

# Review on Opinion Targets and Opinion Words Extraction Techniques from Online Reviews

Mr. Prashant M. Jayle<sup>1</sup>, Prof. Sneha U. Bohra<sup>2</sup>

<sup>1</sup>Student of Master of Engineering in (CSE), G.H. Raisoni college of Engineering and Management, Amravati, India

<sup>2</sup>Assistant professor Department of (CSE), G.H. Raisoni College of Engineering and Management, Amravati, India

-----\*\*\*-----

**Abstract:-** The purchase and sales of huge number of product by people by seating at their place is growing on increasing day-by-day. This activity is performed successfully through the detail summary of customer's feedback and product review who have used it earlier. To analyze this kind of data, opinion mining as technique of data mining is mainly used as an important parameter. As there are thousands of reviews generated for particular product, it is necessary to perform some effective technique that will give the fine grained output on the feedback and reviews provided by the users. For this the necessary task is to identify perfect opinion words and opinion targets. After detecting these opinion words and identifying opinion targets of relation between them it is possible to analyze the comment in both positive and negative way. This kind of work is done by different researchers for performing the opinion mining from online reviews. This paper performs the study of different technique used for opinion mining and sentiment analysis. There are different techniques as supervised and unsupervised word alignment models, nearest-neighbor identification, etc. suggested by different authors used for proper identification of products reviews are briefly explain here. After studying different techniques, the most suitable techniques are identified that performs efficient mining in large amount of database.

**Keywords:** Data mining, Opinion targets, opinion words, opinion mining, Sentiment analysis, Natural Language Processing (NLP).

## I. INTRODUCTION

Now a days, most of our daily activities are perform online. As like many of the banking transactions, all type of ticket booking, buying and selling of products, etc. With this, purchase of huge number of product by people by seating at their place is growing on increasing day-by-day. But how they come to know about that product's usefulness and performance, this is done by thousands of reviews that are springing up on the Web. From these reviews, customers can obtain first-hand use of product information and direct

supervision of their purchase actions with the help users who have already used it. These reviews, are helpful not only to the users but also for the manufacturers, as they can obtain immediate feedback that lead to the opportunities for improving the quality of their products in a timely manner [1]. Thus, mining opinions from online reviews has become an increasingly urgent and important activity that has attracted a great deal of attention from researchers.

To increase customer satisfaction, it has become a necessary for online vendors to make their customers to review or to express opinions about products. Now, as there are thousands of reviews are present about particular product then, reading through all customer reviews is impractical for both customers and manufacturers. The web contains different kinds of opinions about product, as a result the problem of opinion mining from their reviews become an increasingly important activity. Opinion mining has been identified as an important research area in the field Natural Language Processing (NLP) [2]. Opinion mining is a type of natural language processing for tracking the mood of public from the reviews they provided 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 always. In most cases, customers expect to find fine grained sentiments about an aspect or feature of a product that is reviewed.

Generally, data mining is the technique used for searching from hidden patterns present in huge amount of data. Data mining scans via a huge volume of data to find out user intended patterns and correlations between several patterns. This requires the use of data analysis tool to determine previously unknown, valid patterns and relationships from the data. As online reviews are present in counts of thousands and from this much number finding out positive (good) reviews and negative (bad) reviews i.e analyzing opinions of the people requires some effective techniques. Thus, data mining technique is the way of getting analysis and prediction results more than gathering and running data. Opinion mining is one of the important factor in the domain of data mining and sentiment analysis. The

opinion mining is used to analyze the people's opinions, emotions, assessments and attitudes about something. Along with the explosive growth of user created reviews, messages, web sites and social networks it has become a significant media for where millions of users can communicate their opinions [3]. And by applying proper and valid techniques and algorithms of data mining we have to extract proper concluding remarks that becomes useful for both customers and manufacturers.

The opinion targets and opinion words is understand by following example:

"This tab has a colorful and big screen than normal smartphone,

But, its LCD resolution is very disappointing."

In the field of opinion mining, it is important to mine opinion targets, which is described as the objects or classes on which customers have articulated their opinions, classically as nouns, adjectives or phrases. In the above example, "screen" and "LCD resolution" are two opinion targets. This kind of list can be generated as an opinion target list from lots of online product reviews. 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. As a result, opinion targets are most of the time product features or attributes and opinion words are expressions or adjectives that users mainly gives in positive or negative manner about products target.

