ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

An Improved Method for Opinion Target and Word Extraction using Semantic Labeling

R.Hemavathi @ Jeyalaxmi[1], B.Benita[2], Dr.D.C.Joy Winnie Wise[3], Mrs.J.Doulas[4]
[1]PG Scholar,[2] Assistant Professor,[3] Professor & Head of C.S.E.dept.,[4] Assistant Professor,
Francis Xavier Engineering College,Tirunelveli, Tamil Nadu,
[1]hemjey16@gmail.com,

Abstract*— The main objective of this paper is co-extracting opinion targets and opinion words by using a modified word alignment model. This process mainly used for bad comments removal from VIP facebook account. Nowadays facebook is one of the popular social networking used by everyone. The main purpose of the facebook is to share communicate with the whole friends at a time. But today facebook account is used to spread the bad thoughts and opinion in other facebook accounts. It may lead very serious problems and affect the people's privacy. So this paper focuses to identify the bad comments and remove that. To do this, the paper is to focus on detecting opinion relations between opinion targets and opinion words. To detect the opinion relations this project uses the word alignment model. This method captures opinion relations more precisely and therefore is more effective for opinion target and opinion word extraction. Next, 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. In this co-ranking algorithm the items with higher ranks are extracted. Finally the Semantic Labeling is used for finding the synonyms of the given word. This is also compared with bad words to block the comments. This method provides best result. To analyse the performance of the precision rate, recall rate, sensitivity, specificity and f-measure rate are used. The proposed method is compared with the several existing methods. From the comparison it is shown that the proposed method provides better performance than the existing methods.

Keywords—Opionion Word; Opionion Target; Modified Word Alignment Model; knn classifier; Preprocessing, Semantic Labeling

I. INTRODUCTION

In the current digital based economy a large amount of information is available in the form of textual data which can often be used more easily if it is categorized or classified into some predefined classes. In any business or industrial environment corporate information may be available in multiple different formats, about 80% of which is in text documents. This information exists in the form of descriptive data formats which include service reports about repair information, manufacturing quality documentation, customer help desk notes and product reviews and opinions. Social media is one of the biggest forums to express opinions. Sentiment analysis is the procedure by which information is extracted from the opinions, appraisals and emotions of people in regards to entities, events and their attributes. Sentiment analysis is also known as opinion mining. Opinion Mining is to analyse and classify the user generated data like reviews, blogs, comments, articles etc.

Opinion mining can be divided into two categories: sentence-level extraction and corpus-level extraction. Much research focused on corpus-level extraction. Hu and Liu [1] 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 co-occurrence information or nearest-neighbor rules to detect opinion relations among words could not obtain precise results. Thus, [2] 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 of [1]. Wang and Wang [4] adopted the co-occurrence frequency of opinion targets and opinion words to indicate their opinion associations. 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 [1], [4], [5] and syntactic patterns [2], [6]. From the view of sentiment analysis status, many researchers devoted their research to opinion target extraction [7-15]; they employed different methods for this research. Some researchers used the method based on rules/template to extract opinion targets. Yi et al. [16] used three progressive limited levels of part-of-speech to extract real opinion targets from candidate opinion targets. Hu and Liu [17] used association rule mining based on the Apriori algorithm to extract opinion targets; they distinguished high frequency opinion targets based on the co-occurrence of the opinion targets, and employed pruning rules to improve the accuracy and coverage. Popescu et al. [18] proposed a method by computing a Point-wise Mutual Information score of a noun, and then the Bayesian classification was employed to extract product features. Xu et al. [19] developed a method based on heuristic rules for NTCIR-8 tasks of opinion holder extraction and opinion target extraction. The core of this method is constructing rules and using pattern matching to extract opinion targets. Rule/template based method uses pattern matching to extract opinion targets, and these rules/templates established in the process are easy to understand. But it is difficult to guarantee the systematic and logic of these rules/templates, and they

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

have a higher domain-related property which is hard to be transplanted. Some researchers use natural language processing approach extracting opinion targets. Liu et al. [20] got the candidate opinion targets by the syntactic analysis result, and then employed Point-wise Mutual Information algorithm and noun pruning rules to filter the candidate opinion targets.

To detect the opinion relations this project uses the word alignment model. This method captures opinion relations more precisely and therefore is more effective for opinion target and opinion word extraction. Next, 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. In this co-ranking algorithm the items with higher ranks are extracted. And the object word is classified as good or bad. Finally the Semantic Labeling is used for finding the synonyms of the given word. This is also compared with bad words to block the comments. And then the bad comments are removed. This concept mainly used for bad comments removal from VIP facebook account.

