation the proposed model. This dataset consists of 7,686 reviews devided into 6,086 reviews in training data and 1,600 reviews in testing data from two domains, Laptop and Restaurant.

Table 1. SemEval 2014 dataset.

Domain	Training	Testing
Laptop	3,041	800
Restaurant	3,045	800
Total	6,086	1,600

---

[1] http://alt.qcri.org/semeval2014/task4/

4.2   Tools for experiments

Our implement Bi-IndyLSTM-CRF model using Tensorflow, version 1.10.1 allows deployment of computation on NVIDIA Tesla K80 GPU[2]. Tesla K80 is constructed to deliver superior performance in recent machine learning and deep learning applications with a range of features like dual-GPU design and Dynamic GPU Boost. Thus, it is absolutely suitable platform for our proposed model.

4.3   Experimental results

We perform experiments for our aspect extraction model in the laptop and the restaurant domain of the SemEval 2014 ABSA contest and evaluate our model using F1-score. Our model for extracting aspect in opinion documents integrates Bi-IndyLSTM and CRF that is presented in Figure 2. We use 300-dimensional feature vector for Bi-IndyLSTM, the *adam* optimizer with learning rate 0.001, *dropout* 0.5 and a *batch size* of 32.
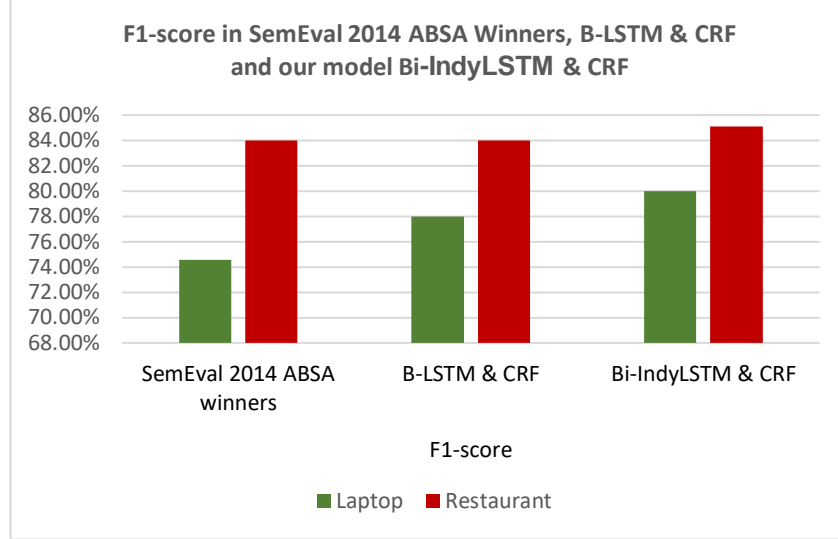
Table 2. F1-score in SemEval 2014 ABSA Winners, B-LSTM & CRF and our model Bi-IndyLSTM & CRF.

Domain	F1-score		
	SemEval 2014 ABSA winners	*B-LSTM & CRF*	*Bi-IndyLSTM & CRF*
Laptop	74.55%	78%	**80%**
Restaurant	84.01%	84%	**85.1%**

The experimental results highlight the differences about accuracy of three models: the Bi-IndyLSTM-CRF, the winners of the SemEval 2014 ABSA and the B-LSTM-CRF clearly in the two domain, Laptop and Restaurant. In detail, Table 2 shows that our model outperforms the previous methods, model of the winners of the SemEval 2014 ABSA contest and B-LSTM-CRF of Athanasios Giannakopoulos et al. for aspect extraction task

- 5.45% and 1.09% higher than SemEval 2014 ABSA winners in the Laptop and Restaurant domain respectively.

- 2% and 1.1% higher than B-LSTM-CRF model in the Laptop & Restaurant domain respectively.

---

[2] https://www.nvidia.com/en-us/data-center/tesla-k80/

Fig. 3. Comparing the accuracy of SemEval 2014 ABSA winners, B-LSTM & CRF and Bi-IndyLSTM & CRF.

As shown in Fig. 3, the differences between the SemEval 2014 ABSA winners, the B-LSTM & CRF and our model Bi-IndyLSTM & CRF are visually expressed.

5   Conclusion

We have proposed a deep learning approach to aspect extraction task in opinion mining. Our model used a Bi-IndyLSTM and CRF integrated architecture that comprises the word embedding layer, GloVe, to make the feature vectors for words in opinion sentence; two IndyLSTM layers, one forward IndyLSTM and one backward IndyLSTM, get the feature vectors from input and take full advantage of information from the previous and the next tokens of each word based on the above Bi-IndyLSTM framework; and one final CRF layer for supporting structure prediction and labelling aspect terms in sequence data. Our proposed model achieved significant improvement in performance over the previous state-of-the-art approaches.

As future work, we plan to perform the different pre-training word embedding technologies for feeding our Bi-IndyLSTM and CRF model. Moreover, we would like to explore alternative variant models of RNN for our problem to get the more effective results. With the innovations of novel RNN-based technologies for natural language processing and opinion mining task, a potential future approach can be able to solve our issue efficiently and result in the significant higher accuracy.