To mine and examine opinions from online reviews, it is unacceptable to simply attain the overall sentiment about a product. In many scenarios, users supposed to discover fine grained opinions about a characteristic or feature of manufactured goods that is examined. For this, one of the important technique that is identified is the word alignment model along with different approaches and techniques are used to investigate the number of document reviews more significantly. It usually used to estimate the opinion targets and opinion words [4]. Finally, candidates with higher confidence are extracted as opinion targets or opinion words. In further sections of this paper, all necessary attributes for opinion mining and sentiment analysis are studied in section II. Section III, shows different levels of opinion mining. In section IV, different techniques used by different authors for this opinion mining and sentiment analysis are studied and briefly explain as Literature review. Finally, in section V we conclude the paper.

## II. METHODOLOGIES USED

### A. Data Mining

Data mining is an interdisciplinary subfield of computer science. It is the computational process of discovering patterns in large data sets. It involves methods at the intersection of artificial intelligence, machine learning,

statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into a user interested and understandable structure that can be used further. This fits for the case of extracting opinions from thousands of reviews. Data mining not only involves the raw analysis step but also, it involves some data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and updating [1]. Data mining mainly performs the analysis step called as "knowledge discovery from databases (KDD)" process.

### B. Sentiments

Sentiment can refer to activity of five material senses that is hearing, sight, touch, smell, and taste associating them with or as something considered transcendental. It can also be defined as [6],

An attitude, thought, or judgment prompted by feeling something and/or by performing predilection.

A specific view or notion: opinion, Exhibition or manifestation of feeling or sensibility, or appeal to the tender emotions, in literature, art, or music.

A romantic or nostalgic feeling verging on sentimentality Refined feeling, delicate sensibility especially as expressed in a work of arts

### C. Opinion mining / Sentiment analysis

Opinion mining also called as Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information from the source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service. These sentiment analysis involves building a system to collect and categorize opinions about a product. Automated opinion mining that mainly has to use machine learning, a type of artificial intelligence (AI), to mine text for sentiment [7]. Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or it can have overall contextual polarity of complete document. The attitude may be someone's judgment or evaluation, affective state or the intended emotional communication.

### D. Natural Language Processing (NLP)

Natural language processing is a field of computer science, artificial intelligence, and computational linguistics that concerned with the interactions between computers and human being i.e. (natural) languages. Mainly this NLP is related to the area of human-computer interaction.

NLP is a way for computers to analyze, understand and derive meaning from human language in some smart and useful way [8]. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation. Some researchers used NLP in opinion mining from online reviews that helps customers in purchasing that products and manufacturers to make necessary changes in that products.

### E. Word Alignment Model

This model develops opinion relation recognition as a word alignment process. It uses the word-based alignment model to execute monolingual word alignment. This is extensively used in several tasks such as collocation extraction of user's opinions. In specific, if it to directly concerns the typical alignment model to this task, an opinion target candidate (noun or noun phrase) will arrange with the inappropriate words rather than potential opinion words (adjectives or verbs), such as prepositions and conjunctions [9]. Therefore, the scenario introduced some restrictions in the alignment model as follows:

- 1) Nouns or noun phrases (adjectives/verbs) should be arranged with adjectives/verbs or a null word. Arranging to a null word implies that this word either has no modifier or modifies nothing;
- 2) Other dissimilar words, such as prepositions, conjunctions and adverbs, could only support with themselves.

### F. Partially-Supervised Word Alignment Model

The typical word alignment model is classically trained in a completely unsupervised manner, which may not obtain precise alignment results. Thus, to progress alignment process, the algorithm execute a partial supervision on the statistic model and utilize a partially- supervised alignment model to integrate partial alignment links into the alignment process. In this research, the partial alignment links are considered as conditions for the trained alignment model.

## III. LEVELS OF OPINION MINING

There are mainly two approaches by using which we can perform the task of Opinion mining: Feature based opinion mining and Sentiment classification. In case of sentiment classification, the sentiment or opinion of the user is classified as positive or negative and then it will summarized to generate an opinion. But in case of feature based opinion mining, the features are selected from reviews and the sentiment is extracted after feature selection. There

are mainly three levels of opinion mining; Document level, Sentence level and Entity or Aspect Level. These different levels of analysis are explained briefly as follows [10]:

### 1. Document Level:

As he name suggest the task at document-level analysis is to classify whether a whole document is expressing a positive or negative opinion. This level of opinion mining assumes that each document gives opinion about a single product or single topic.