II. PROPOSED METOD

The proposed method contains three main modules. They are pre-processing, opinion target and word extraction and opinion word classification. 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 pre-processing steps are applied on the given comment to optimize it for further experimentations.

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 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. The proposed model for data pre-processing is shown in Fig.1.

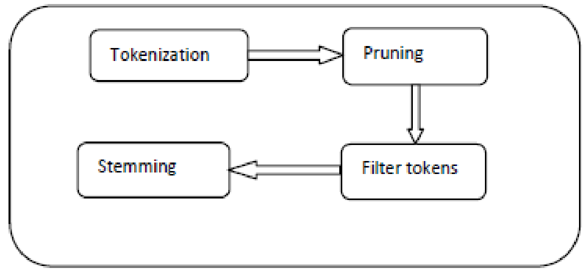


Fig. 1. Process Flow Diagram of Preprocessing

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 also applies 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.

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.

III. EXPRIMENTAL ANALYSIS

A. Performance Analysis

To evaluate the performance of the proposed method several performance metrics are available. This paper uses the Precision Rate, Recall Rate, Sensitivity, Specificity and F-Measure to analyses the performance.

1. Precision Rate

The precision is the fraction of retrieved instances that are relevant to the find.

P_1

where TP = True Positive (Equivalent with Hits)
FP = False Positive (Equivalent with False Alarm)

2. Recall Rate

The recall is the fraction of relevant instances that are retrieved according to the query.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

Where TP = True Positive (Equivalent with Hits)
FP = False Negative (Equivalent with Miss)

3. F-Measure

F-measure is the ratio of product of precision and recall to the sum of precision and recall. The f-measure can be calculated as,

$$F_m = (1 + \alpha) * \frac{Precis}{\alpha*(Preci}$$

4. Sensitivity

Sensitivity also called the true positive rate or the recall rate in some field's measures the proportion of actual positives.

Sens

where, TP – True Positive (equivalent with hit)
FN – False Negative (equivalent with miss)

5. Specificity

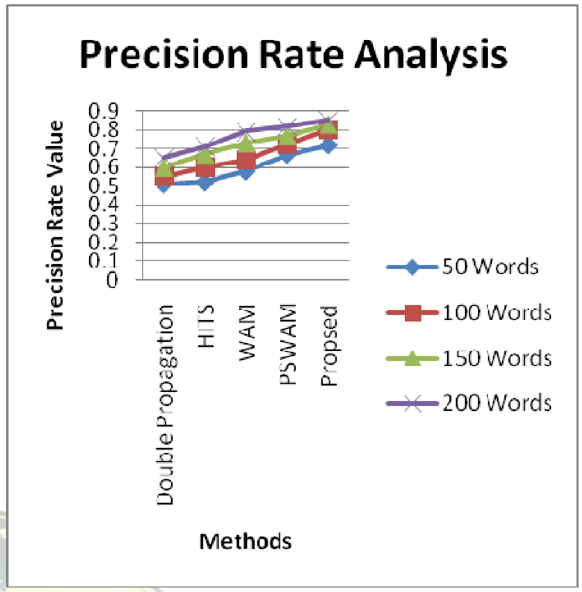
Specificity measures the proportion of negatives which are correctly identified such as the percentage.

Spec

where, TN – True Negative (equivalent with correct rejection)
FP – False Positive (equivalent with false alarm)

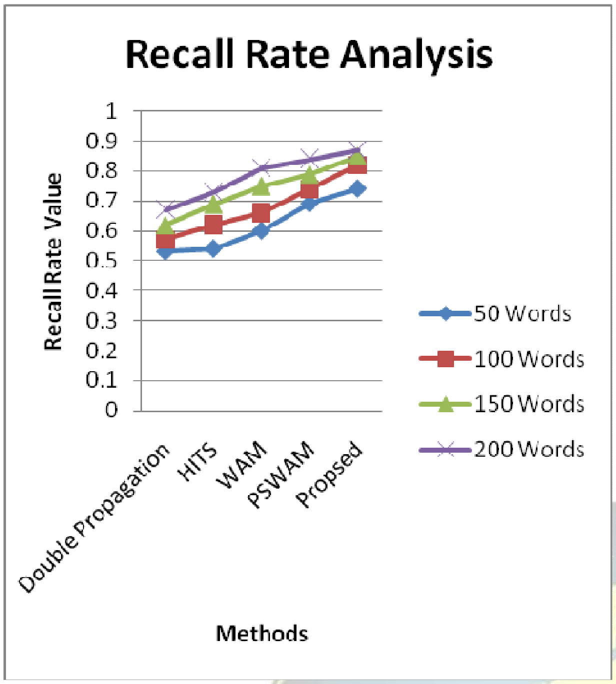
To analysis the performance of the proposed system, it is compared with various techniques by using the performance metrics which are mentioned above. This is shown in the below tables and graphs.