References

1. Ana-Maria Popescu, Oren Etzioni: Extracting product features and opinions from reviews. In: HLT'05 Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 339–346, October 6–8 (2005).
2. Arzoo Katiyar, Claire Cardie: Investigating LSTMs for joint extraction of Opinion entities and relations. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, vol. 1, pp. 919–929. Berlin, August (2016).
3. Athanasios Giannakopoulos, Claudiu Musat, Andreea Hossmann, Michael Baeriswyl: Unsupervised aspect term extraction with B-LSTM & CRF using automatically labelled datasets. ArXiv:1709.05094v1 [cs.CL], September 15th (2017).
4. Bing Liu: Sentiment Analysis and Opinion Mining. Morgan and Claypool Publishers, May (2012).
5. Bishan Yang, Claire Cardie: Extracting opinion expressions with semi-Markov conditional random fields. In: EMNLP-CoNLL'12 Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1335–1345. Jeju Island, Korea, July 12–14 (2012).
6. Chenghua Lin, Yulan He: Joint sentiment/topic model for sentiment analysis. In: CIKM'09 Proceedings of the 18th ACM conference on Information and knowledge management, pp. 375–384. Hong Kong, November 2–6 (2009).
7. Chong Long, Zie Zhang, Xiaoyan Zhu: A review selection approach for accurate feature rating estimation. In: COLING'10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pp. 766–774. Beijing, August 23–27 (2010).
8. Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman, Ng., Chun Jin: Red Opal: product-feature scoring from reviews. In: EC'07 Proceedings of the 8th ACM conference on Electronic commerce, vol. 2, pp. 182–191. June 11–15 (2007).
9. Fangtao Li, Chao Han, Minlie Huang, Xiaoyan Zhu, Ying-Ju Xia, Shu Zhang, Hao Yu: Structure-aware review mining and summarization. In: COLING'10 Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pp. 653–661. Beijing, August 23–27 (2010).
10. Guang Qiu, Bing Liu, Jiajun Bu, Chun Chen: Opinion word expansion and Target extraction through Double Propagation. In: 2011 Association for Computational Linguistics, vol. 37, no. 1 (2011).
11. Jeffrey Pennington, Richard Socher, Christopher D. Manning: GloVe: Global Vectors for Word Representation. In: EMNLP'14 Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 1532–1543. Doha, October 25–29 (2014).
12. John Lafferty, Andrew McCallum, Fernando C. N. Pereira: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th ACM International Conference on Machine Learning, pp. 282–289, June 28th (2001).
13. Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio: Learning phrase representations using RNN Encoder-Decoder for statistical machine translation. ArXiv:1406.1078v3 [cs.CL], September 3rd (2014).
14. Meishan Zhang, Yue Zhang, Duy Tin Vo: Neural networks for open domain targeted sentiment. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 612–621. Lisbon, September 17–21 (2015).
15. Minqing Hu, Bing Liu: Mining opinion features in customer reviews. In: AAAI'04 Proceedings of the 19th national conference on Artificial intelligence, pp. 755–760. July 25–29 (2004).

16. Niklas Jakob and Iryna Gurevych: Extracting opinion targets in a single- and cross-domain setting with conditional random fields. In: EMNLP'10 Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 1035–1045. Cambridge, Massachusetts, October 9–11 (2010).

17. Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto: Extracting aspect–evaluation and aspect-of relations in Opinion mining. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1065–1074. Prague, June (2007).

18. Ozan Irsoi, Claire Cardie: Opinion mining with deep recurrent neural networks. In: EMNLP'14 Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 720–728. Doha, October 25–29 (2014).

19. Pengfei Liu, Shafiq Joty, Helen Meng: Fine-grained Opinion mining with recurrent neural networks and word embeddings. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1433–1443. Lisbon, September 17–21 (2015).

20. Samuel Brody, Noemie Elhadad: An unsupervised aspect-sentiment model for online reviews. In: HLT'10 Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 804–812. Los Angeles, June 02–04 (2010).

21. Shuai Li, Wanqing Li, Chris Cook, Ce Zhu, Yanbo Gao: Independently Recurrent Neural Network (IndRNN): Building A Longer And Deeper RNN. ArXiv: 1803.04831v3 [cs.CV]. May 22th (2018).

22. Soujanya Poria, Erik Cambria, Alexander Gelbukh: Aspect extraction for opinion mining with a deep convolutional neural network: Knowledge-Based Systems, vol. 108, pp. 42–49. September 15th (2016).

23. Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean: Efficient Estimation of Word Representations in Vector Space. ArXiv:1301.3781v3 [cs.CL]. September 7th (2013).

24. Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, Xiaokui Xiao: Recursive Neural Conditional Random Fields for Aspect-based Sentiment Analysis. ArXiv:1603.06679v3 [cs.CL]. September 19th (2016).

25. Yoshua Bengio, P. Simard, Paolo Frasconi: Learning long-term dependencies with gradient descent is difficult. In: IEEE Transactions on Neural Networks, vol. 5, issue 2, pp. 157–166. New York, USA, March (1994).

26. Yuanbin Wu, Qi Zhang, Xuangjin Huang, Lide Wu: Phrase dependency parsing for opinion mining. In: EMNLP'09 Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, vol. 3, pp. 1533–1541. Singapore, August 6–7 (2009).

27. Yue Lu, ChengXian Zhai, Neel Sundaresan: Rated aspect summarization of short comments. In: WWW'09 Proceedings of the 18th ACM International Conference on World Wide Web, pp. 131–140. Madrid, April 20–24 (2009).

28. Yuliya Rubtsova, Sergey Koshelnikov: Aspect extraction from reviews using conditional random fields. In: International Conference on Knowledge Engineering and the Semantic Web, KESW 2015, CCIS, vol. 518, pp. 158–167, October 30th (2015).