### 2. Sentence Level:

In this type of analysis, the task is to identify each sentence and judge that weather it gives a positive, negative or neutral sentiment. Here Neutral is usually means this sentence does not give any opinion. This level of analysis is related to subjective classification, which distinguishes objective sentences that expresses factual information from subjective sentences.

### 3. Entity and Aspect Level:

Aspect level analysis performs fine-grain analysis. Aspect level is earlier called feature level. Instead of looking at language constructs that is documents, sentences, paragraphs, etc., aspect level directly look for opinion. It is based on idea that an opinion consists of a sentiment i.e. positive or negative and a target of opinion.

## IV. LITERATURE REVIEW

There are lots of studies done by different authors that are paying attention on the task of opinion target and opinion word extraction. General textual inspection uses simple part of speech (POS) information that we are using on daily basis in our conversations For example, nouns, adjectives, adverbs, and verbs are the basic form of word-sense description. Some adjectives are good indicators of emotion and guide feature assortment that categorize the sentiment. By performing sentiment analysis and Also, selected phrases elected by pre-specified POS patterns, usually including an adjective or adverb, help detect sentiments. In this section, techniques and methods used by different researchers for opinion mining and sentiment analysis are describe briefly.

The authors Kang Liu, Liheng Xu, and Jun Zhao [1] find out here that, in previous methods of mining the opinion relations between opinion targets and opinion words seems to be a key to collective extraction. To this end, the most adopted techniques have been nearest-neighbor rules and syntactic patterns. Nearest neighbor technique regards the nearest adjective/verb to a noun/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. For addressing this problem, several methods exploited syntactic information, in which the opinion relations among words are decided according to their dependency in the parsing tree. And based on this, several heuristic syntactic patterns were designed. Another important is that, online reviews usually have informal writing styles, that may have grammatical, typographical and punctuation errors. This makes the existing parsing tools, which are usually trained on formal texts such as news reports that leads to generating errors. In the same way, these syntax-based methods, which heavily depend on parsing performance, suffer from parsing errors and often do not work well.

The authors Hao Wang, Chen Zhang, Hongzhi Yin, Fanjiang Xu [11] in this paper, proposes a novel joint framework for opinion words and targets co-extraction. Unlike the existing propagation framework and pipeline framework, their framework combines propagation and refinement in a dynamic and iterative process. In the propagation process, generalized syntactic patterns can discover more potential opinion words and targets, thereby improving the extraction recall of the results. In the refinement process, an Opinion Relation Graph (ORG) model is construct to evaluate all the candidates including opinion words, opinion targets and syntactic patterns. In future they have to focus on improving the extraction accuracy by working on the refinement methods and to reduce the false results to increase the recall which is a side effect of rule refinement. Moreover the ORG model they have discover needs to be improve for more potential relations between opinion words and targets.

The authors DONGYU LI, GUANG CHEN, YAN LI, WEIRAN XU [12] in this paper, proposed a novel system for opinion target extraction. Without syntax parser result, the system got high precision and recall rate. The method is robust and can be easily applied to other dataset with different theme. The study perform here is mainly on Chinese data. For Chinese reviews, they have proposed a module which can effectively find out the new word, and the important Chinese New Word detection will significantly improve the result. The word alignment is based on word frequency and other information, so we introduce the semantic method into the corpus and get better performance. The Machine Translation model is also widely used in other fields in Industry.

The authors Pang and Lee [13], presented survey on sentiment analysis and opinion mining. In this survey they have explained opinion oriented information right of entry, challenges, opinion categorization and their summarization. Many researchers used machine learning methods for emotion examination. The techniques explained by the authors consists of guidance of classifier on the datasets and

that uses the skilled model for classification of new documents. There are some another techniques that uses optional methods such as dictionary of word lexicons.