Methods	*50 Words*	*100 Words*	*150 Words*	*200 Words*
Double Propagation	0.51	0.55	0.60	0.65
HITS	0.52	0.60	0.67	0.71
WAM	0.58	0.64	0.73	0.79
PSWAM	0.67	0.72	0.77	0.82
Propsed	0.76	0.84	0.87	0.89

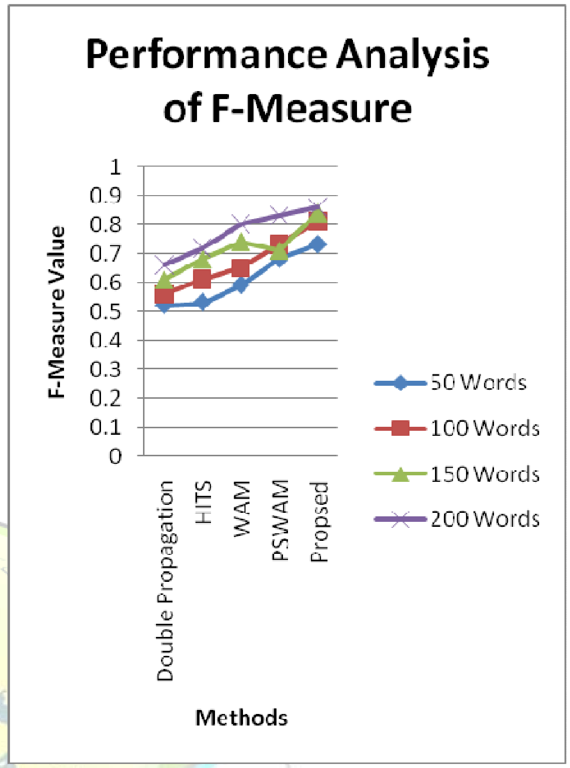


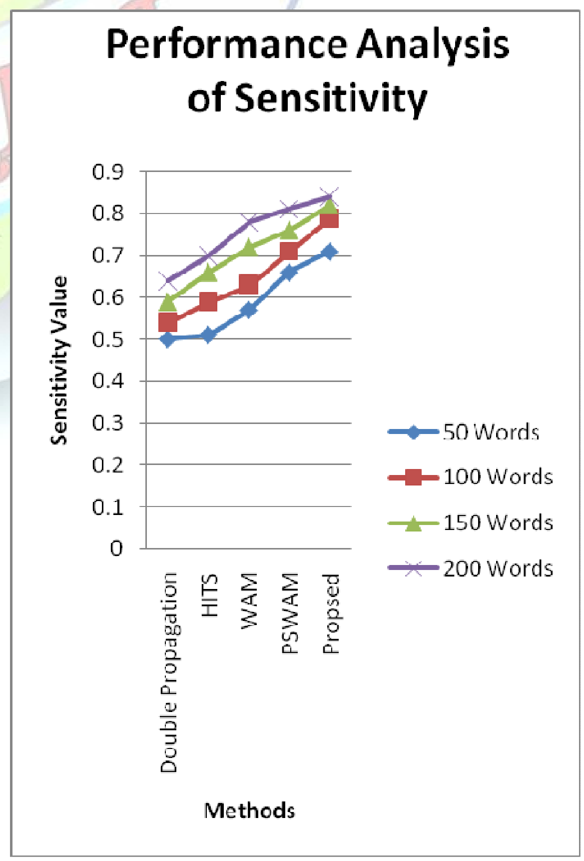
Methods	*50 Words*	*100 Words*	*150 Words*	*200 Words*
Double Propagation	0.53	0.57	0.62	0.67
HITS	0.54	0.62	0.69	0.73
WAM	0.6	0.66	0.75	0.81
PSWAM	0.69	0.74	0.79	0.84
Propsed	0.74	0.82	0.85	0.87

