ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 5, May 2016**

# Target Oriented Opinion Mining Based on Online Review Model

Dhivya Bharathi. N [1], Dr.M.Sabibullah[2,],

Research Scholar, Department of Computer Science, Sudharsan College of Arts and Science, Perumanadu, Pudukkottai District, Tamil Nadu, India.

Associate Professor and Dean, Department of Computer Science, Sudharsan College of Arts and Science, Perumanadu, Pudukkottai District, Tamil Nadu, India.

**ABSTRACT:** Mining opinion targets and opinion words from online reviews are important tasks for fine-grained opinion mining, the key component of which involves detecting opinion relations among words. To this end, this paper proposes a novel approach based on the partially-supervised alignment model, which regards identifying opinion relations as an alignment process. Then, a graph-based co-ranking algorithm is exploited to estimate the confidence of each candidate. Finally, candidates with higher confidence are extracted as opinion targets or opinion words. Compared to previous methods based on the nearest-neighbor rules, our model captures opinion relations more precisely, especially for long-span relations. Compared to syntax-based methods, our word alignment model effectively alleviates the negative effects of parsing errors when dealing with informal online texts. In particular, compared to the traditional unsupervised alignment model, the proposed model obtains better precision because of the usage of partial supervision. In addition, when estimating candidate confidence, we penalize higher-degree vertices in our graph-based co-ranking algorithm to decrease the probability of error generation. Our experimental results on three corpora with different sizes and languages show that our approach effectively outperforms state-of-the-art methods.

**KEYWORDS:** Opinion mining, opinion targets extraction, opinion words extraction

## I. INTRODUCTION

With the rapid development of Web 2.0, a huge number of product reviews are springing up on the Web. From these reviews, customers can obtain first-hand assessments of product information and direct supervision of their purchase actions. Meanwhile, manufacturers can obtain immediate feedback and opportunities to improve the quality of their products in a timely fashion. Thus, mining opinions from online reviews has become an increasingly urgent activity and has attracted a great deal of attention from researchers. To extract and analyze opinions from online reviews, it is unsatisfactory to merely obtain the overall sentiment about a product. In most cases, customers expect to find finegrained sentiments about an aspect or feature of a product that is reviewed.

For example: “This phone has a colorful and big screen, but its LCD resolution is very disappointing.” Readers expect to know that the reviewer expresses a positive opinion of the phone’s screen and a negative opinion of the screen’s resolution, not just the reviewer’s overall sentiment. To fulfill this aim, both opinion targets and opinion words must be detected. First, however, it is necessary to extract and construct an opinion target list and an opinion word lexicon, both of which can provide prior knowledge that is useful for fine-grained opinion mining and both of which are the focus of this paper.

An opinion target is defined as the object about which users express their opinions, typically as nouns or noun phrases. In the above example, “screen” and “LCD resolution” are two opinion targets. Previous methods have usually generated an opinion target list from online product reviews. As a result, opinion targets usually are product features or attributes. Accordingly this subtask is also called as product feature extraction. In addition, opinion words are the words that are used to express users’ opinions. In the above example, “colorful”, “big” and “disappointing” are three opinion words. Constructing an opinion words lexicon is also important because the lexicon is beneficial for identifying opinion expressions.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

For these two subtasks, previous work generally adopted a collective extraction strategy. The intuition represented by this strategy was that in sentences, opinion words usually co-occur with opinion targets, and there are strong modification relations and associations among them (which in this paper are called opinion relations or opinion associations). Therefore, many methods jointly extracted opinion targets and opinion words in a bootstrapping manner. For example, "colorful" and "big" are usually used to modify "screen" in the cell-phone domain, and there are remarkable opinion relations among them. If we know "big" to be an opinion word, then "screen" is very likely to be an opinion target in this domain. Next, the extracted opinion target "screen" can be used to deduce that "colorful" is most likely an opinion word. Thus, the extraction is alternatively performed between opinion targets and opinion words until there is no item left to extract.

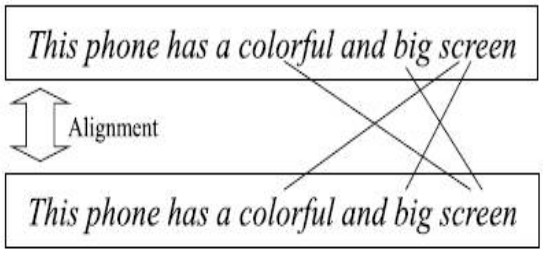


Fig. 1. Mining opinion relations between words using the word alignment model.

