Aspect Extraction with Bidirectional GRU and CRF

Trang Uyen Tran
Faculty of Informatics
The University of Education
Danang University
Danang, Vietnam
trang.tranuyen@gmail.com

tutrang@ued.udn.vn

Ha Thanh Thi Hoang
Faculty of Statistics and Informatics
The university of Economic
Danang University
Danang, Vietnam
ha.htt@due.udn.vn

Hiep Xuan Huynh
College of Information and
Communication Technology
Can Tho University
Cantho, Vietnam
hxhiep@ctu.edu.vn

***Abstract*— Opinion mining or sentiment analysis used to understand the community's opinions on a particular product. Sentiment analysis involves building the opinion collection and classification system. One of the most crucial tasks of sentiment analysis is the ability to extract aspects or features that opinions expressed on. There are many approaches and techniques used to explore these features from unstructured comments. We proposed a different approach to the above mentioned aspect extraction task in sentiment analysis using a deep learning model combining Bidirectional Gated Recurrent Unit (BiGRU) and Conditional Random Field (CRF). This model is trained on labeled data to extract and classify feature sets in comments. Our model uses a BiGRU neural network with word embeddings achieved by training GloVe on the SemEval 2014 dataset. The SemEval 2014 dataset include 7,686 reviews on two domains, Laptop and Restaurant. Experimental results showed that our aspect extraction model in sentiment analysis using BiGRU-CRF achieved significantly better accuracy than the state-of-the-art methods.**

Keywords—sentiment analysis, opinion mining, aspect extraction, feature extraction, BiGRU, CRF.

I. Introduction

The fast and diverse growth of social media like user experience reviews, forums, blogs or social network has enabled many individuals and organizations using these public views on their decision-making for a particular product or service. However, finding and aggregating these opinions are difficult tasks because of the tremendous amount of data available today and limited readability which make it difficult for users to accurately identify and synthesize information and reviews from relevant documents to the issue they are looking for. For these reasons, it is necessary to build the automated opinion synthesizing and mining system so that users can get specific opinion polarity on their interested issues easily.

The opinion mining (OM) [4] is a field of research on the ability to extract and categorize opinions of the community about entities and attributes related to these entities such as products and services…that support for tracking the mood of the community. Opinion mining is used in many various areas: businesses and organizations seek to catch the public views about their products and services; individuals want to know the opinions of their predecessors have experienced about the products or services they are interested in; the sentiments of the masses about a political figure before an election. Therefore, the need to exploit the ability to detect and understand the views automatically is increasingly interested in research. It is necessary to extract aspect of the entity for identifying the opinion or sentiment expressed on it. This task maybe the major issue in the specific tasks of sentiment analysis. The innovations of proposed aspect extraction models and the existing aspect extraction techniques are absolutely the core research directions of sentiment analysis.

In this paper, we propose to apply the deep learning approach for aspect extraction task in sentiment text using an integrated model of BiGRU and CRF. Our model uses BiGRU with word embeddings [24] obtained by training GloVe [11] on 7,686 reviews of the SemEval 2014 dataset. The reasons for our choosing GRU model are that: (i) GRU is more computational efficient than LSTM thanks to its structure: GRU has only two gates compared with three gates of LSTM. This makes the GRU's processing speed faster and simpler in its structure and thus easier to modify, for example adding new gates in case of additional input to the network; (ii) unlike CNN, a different deep learning technique that has been recently applied for aspect extraction models, GRU also can extract long semantic patterns without tuning the parameter when training the model. The experimental results show that our BiGRU 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 II presents the related works for aspect extraction task in sentiment analysis. Our proposed approach for aspect extraction with BiGRU and CRF architecture is described in Section III. Section IV presents our experiments using a benchmark dataset and received results. Finally, Section V focuses on our conclusions and discussions about future works.

II. Related Works

Most of recent researches focus on four main approaches for extracting aspect [4]: based on the frequency of nouns and noun phrases; based on relationships between features and opinions; topic modeling and supervised learning.