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016



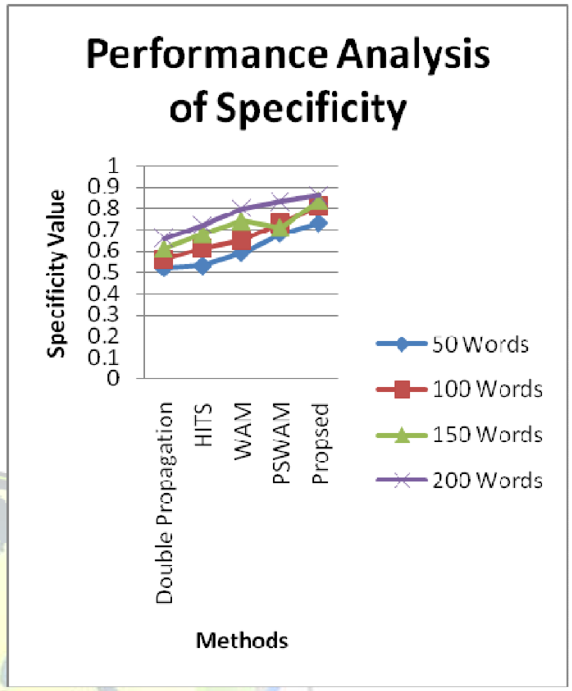
Methods	*50 Words*	*100 Words*	*150 Words*	*200 Words*
Double Propagation	0.52	0.56	0.61	0.66
HITS	0.53	0.61	0.68	0.72
WAM	0.59	0.65	0.74	0.8
PSWAM	0.68	0.73	0.71	0.83
Propsed	0.76	0.84	0.87	0.89





ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 19, April 2016

Methods	50 Words	100 Words	150 Words	200 Words
Double Propagation	0.52	0.56	0.61	0.66
HITS	0.53	0.61	0.68	0.72
WAM	0.59	0.65	0.74	0.8
PSWAM	0.68	0.73	0.71	0.83
Propsed	0.76	0.84	0.87	0.89



Methods	50 Words	100 Words	150 Words	200 Words
Double Propagation	0.5	0.54	0.59	0.64
HITS	0.51	0.59	0.66	0.7
WAM	0.57	0.63	0.72	0.78
PSWAM	0.66	0.71	0.76	0.81
Propsed	0.74	0.82	0.85	0.87

IV. CONCLUSION

This paper proposed a new algorithm for extracting opinion targets and opinion words by using a modified word alignment model. This process mainly used for bad comments removal from VIP facebook account. First the opinion relations between opinion targets and opinion words are extracted. To detect the opinion relations this project uses the word alignment model. This method captures opinion relations more precisely and therefore is more effective for opinion target and opinion word extraction. Next, 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. In this co-ranking algorithm the items with higher ranks are extracted. Finally the Semantic Labeling is used for finding the synonyms of the given word. This is also compared with bad words to block the comments. This method provides best result than the existing methods.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

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

[6] 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.

[7] H. Liu, Y. Zhao, B. Qin and T. Liu, "Comment Target Extraction and Sentiment Classification", Journal of Chinese Information Processing, vol. 24, no. 1, (2010), pp. 84-88.

[8] J. Zhu, H. Wang and M. Zhu, "Aspect-based opinion polling from customer reviews", Affective Computing, IEEE Transactions on, vol. 2, no. 1, (2011), pp. 37-49.

[9] I. Chetviorkin and N. Loukachevitch, "Extraction of Domain-specific Opinion Words for Similar Domains", Proceedings of the Workshop on Information Extraction and Knowledge Acquisition held in conjunction with RANLP (2010) October 26-30, Toronto, Canada.

[10] T. Ma and X. Wan, "Opinion target extraction in Chinese news comments", Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, (2010), pp. 782-790.

[11] F. Xianghua, L. Guo and G. Yanyan, "Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon", Knowledge-Based Systems, vol. 37, (2013), pp. 186-195.

[12] A. Duric and F. Song, "Feature selection for sentiment analysis based on content and syntax models", Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, Association for Computational Linguistics, (2011), pp. 96-103.

[13] C. Quan and F. Ren, "Unsupervised product feature extraction for feature-oriented opinion determination", Information Sciences, vol. 272, (2014), pp. 16-28.

[14] L. Chen, L. Qi and F. Wang, "Comparison of feature-level learning methods for mining online consumer reviews", Expert Systems with Applications, vol. 39, no. 10, (2012), pp. 9588-9601.

[15] F. He, Y. He, N. Liu, J. b. Liu and M. Peng, "A Micro-blogging Short Text Oriented Multi-class Feature Extraction Method of Fine-grained Sentiment Analysis", Acta Scientiarum Naturalium Universitatis Pekinensis, vol. 1 (2014), pp. 48-54.