Although there are many variants of bootstrapping-based approaches, we notice that these methods still have some limitations as follows:

1) In previous methods, mining the opinion relations between opinion targets and opinion words was the key to collective extraction. To this end, the most adopted techniques have been nearest-neighbor rules and syntactic patterns. Nearest neighbor rules regard the nearest adjective/verb to a noun phrase in a limited window as its modifier. Clearly, this strategy cannot obtain precise results because there exist long-span modified relations and diverse opinion expressions. To address this problem, several methods exploited syntactic information, in which the opinion relations among words are decided according to their dependency relations in the parsing tree. Accordingly several heuristic syntactic patterns were designed. However, online reviews usually have informal writing styles, including grammatical errors, typographical errors, and punctuation errors. This makes the existing parsing tools, which are usually trained on formal texts such as news reports, prone to generating errors. Accordingly, these syntax-based methods, which heavily depend on parsing performance, suffer from parsing errors and often do not work well. To improve the performance of these methods, we can specially design exquisite, high-precision patterns. However, with an increase in corpus size, this strategy is likely to miss more items and has lower recall. Therefore, how to precisely detect the opinion relations among words is a considerable challenge in this task.

2) The collective extraction adopted by most previous methods was usually based on a bootstrapping framework, which has the problem of error propagation. If some errors are extracted by iteration, they would not be filtered out in subsequent iterations. As a result, more errors are accumulated iteratively. Therefore, how to alleviate, or even avoid, error propagation is another challenge in this task. To resolve these two challenges, this paper presents an alignment-based approach with graph co-ranking to collectively extract opinion targets and opinion words.
Our main contributions can be summarized as follows:

1) To precisely mine the opinion relations among words, we propose a method based on a monolingual word alignment model (WAM). An opinion target can find its corresponding modifier through word alignment. For example in Fig. 1, the opinion words "colorful" and "big" are aligned with the target word "screen". Compared to previous nearest-neighbor rules, the WAM does not constrain identifying modified relations to a limited window; therefore, it can capture more complex relations, such as long-span modified relations. Compared to syntactic patterns, the WAM is more robust because it does not need to parse informal texts. In addition, the WAM can integrate several intuitive

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

factors, such as word co-occurrence frequencies and word positions, into a unified model for indicating the opinion relations among words. Thus, we expect to obtain more precise results on opinion relation identification. The alignment model used has proved to be effective for opinion target extraction. However, for opinion word extraction, there is still no straightforward evidence to demonstrate the WAM's effectiveness.

2) We further notice that standard word alignment models are often trained in a completely unsupervised manner, which results in alignment quality that may be unsatisfactory. We certainly can improve alignment quality by using supervision. However, it is both time consuming and impractical to manually label full alignments in sentences. Thus, we further employ a partially-supervised word alignment model (PSWAM). We believe that we can easily obtain a portion of the links of the full alignment in a sentence. These can be used to constrain the alignment model and obtain better alignment results. To obtain partial alignments, we resort to syntactic parsing. Although existing syntactic parsing algorithms cannot precisely obtain the whole syntactic tree of informal sentences, some opinion relations can still be obtained precisely by using high-precision syntactic patterns. A constrained EM algorithm based on hill-climbing is then performed to determine all of the alignments in sentences, where the model will be consistent with these links as much as possible. In this way, some errors induced by completely unsupervised WAMs will be corrected. For example, in Fig. 2, "kindly" and "courteous" are incorrectly identified as modifiers for "foods" if the WAM is performed in a wholly unsupervised manner. However, by using some syntactic patterns, we can assert that "courteous" should be aligned to "services". Through the PSWAM, "kindly" and "courteous" are correctly linked to "services". This model not only inherits the advantages of the word alignment model for opinion relation identification, but it also has a more precise performance because of the use of partial supervision. Thus, it is reasonable to expect that the PSWAM is likely to yield better results compared to traditional methods for extracting opinion targets and opinion words.

3) To alleviate the problem of error propagation, we resort to graph co-ranking. Extracting opinion targets/words is regarded as a co-ranking process. Specifically, a graph, named as Opinion Relation Graph, is constructed to model all opinion target/word candidates and the opinion relations among them. A random walk based co-ranking algorithm is then proposed to estimate each candidate's confidence on the graph. In this process, we penalize high-degree vertices to weaken their impacts and decrease the probability of a random walk running into unrelated regions on the graph. Meanwhile, we calculate the prior knowledge of candidates for indicating some noises and incorporating them into our ranking algorithm to make collaborated operations on candidate confidence estimations. Finally, candidates with higher confidence than a threshold are extracted. Compared to the previous methods based on the bootstrapping strategy, opinion targets/words are no longer extracted step by step. Instead, the confidence of each candidate is estimated in a global process with graph co-ranking. Intuitively, the error propagation is effectively alleviated.

