

# Co-Extracting Opinion Relations from Online Reviews Based on the Word Alignment Model

**Mrs.Nibedita Pangrahi**, Assistant professor, Department of Computer Science and Engineering,  
Vivekananda Institute of Technology, Bangalore-74.

**Chaithanya M Murthy**, Student, Department of Computer Science and Engineering,  
Vivekananda Institute of Technology, Bangalore-74.

chaithanya.murthy@gmail.com

**Chaitra B R**, Student, Department of Computer Science and Engineering,  
Vivekananda Institute of Technology, Bangalore-74

Chaitrabr69@gmail.com

**Harshitha V**, Student, Department of Computer Science and Engineering,  
Vivekananda Institute of Technology, Bangalore-74.

hvaradaraj3@gmail.com

**Neha G A**, Student, Department of Computer Science and Engineering,  
Vivekananda Institute of Technology, Bangalore-74

nehaanandbabu@gmail.com

**Abstract—** The important task of opinion mining is to mine opinion targets and opinion words from a huge number of product reviews. The one of the approach proposes a method based on partially supervised word alignment model, in which opinion relations identification is consider as an alignment process. Calculating opinion association among words is an important for constructing Co-Ranking graph; to find confidence of each candidate graph based co-ranking algorithm is used. Higher confidence candidate are extracted as opinion targets or opinion words. Prior knowledge also consider in finding confidence of candidate as being opinion target or opinion word. Previous methods are based on syntax based, compared to these methods proposed model minimizes negative effects of parsing errors. Due to use of partial supervision proposed model achieves better accuracy compared to unsupervised word alignment model. Final task is to extractive summary generation from Opinion Targets and Opinion Words with Word Alignment Model.

**Keywords—** Opinion mining, Opinion target extraction, Opinion word extraction, Text Mining

## I. Introduction

With the rapid growth, of e-commerce, a number of customers have taken interest in online shopping and more and more products are sold on the web, so that there are a huge number of product reviews are coming on the web. Online reviews are

often combined with numerical ratings provided by users for product aspects. Reading through all customer reviews is impractical, for popular items, as in some cases the number of reviews can be up to hundreds or in some cases thousands. To increase customer satisfaction, it has become a necessary for online vendors to make their customers to review or to express opinions about products. These reviews are not only provide customers useful information but also important for merchants to get immediate feedback from customers immediately. The web contains number of opinions about product, as a result the problem of opinion mining become an increasingly important activity. Opinion mining has been an important research area in NLP. Opinion mining is a type of natural language processing for tracking the mood of public about a particular product. Opinion mining, which is also called sentiment analysis, involves building a system to collect and categorize opinions about a product. Overall sentiment polarity of a product is not just satisfied. In most cases, customers expect to find fine grained sentiments about an aspect or feature of a product that is reviewed. For example:

*“This phone has a colourful and big screen, but its LCD resolution is very disappointing.”*

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. In addition, opinion words are the words that are used to express users’ opinions. In the above example,

“colourful”, “big” and “disappointing” are three opinion words.

## II. Material and Methodology

In this section, we present the main framework of our method i.e. extracting opinion targets/words as a co-ranking process. We assume that all nouns/noun phrases in sentences are opinion target candidates, and all adjectives/verbs are regarded as potential opinion words, which are widely adopted by previous methods [iv], [v], [vii], and [vii]. Each candidate will be assigned a confidence, and candidates with higher confidence than a threshold are extracted as the opinion targets or opinion words. If a word is likely to be an opinion word, the nouns/ noun phrases with which that word has a modified relation will have higher confidence as opinion targets. We can see that the confidence of a candidate (opinion target or opinion word) is collectively determined by its neighbours according to the opinion associations among them of each candidate. Existing system on opinion mining have applied various methods for extracting opinion targets and opinion words. [i]. Extracting opinion targets and opinion words using word alignment model using partially supervised word alignment model [i]:

The proposed method contains three main modules. They are pre-processing, opinion target and word extraction and opinion word classification. The overall diagram of the proposed method is shown in Fig.1. In pre-processing the given comment is processed and eliminates stop word and stemming. Extracting opinion targets/words as a co-ranking process. Assume all nouns/noun phrases in sentences are opinion target candidates, and all adjectives/verbs are regarded as potential opinion words. And then the opinion word is classified as good or bad.

**A. Pre-Processing** This is the first step of the proposed method. Several preprocessing steps are applied on the given comment to optimize it for further experimentations. The proposed model for data pre-processing is shown in Fig.2. Tokenization process splits the text of a document into sequence of tokens. The splitting points are defined using all non-letter characters. This results in tokens consisting of one single word (unigrams). The movie review data set was pruned to ignore the too frequent and too infrequent words. Absolute pruning scheme was used for the task. Length based filtration scheme was applied for reducing the generated token set.

The parameters used to filter out the tokens are the minimum length and maximum length. The parameters define the range for selecting the tokens. Stemming defines a technique that is

used to find the root or stem of a word. The filtered token set undergoes stemming to reduce the length of words until a minimum length is reached.

This resulted in reducing the different grammatical forms of a word to a single term.