Hu and Liu (2004) [15] first studied aspect extraction from sentiment text using a data mining algorithm. They counted occurrence frequencies of nouns and noun phrases that were identified by a POS tagger and only the frequent ones were kept. They are likely to be aspects. Popescu and

978-1-5386-9313-1/19/$31.00 ©2019 IEEE

Etzioni (2005) [1] improved the above algorithm by detecting whether noun or noun phrase is an aspect or not. They used PMI measure to evaluate the discovered noun phrase by computing the PMI scores between the phrase and meronymy discriminators associated with the product class known in advance. Scaffidi et al. (2007) [8] considered the extracted nouns and noun phrases frequency from a review corpus with their occurrence rates in a generic English corpus to find true aspects. They assumed that product aspects are more frequent in product reviews than in a general natural language text. Long, Zhang and Zhu (2010) [7] used the frequency-based method to extract aspects. Then, 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. Hu and Liu (2004) [16] applied this method for extracting infrequent aspects. They proposed that: different aspects can be modified or depicted by the same opinion word. So that, if a sentence does not have a regular aspect but has some opinion words, the nearest noun or noun phrase to each opinion word is extracted. For example: "The laptop is great", if we know that "great" is an opinion word, "laptop" is extracted as an aspect. Kobayashi, Inui and Matsumoto (2007) [18] used a dependency parser to recognize such dependency relations of individual words for aspect extraction task. Y. Wu, Zhang, Huang and L. Wu (2009) [28] proposed to use a phrase dependency parser for noun phrases and verb phrases extraction that maybe aspects. Qiu, Liu, Bu and Chen (2011) [10] proposed opinion aspect extraction approach using domain dependent corpus so that domain-dependent opinion aspect can be found. This method utilized the syntactic relationship between opinion and aspect with a small set of opinion seed words to extract concurrent opinions and aspects. Thanks to the relationship between the opinion and its aspect, the aspect can be identified by the definite opinion and vice versa.

Using topic model for extracting aspect is one of the most popular approaches in recent year. Topic model is essentially the method of detecting topics that are aspects in the sentiment analysis context. Mei, Ling, Wondra, Su and Zhai (2007) [21] applied an aspect-sentiment mixture model based on a topic model for sentiment analysis by using probabilistic Latent Semantic Analysis (pLSA). Lin and He (2009) [6] proposed a joint topic-sentiment model by Latent Dirichlet Allocation (LDA). However, extracted aspect and opinion words from their model were not clearly separated. Lu, Zhai and Sundaresan (2009) [29] extracted aspect using structured pLSA topic model. It can model the dependency structure of phrases in short reviews. Moreover, this model can predict the rating for each extracted aspect by combining all rating of the comment and the classification result of a learned aspect classifier. Brody and Elhadad (2010) [22] applied topic model to identify aspect and then identify aspect-specific opinion words by only considering adjectives.

Extracting aspect based on supervised learning approach is also recently interested. The most decisive methods are based on sequential learning using machine learning techniques such as Hidden Markov Models (HMM) and Conditional Random Fields (CRF). Jin and Ho (2009) [25] extracted aspects and opinion phrases by learning patterns with lexicalized HMM model. Jakob and Gurevych (2010) [17] applied CRF for domain independent feature extraction tasks such as tokens, POS tags, syntactic dependencies. Li et al. (2010) [9] also used the same approach to extract both features and opinions by combining two CRF variations such as Skip-CRF and Tree-CRF. M. Zhang, Y. Zhang and Vo (2015) [14] expanded CRF model using neural network to extract both relevant aspects and opinions. This CRF replaces the original discrete aspects in the CRF by continuous word embedding mechanism and adds a neural layer between input and output nodes. Wang, Pan, Dahlmeier and Xiao (2016) [26] proposed 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 (2016) [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. Irsoy and Cardie (2014) [19] explored the application of the two-dimensional deep recurrent neural network (RNN) for the aspect extraction, which outperformed traditional shallow RNNs with the same number parameters and the previous CRF methods. Liu, Joty and Meng (2015) [20] presented a common class of the discriminant models based on RNN and word embedding. Authors have used word embedding trained from three external sources in different RNN architectures including Elman-type, Jordan-type, LSTM and the variants of them for opinion target identification task. Poria, Cambria, Gelbukh (2016) [23] applied a deep convolutional neural network (CNN) for aspect extraction. They 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. Giannakopoulos, Musat, Hossmann and Baeriswyl (2017) [3] used B-LSTM and CRF 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.