The authors Hui Song, Jianfeng Chu, Yun Hu, Xiaoqiang Liu [14], in this paper design a system to extract opinion targets and opinion words from online comments and propose a novel approach for the extraction of subjective from implicit sentences. In order to extract it, they use the Wikipedia to build the associations between words and then use the cosine theorem to compute the words' similarity degree. After all they can get the synonyms and use them to represent the implicit sentence's real subjective word. The experiment results demonstrate that this method is effective. The limitations of this approach is, the result is not definitely high because of some limitations. One bottleneck is that Chinese parsing results is far from satisfactory. And some comments don't obey the grammatical rules and popular network vocabulary is also used in the comments. So one still has to do improvement in this approach for handling lots of issues occurring.

The authors G. Qiu, L. Bing, J. Bu, and C. Chen [15] proposed a Double Propagation method to expand sentiment words and opinion targets iteratively, where they also exploited syntactic relations between words. The main limitation of Qiu's method is that, the patterns based on dependency parsing tree may introduce many noises for the large corpora. So for that extended Qiu's method, besides the patterns used in old version is used. In this method authors adopted some other special designed patterns to increase recall.

The authors Jaysri Vilas Borole, Nilesh Vani [16] observed that previous studies focused on opinion target extraction, which can be divided into two main categories: supervised and unsupervised methods. In supervised approaches, the opinion target extraction task was regarded as sequence labeling task. The main limitation of these methods is that labeling training data for each domain is time consuming and many times impracticable. In another part, mainly is unsupervised methods, similar to ours, most approaches regarded opinion words as the important indicators for opinion targets. An association rule mining algorithm and frequency information is used here to extract frequent explicit product features in a bootstrapping process.

The authors Jin and Ho and Li *et al.* [17] both proposed to use supervised sequential labeling methods for topic and opinion extraction. Results showed that the supervised learning methods can achieve state-of-the-art performance on lexicon extraction. However, the limitations occurs as these methods need to manually annotate a lot of training data in each



domain. Recently, Qiu *et al.* proposed a rule-based semi-supervised learning methods for lexicon extraction. However, their method requires to manually define some *general* syntactic rules among sentiment and topic words. In addition, it still requires some annotated words in the target domain to be identified.

The authors P.Bharathi, PCD.Kalaivaani [18] in this paper, proposed a sentiment classifier to overcome the problem of domain dependency by using an automatically extracted sentiment sensitive thesaurus. The data used here is the labeled data from multiple source domains and unlabeled data from a source and target domains to compute the relatedness of features and construct a sentiment sensitive thesaurus to solve feature mismatch problem. This is a hybrid model that is created based on weakly supervised learning techniques. So it is portable to all other domains. It produces good performance results which demonstrate the flexibility of hybrid model for sentiment analysis task. But there is some work is needs to be done that overcomes the limitations of this approach. One is, after identifying document sentiment one is also able to identify the semantic orientation of specific components of the review which will help to improve the mining results. Another limitation is incremental learning of hybrid model parameters should get possible as this purpose has to face with new data.

The authors Hui Zhang, Qiyun Zhao, Hao Wang, Chen Zhang, Fanjiang Xu [19] in this paper, proposes a novel dynamic propagation and refinement framework for opinion words and targets co-extraction. The methods proposed by authors is unlike the existing propagation framework and pipeline framework, as this methods combines propagation and refinement in an iterative process. The authors constructed a Opinion Relation Graph (ORG) for evaluating dependency patterns and the relations between opinions. The authors for pruning the false results adopted an automatic rule refinement to and update the rules for extraction for improving the performance. The experimental results indicate that this method achieves better performance over current state-of-the-art unsupervised methods.

## V. CONCLUSION

As we have seen the purchase and sales of huge number of product online by people is growing on increasing. And this activity is performed successfully by reading product reviews given by customers who have used it earlier. So it becomes necessary to develop some techniques and methods for analyzing customer's reviews as they are present in thousands. There are different terms and methods like opinion mining, sentiment analysis, word

alignment model, etc. associated with this concept, this paper performs review on most of these important terms. Along with important terms, this paper performs the review on work done by different researchers earlier on this concept. There are different levels of opinion mining and according to our requirement opinions are generated. There are different techniques as supervised and unsupervised word alignment models, nearest-neighbor identification, etc. suggested by different researchers used for proper identification of products reviews are briefly explain here. All the techniques have some of their advantages and drawback, but as the requirement of the opinion mining is of much important, so it is necessary to develop and make improvements in techniques for opinion mining and sentiment analysis task.