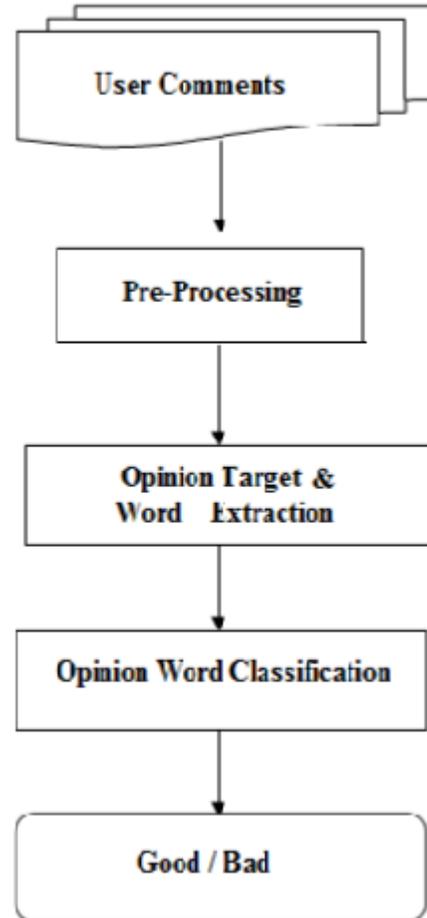


Fig. 1. Overall Diagram of the Proposed Method

## B. Opinion Targets and Opinion Words Extraction

The modified word alignment model assumes that all nouns/noun phrases in sentences are opinion target candidates, and all adjectives/verbs are regarded as potential opinion words. A noun/noun phrase can find its modifier through word alignment.

The proposed word alignment model apply a partially-supervised modified word alignment model. It performs modified word alignment in a partially supervised framework.

After that, obtain a large number of word pairs, each of which is composed of a noun/noun phrase and its modifier. And then calculate associations between opinion target candidates and opinion word candidates as the weights on the edges.

The modified word alignment model example is shown in Fig.3.

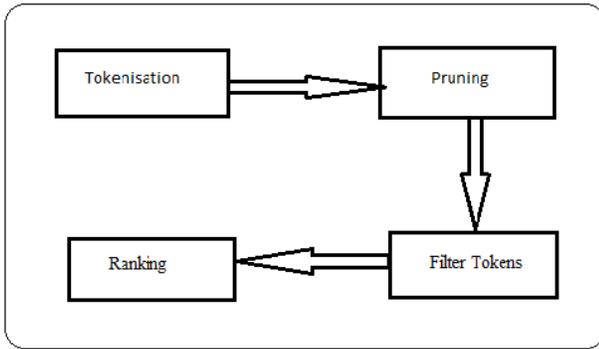


Fig. 2. Process Flow Diagram of Preprocessing

**C. Opinion Words Classification** After extraction opinion word and target the next step is to classify the opinion word. In this process the opinion word is classified as good or bad. The knn classifier is used to classify the opinion word. Once it is classified as bad then the comment is removed. In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

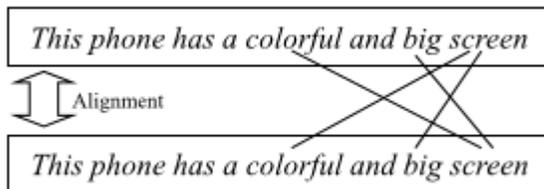


Fig.3. Opinion Relations between Words and Targets using Modified Word Alignment Model

### III. Results and Tables

Features	Screen	Touchpad	Display	Camera	Sound	Total
Review	A1	A2	A3	A4	A5	
Rv1	1	0	0	1	1	3
Rv2.	1	1	1	0	1	4
Rvn	1	0	0	1	0	2
Total features in review data set	33	21	19	15	12	

Table 1. Shows Sample Feature Extraction Matrix (FEM)

A1, A2- - - An are the aspects of the product. Rv1, Rv2- - - Rvn are the reviews about the product in Table 1. For each review, number of features is extracted and desired score get associated with the feature. Binary distribution is used in generating the feature extraction matrix (FEM). The state of the feature is set to 1 by the algorithm, if it is present in the review otherwise, it is 0. In this way, the total number of features is determined, that are associated to each product review. The FEM (feature extraction matrix) will help further in counting the number of positive and negative reviews for each feature. Also, specific number of reviews can be identified for the feature. The information can be used to determine the popular reviews and also the popular aspects of the entity, which are positively voted by the customers. The aspects can be arranged according to the popularity level. The aspects may be compared to the relative products as well.

### IV. Conclusion

This paper 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 coextracting opinion targets and opinion words.

### ACKNOWLEDGEMENT

We feel great pleasure in submitting this paper “Co-Extracting Opinion Relations from Online Reviews Based on the Word Alignment Model”. We wish to express true sense of gratitude towards our guide, **Prof, Mrs.Nibedita Panigrahi** Who at very discrete step in study of paper, contributed her valuable guidance and helped to solve every problem that arose. Our great obligation would remain due towards **Prof, Mrs.R.Chandramma**, Head of Department of Computer

Science and Engineering who was a constant inspiration during our project. Our sincere thanks to Principal **Dr. P GiridharaReddy**, proved to be a constant motivation for the

knowledge acquisition and moral support during our course curriculum

## REFERENCES

- [i] Kang Liu, Liheng Xu, and Jun Zhao, "Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 3 MARCH 2015.
- [ii] 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.
- [iii] 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
- [iv] 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.
- [v] 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
- [vi] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada, 2005, pp. 339–346
- [vii] G. Qiu, L. Bing, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," Computational Linguistics, vol. 37, no. 1, pp. 9–27, 2011.
- [viii] Z. Hai, K. Chang, J.-J. Kim, and C. C. Yang, "Identifying features in opinion mining via intrinsic and extrinsic domain relevance," IEEE Trans. Knowledge Data Eng., vol. 26, no. 3, p. 623–634, 2014.
- [ix] Z.-H. Zhou and M. Li, "Semi-supervised regression with cotraining," in Proc. 15th Int. Joint Conf. Artif. Intell., Edinburgh, Scotland, U.K
- [x] J. Zhu, H. Wang, B. K. Tsou, and M. Zhu, "Multi-aspect opinion polling from textual reviews," in Proc. 18th ACM Conf. Inf. Knowl. Manage., Hong Kong, 2009, pp. 1799–1802.