III. METHODOLOGY

Recurrent neural networks (RNNs), 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 [27]. As a result, RNNs are not sufficient for learning long-term dependencies. Based on these disadvantages of RNNs, we propose using Gated Recurrent Unit (GRU), a variant of RNN as the solution for the above mentioned extraction task.

A. Gated Recurrent Unit (GRU)

GRU [13] is one of the most widely used network models of RNN. GRU was designed to solve vanishing gradient problem through a gating mechanism. Different from LSTM, another variant of RNN with three gates, there are only two gates for GRU layer: a reset gate r and an update gate z.

Basically, these are two vectors which decide what information should be passed to the output. The most useful thing of them is that they can be trained to keep information from long ago without washing it through time or remove information which is irrelevant to the prediction. The reset gate *r* determines how to combine the new input with the previous memory, and the update gate *z* defines how much of the previous memory to keep around. GRU uses this gating mechanism to learn long-term dependencies. The function of each gate is described as followed:

$$z = \sigma(x_t U^z + s_{t\text{-}1} W^z)$$
$$r = \sigma(x_t U^r + s_{t\text{-}1} W^r) \quad (1)$$
$$h = tanh(x_t U^h + (s_{t-1} \circ r) W^h)$$
$$s_t = (1 - z) \circ h + z \circ s_{t\text{-}1}$$

z : update gate

r : reset gate

x_t : input vector

s_t : cell state

W : input weight

U : output weight

σ: sigmoid function $\in$ *[0, 1]* controls how much information can be through

B. *Bidirectional Gated Recurrent Unit (BiGRU)*

A main issue of unidirectional GRU 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 GRU (BiGRU) can solve this issue.

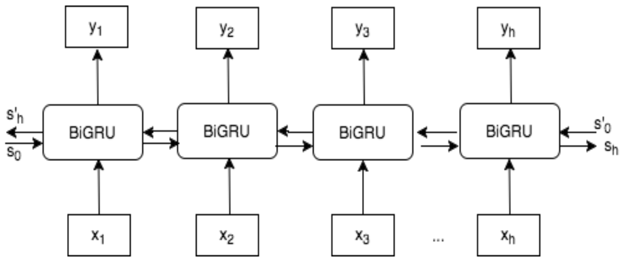

Fig. 1. Bidirectional GRU model

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

C. *Conditional Random Field (CRF)*

CRF [12] is a statistical modeling method applied in pattern recognition and machine learning and used for structured prediction. CRF is a type of discriminative undirected probabilistic graphical model and used for labeling or parsing of sequential data. CRF has got successful results in discriminative opinion mining tasks such as opinion and aspect term extraction [5][30]. However, the key for CRF's success is that feature set. How to apply the appropriate feature set and feature function expansion which often requires a lot of engineering effort for each task in hand.

D. *Proposed BiGRU-CRF model for extracting aspect*

Our network contain three layers: one first word embedding layer; one BiGRU layer composed of two components: forward GRU and backward GRU; and the final CRF layer.