## REFERENCES

- [1] 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.
- [2] Mrs.Nibedita Pangrahi, Chaithanya M Murthy, "Co-Extracting Opinion Relations from Online Reviews Based on the WordAlignment Model", International Journal of Engineering Research ISSN: 2319-6890(online),2347-5013(print), Volume No.5 Issue: Special 4, pp: 790-991, 20 May 2016.
- [3] Manju, S. Revathi, E. V. R. M. Kalaimani, R. Bhavani. Product Aspect Ranking Using Semantic Oriented Sentiment Classifier. International Journal of Scientific Engineering and Research (IJSER). 2014; 2(10), 25-28.
- [4] Liu, Kang, LihengXu, Jun Zhao. Extracting opinion targets and opinion words from online reviews with graph coranking, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. 2014.
- [5] *Data Mining Curriculum". ACM SIGKDD. 2006-04-30. Retrieved 2014-01-27.*
- [6]Sentiment [Online] available: <http://www.dictionary.com/browse/sentiment>
- [7]Opinion mining (sentiment mining)[Online] available: <http://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining>
- [8]Introduction to Natural Language Processing (NLP) 2016 [Online] available: [http://blog.algorithmia.com/introduction-natural-language-processing-nlp/\(11 August 2016\).](http://blog.algorithmia.com/introduction-natural-language-processing-nlp/(11 August 2016).)
- [9] J. Yesudoss, T. Banusankari, "An efficient word alignment model for co-extracting opinion targets and opinion words from online reviews", *Indian Journal of Innovations and Developments*, Vol 4 (7), November, 2015.

- [10]. Vibhuti Patel, Mital Panchal, "A survey on Opinion Mining Methods from Online Reviews", International Journal of Scientific Research in Science, Engineering and technology, In December, 2015.
- [11] Hao Wang, Chen Zhang, Hongzhi Yin, Wei Wang, Jun Zhang, Fanjiang Xu, "A Unified Framework for Fine-Grained Opinion Mining from Online Reviews", *49th Hawaii International Conference on System Sciences*, 1530-1605/16 \$31.00 © 2016 IEEE.
- [12] DONGYU LI, GUANG CHEN, YAN LI, WEIRAN XU, "MINING CONSUMER'S OPINION TARGET BASED ON TRANSLATION MODEL AND WORD REPRESENTATION", *Beijing University of Posts and Telecommunications*, Beijing, 100876, 978-1-4799-7208-1/14/\$31.00 ©2014 IEEE.
- [13] Bo Pang, Lillian Lee, "Opinion Mining and Sentiment Analysis", *Foundations and Trends in Information Retrieval* Vol. 2, Nos. 1-2 (2008).
- [14] Hui Song, Jianfeng Chu, Yun Hu, Xiaoqiang Liu, "Semantic Analysis and Implicit Target Extraction of Comments from E-commerce Websites", 978-1-4799-2883-5/13 © 2013 IEEE.
- [15] 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.
- [16] Jayshri Vilas Borole, Nilesh Vani, "Extraction of Hidden Opinion Based On Sentiment Analysis Using Word Alignment Model: A Review", *International Journal of Computer Science and Information Technologies (IJCSIT)*, Vol. 7 (1), 2016, 422-426.
- [17] Fangtao Li, Sinno Jialin Pan, Ou Jin, Qiang Yang and Xiaoyan Zhu, "Cross-Domain Co-Extraction of Sentiment and Topic Lexicons", *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 410-419, Jeju, Republic of Korea, 8-14 July 2012.
- [18] P.Bharathi, PCD.Kalaivaani, "Sentiment Classification Using Weakly Supervised Learning Techniques", *ICICES2014 - S.A. Engineering College, Chennai, Tamil Nadu, India*, ISBN No.978-1-4799-3834-6/14/\$31.00©2014 IEEE.
- [19] Hui Zhang, Qiyun Zhao, Hao Wang, Chen Zhang, Fanjiang Xu, "Mining Opinion Words and Targets from Online Reviews in a hybrid Framework", *International Conference on Web Intelligence and Intelligent Agent Technology*, 2015 IEEE/WIC/ACM, 978-1-4673-9618-9/15 \$31.00 © 2015 IEEE.