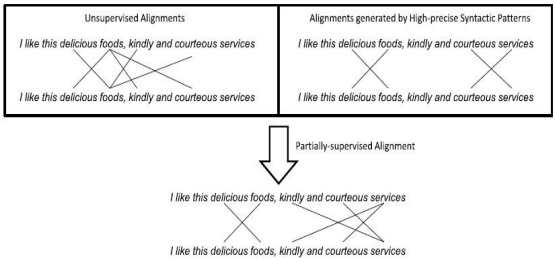


Fig. 2. Mining opinion relations between words using partially supervised alignment model.

## II. RELATED WORK

Opinion target and opinion word extraction are not new tasks in opinion mining. There is significant effort focused on these tasks. They can be divided into two categories: sentence-level extraction and corpuslevel extraction according to their extraction aims. In sentence-level extraction, the task of opinion target/ word extraction is to identify the opinion target mentions or opinion expressions in sentences. Thus, these tasks are usually regarded as sequence-labeling problems. Intuitively, contextual words are selected as the features to indicate opinion targets/words in sentences. Additionally, classical sequence labeling models are used to build the extractor, such as CRFs and HMM. Jin and Huang

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 5, May 2016**

proposed a lexicalized HMM model to perform opinion mining. Both used CRFs to extract opinion targets from reviews. However, these methods always need the labeled data to train the model. If the labeled training data are insufficient or come from the different domains than the current texts, they would have unsatisfied extraction performance.

Although proposed a method based on transfer learning to facilitate crossdomain extraction of opinion targets/words, their method still needed the labeled data from out-domains and the extraction performance heavily depended on the relevance between in-domain and out-domain. In addition, much research focused on corpus-level extraction. They did not identify the opinion target/word mentions in sentences, but aimed to extract a list of opinion targets or generate a sentiment word lexicon from texts. Most previous approaches adopted a collective unsupervised extraction framework.

As mentioned in our first section, detecting opinion relations and calculating opinion associations among words are the key component of this type of method. Wang and Wang adopted the co-occurrence frequency of opinion targets and opinion words to indicate their opinion associations. Hu and Liu exploited nearest-neighbor rules to identify opinion relations among words. Next, frequent and explicit product features were extracted using a bootstrapping process. Only the use of cooccurrence information or nearest-neighbor rules to detect opinion relations among words could not obtain precise results. Thus, exploited syntax information to extract opinion targets, and designed some syntactic patterns to capture the opinion relations among words. The experimental results showed that their method performed better than that.

Moreover, proposed a method, named as Double Propagation, that exploited syntactic relations among words to expand sentiment words and opinion targets iteratively. Their main limitation is that the patterns based on the dependency parsing tree could not cover all opinion relations. Therefore, Zhang et al. extended the work. Besides the patterns used in Zhang et al. further designed specific patterns to increase recall. Moreover, they used an HITS algorithm to compute opinion target confidences to improve precision. Liu et al. focused on opinion target extraction based on the WAM. They used a completely unsupervised WAM to capture opinion relations in sentences. Next, opinion targets were extracted in a standard random walk framework. Liu's experimental results showed that the WAM was effective for extracting opinion targets.

Nonetheless, they present no evidence to demonstrate the effectiveness of theWAMon opinion word extraction. Furthermore, a study employed topic modeling to identify implicit topics and sentiment words. The aims of these methods usually were not to extract an opinion target list or opinion word lexicon from reviews. Instead, they were to cluster for all words into corresponding aspects in reviews, which was different from the task in this paper. These methods usually adopted coarser techniques, such as frequency statistics and phrase detection, to detect the proper opinion targets/words. They put more emphasis on how to cluster these words into their corresponding topics or aspects.

## III. CONCLUSION AND FUTURE WORK

This system proposes a novel method for co-extracting opinion targets and opinion words by using a word alignment model. Our main contribution is focused on detecting opinion relations between opinion targets and opinion words. Compared to previous methods based on nearest neighbor rules and syntactic patterns, in using a word alignment model, our method captures opinion relations more precisely and therefore is more effective for opinion target and opinion word extraction. Next, we construct an Opinion Relation Graph to model all candidates and the detected opinion relations among them, along with a graph co-ranking algorithm to estimate the confidence of each candidate. The items with higher ranks are extracted out. The experimental results for three datasets with different languages and different sizes prove the effectiveness of the proposed method. In future work, we plan to consider additional types of relations between words, such as topical relations, in Opinion Relation Graph. We believe that this may be beneficial for co-extracting opinion targets and opinion words.