Adding a CRF layer on top of BiGRU to capture dependencies can produce state-of-the-art performance. In detail, the CRF layer can add some rules of constrains to the final predicted labels, output results of BiGRU 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. We let $x_i \in R^k$ be the *k-dimensional* word vector corresponding to the *i-th* word in the sentence. A sentence of length *h* is represented as:

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

h is maximum length of the sentence. Each word x_i is represented by embedding vectors $(w_1, w_2 ... w_f)$. For word embeddings, we use pre-trained word vectors from GloVe. GloVe is an unsupervised learning technology for learning word representation. The purpose of training is to use statistical information to find similarities among words and based on co-occurrence matrix and statistical information. We use GloVe to provide 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, the first token gets the *B* label and the rest get the *I* label. For example:
I/O love/O the/O operating/B system/I

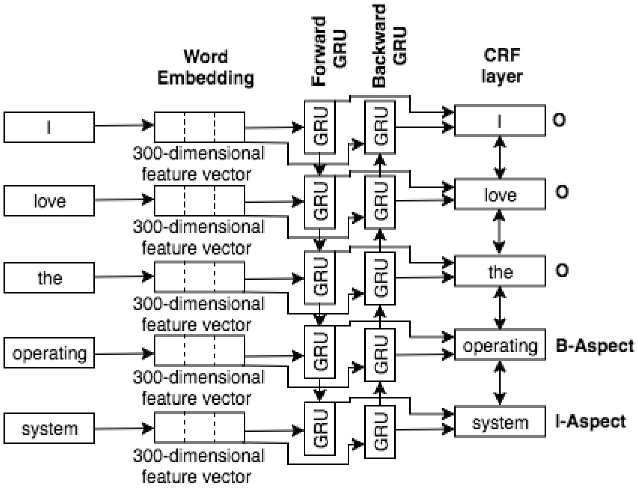

Fig. 2. BiGRU-CRF model for extracting aspect

For each token of an opinion sentence, through GloVe word embedding layer we create the 300-dimensional feature vector and fed to the BiGRU layer. Taking full advantage of the word morphology and the structure of the opinion sentence, BiGRU 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 GRU make themselves responsible for extracting features using the previous and the next tokens of each word. This combination is intended to improve the accuracy of the model.

IV. Experiments And Results

A. Dataset

We use the SemEval 2014 dataset[1] presented in TABLE I as a corpora 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 I. SemEval 2014 dataset

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

B. Tools for experiments

Our implement BiGRU-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.

C. 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 BiGRU and CRF that is presented in Fig. 2. We use 300-dimensional feature vector for BiGRU, the *adam* optimizer with learning rate 0.001, *dropout* 0.5 and a *batch size* of 32.

TABLE II. F1-score in SemEval 2014 ABSA Winners, B-LSTM & CRF and our model BiGRU & CRF

Domain	F1-score		
	SemEval 2014 ABSA winners	***B-LSTM & CRF***	***BiGRU & CRF***
Laptop	74.55%	78%	**78.5%**
Restaurant	84.01%	84%	**85%**

TABLE II shows that our proposed model outperforms the previous state-of-the-art methods, model of the winners of the SemEval 2014 ABSA contest and B-LSTM-CRF of Athanasios Giannakopoulos, Claudiu Musat, Andreea Hossmann and Michael Baeriswyl for aspect extraction task

- 3.95% and approximately 1% higher than SemEval 2014 ABSA winners in the Laptop and Restaurant domain respectively.
- 0.5% and 1% higher than B-LSTM-CRF model in the Laptop & Restaurant domain respectively.

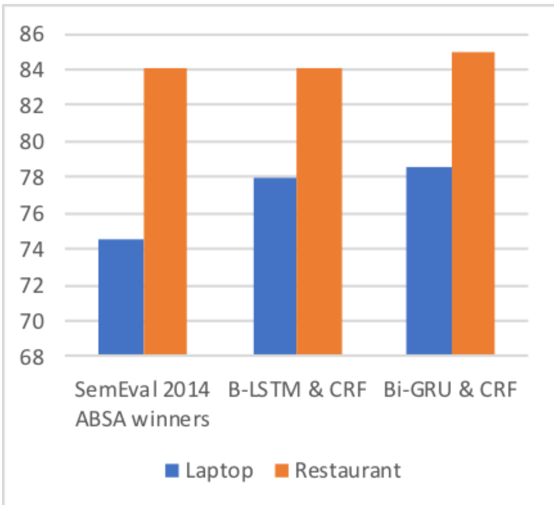

Fig. 3. Comparing the accuracy of SemEval 2014 ABSA winners, B-LSTM & CRF and BiGRU & CRF.