## REFERENCES

[1] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Seattle, WA, USA, 2004, pp. 168–177.

[2] F. Li, S. J. Pan, O. Jin, Q. Yang, and X. Zhu, "Cross-domain coextraction of sentiment and topic lexicons," in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Jeju, Korea, 2012, pp. 410–419.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

**Vol. 4, Issue 5, May 2016**

[3] L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain, "Extracting and ranking product features in opinion documents," in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 1462–1470.

[4] K. Liu, L. Xu, and J. Zhao, "Opinion target extraction using wordbased translation model," in Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn., Jeju, Korea, Jul. 2012, pp. 1346–1356.

[5] M. Hu and B. Liu, "Mining opinion features in customer reviews," in Proc. 19th Nat. Conf. Artif. Intell., San Jose, CA, USA, 2004, pp. 755–760.

[6] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process., Vancouver, BC, Canada, 2005, pp. 339–346.

[7] G. Qiu, L. Bing, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," Comput. Linguistics, vol. 37, no. 1, pp. 9–27, 2011.

[8] B. Wang and H. Wang, "Bootstrapping both product features and opinion words from chinese customer reviews with crossinducing," in Proc. 3rd Int. Joint Conf. Natural Lang. Process., Hyderabad, India, 2008, pp. 289–295.

[9] B. Liu, Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, series Data-Centric Systems and Applications. New York, NY, USA: Springer, 2007.

[10] G. Qiu, B. Liu, J. Bu, and C. Che, "Expanding domain sentiment lexicon through double propagation," in Proc. 21st Int. Jont Conf. Artif. Intell., Pasadena, CA, USA, 2009, pp. 1199–1204.

[11] R. C. Moore, "A discriminative framework for bilingual word alignment," in Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process., Vancouver, BC, Canada, 2005, pp. 81–88.

[12] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," in Proc. Conf. Web Search Web Data Mining, 2008, pp. 231–240.

[13] F. Li, C. Han, M. Huang, X. Zhu, Y. Xia, S. Zhang, and H. Yu, "Structure-aware review mining and summarization." in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 653–661.

[14] Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in Proc. Conf. Empirical Methods Natural Lang. Process., Singapore, 2009, pp. 1533–1541.

[15] T. Ma and X. Wan, "Opinion target extraction in chinese news comments." in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 782–790.

[16] Q. Zhang, Y. Wu, T. Li, M. Ogihara, J. Johnson, and X. Huang, "Mining product reviews based on shallow dependency parsing," in Proc. 32nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, Boston, MA, USA, 2009, pp. 726–727.

[17] W. Jin and H. H. Huang, "A novel lexicalized HMM-based learning framework for web opinion mining," in Proc. Int. Conf. Mach. Learn., Montreal, QC, Canada, 2009, pp. 465–472.

[18] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," J. ACM, vol. 46, no. 5, pp. 604–632, Sep. 1999.

[19] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai, "Topic sentiment mixture: modeling facets and opinions in weblogs," in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 171–180.

[20] I. Titov and R. McDonald, "A joint model of text and aspect ratings for sentiment summarization," in Proc. 46th Annu. Meeting Assoc. Comput. Linguistics, Columbus, OH, USA, 2008, pp. 308–316.

[21] W. X. Zhao, J. Jiang, H. Yan, and X. Li, "Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid," in Proc. Conf. Empirical Methods Natural Lang. Process., Cambridge, MA, USA, 2010, pp. 56–65.

[22] A. Mukherjee and B. Liu, "Modeling review comments," in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, Jeju, Korea, Jul. 2012, pp. 320–329.

[23] P. F. Brown, V. J. D. Pietra, S. A. D. Pietra, and R. L. Mercer, "The mathematics of statistical machine translation: Parameter estimation," Comput. Linguist., vol. 19, no. 2, pp. 263–311, Jun. 1993.

[24] Z. Liu, H. Wang, H. Wu, and S. Li, "Collocation extraction using monolingual word alignment method," in Proc. Conf. Empirical Methods Natural Lang. Process., Singapore, 2009, pp. 487–495.

[25] Z. Liu, X. Chen, and M. Sun, "A simple word trigger method for social tag suggestion," in Proc. Conf. Empirical Methods Natural Lang. Process., Edinburgh, U.K., 2011, pp. 1577–1588.

## BIOGRAPHY

**Dhivya Bharathi. N. (**Research Scholar**)** doing M.Phil., Computer Science, Department of Computer Science, Sudharsan College of Arts and Science, Perumanadu-622104, Pudukkottai District, Tamil Nadu, India.. Her areas of interests are as follows: Area of Interest- Data Mining, Communication System, Wireless communication and networking.