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

V. Conclusion

We have proposed a deep learning approach to aspect extraction task in opinion mining. Our model used a BiGRU and CRF integrated architecture that comprises the word embedding layer, GloVe, to make the feature vectors for words in opinion sentence; two GRU layers, one forward GRU and one backward GRU, 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 BiGRU framework; and one final CRF layer for supporting structure prediction and labelling aspect terms in

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

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

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 BiGRU 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 and Oren Etzioni, "Extracting product features and opinions from reviews", 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 and Claire Cardie, "Investigating LSTMs for joint extraction of Opinion entities and relations", 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 and 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 and Claire Cardie, "Extracting opinion expressions with semi-Markov conditional random fields", 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 and Yulan He, "Joint sentiment/topic model for sentiment analysis", 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 and Xiaoyan Zhu, "A review selection approach for accurate feature rating estimation", 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 and Chun Jin, "Red Opal: product-feature scoring from reviews", 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 and Hao Yu, "Structure-aware review mining and summarization", 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 and Chun Chen, "Opinion word expansion and Target extraction through Double Propagation", 2011 Association for Computational Linguistics, vol. 37, no. 1, 2011.

[11] Jeffrey Pennington, Richard Socher, and Christopher D. Manning, "GloVe: Global Vectors for Word Representation", 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 and Fernando C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", 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 and 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 and Duy Tin Vo, "Neural networks for open domain targeted sentiment", Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp.612–621, Lisbon, September 17–21, 2015.

[15] Minqing Hu and Bing Liu, "Mining opinion features in customer reviews", AAAI'04 Proceedings of the 19th national conference on Artificial intelligence, pp.755–760, July 25–29, 2004.

[16] Minqing Hu and Bing Liu, "Mining and summarizing customer reviews", KDD'04 Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.168–177, Seattle, WA, USA, August 22–25, 2004.

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

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

[19] Ozan Irsoi and Claire Cardie, "Opinion mining with deep recurrent neural networks", EMNLP'14 Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp.720–728, Doha, October 25–29, 2014.

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

[21] Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su and ChengXian Zhai, "Topic sentiment mixture: Modeling facets and Opinions in Weblogs", Proceedings of the 16th ACM International Conference on World Wide Web, pp.171–180, New York, 2007.

[22] Samuel Brody and Noemie Elhadad, "An unsupervised aspect-sentiment model for online reviews", 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.

[23] Soujanya Poria, Erik Cambria and Alexander Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network", Knowledge-Based Systems, vol. 108, pp.42–49, September 15th, 2016.

[24] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, "Efficient Estimation of Word Representations in Vector Space", arXiv:1301.3781v3 [cs.CL], September 7th, 2013.

[25] Wei Jin and Hung Hay Ho, "A novel lexicalized HMM-based learning framework for web opinion mining", ICML'09 Proceedings of the 26th Annual International Conference on Machine Learning, pp.465–472, Montreal, Quebec, June 14–18, 2009.

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

[27] Yoshua Bengio, P. Simard and Paolo Frasconi, "Learning long-term dependencies with gradient descent is difficult", IEEE Transactions on Neural Networks, vol. 5, issue 2, pp. 157–166, New York, USA, March, 1994.

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

[29] Yue Lu, ChengXian Zhai and Neel Sundaresan, "Rated aspect summarization of short comments", WWW'09 Proceedings of the 18th ACM International Conference on World Wide Web, pp.131–140, Madrid, April 20–24, 2009.

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