Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model

Prof. Satish Bhojannawar[1], Madhuri Bandarkavate[2], Kavita Kokatnur[3], Sandya Nandagave[4], Savakka Magadum[5]
Assistant Professor[1], Student[2, 3, 4, 5]
Department of CSE
KLE College of Engineering and Technology Chikodi, Karnataka, India

Abstract:
Any organization needs to conduct surveys and collect reviews, in order to improve their product quality. There are number of websites which deals with product reviews. All these reviews are nothings but the opinions of people all over the world about different products. These reviews are very huge and difficult to analyze. Opinion mining is a discipline which deals with analysis of such reviews. 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. Partially-supervised alignment model is regards to identifying opinion relations as an alignment process. Then, a graph-based co-ranking algorithm is exploited to estimate the confidence of each candidate.

Index Terms: Opinion mining, opinion targets extraction, opinion words extraction.

I. INTRODUCTION

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 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." 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. 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. Opinion words are the words that are used to express users' opinions. In the above example, "colorful", "big" and "disappointing" are three opinion words.

II. LITERATURE SURVEY

[1] 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 corpus level 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. Intuitively, contextual words are selected as the features to indicate opinion targets/words in sentences. Jin and Huang proposed a lexicalized HMM model to perform opinion mining. If the labeled training data are insufficient or come from the different domains than the current texts, they would have unsatisfied extraction performance. The co-occurrence frequency of opinion targets and opinion words to indicate their opinion

associations. 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, exploited syntax information to extract opinion targets, and designed some syntactic patterns to capture the opinion relations among words [2]. Early works on opinion mining focused on the polarity of opinion, positive or negative; this Kind of opinion mining is called sentiment analysis. Another type of opinion mining focused on finding the detailed information of a product from reviews; this approach is a kind of information extraction. Recent research has focused on assessing the review quality before mining the opinion. Explored the use of some semantic features for review helpfulness ranking [3].

III. EXISTING APPROACH

1) 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/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. 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.

IV. PROPOSED APPROACH

Mining the opinion relations among words, is based on a monolingual word alignment model (WAM). An opinion target can find its corresponding modifier through word alignment.

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 factors, such as word co-occurrence frequencies and word positions, into a unified model for indicating the opinion relations among words. 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. 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. To alleviate the problem of error propagation, 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. 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.

VI. METHODOLOGY

Data Collection

To determine the polarity of the sentences, based on aspects, large numbers of reviews are collected from the Web. There are lots of websites on the Internet where the large numbers of customer reviews are available. Amazon website (www.amazon.com), Face Book, Twitter and other online websites is used to collect the reviews.

POS Tagging

After collecting the reviews, they are sent to the POS tagging module where POS tagger tags all the words of the sentences to their appropriate part of speech tag. POS tagging is and important phase of opinion mining, it is necessary to determine the features and opinion words from the reviews.POS tagging can be done manually or with the help of POS tagger. Manual POS tagging of the reviews take lots of time. Here, POS tagger is used to tag all the words of reviews.
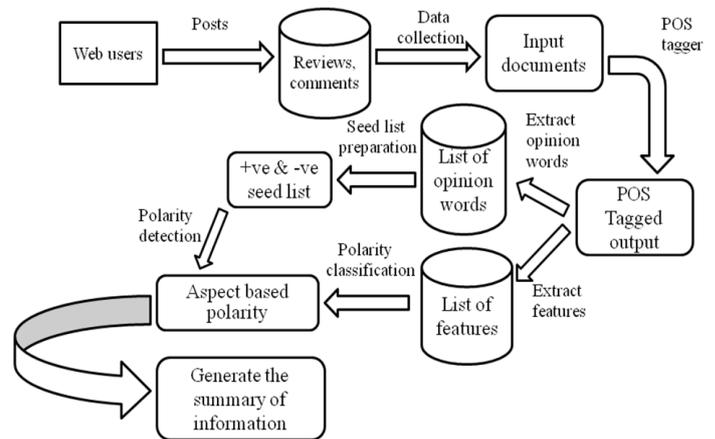

Figure.1. Architecture Diagram

Feature Extraction
All the features are extracted from the reviews and stored in a database then its corresponding opinion words are extracted from these reviews.

Extracting Opinion Words & Seed List Preparation

Initially some of the common opinion words along with their polarity are stored in the seed list. All the opinion words are extracted from the tagged output. The extracted opinion words matched with the words stored in seed list .If the word is not found in the seed list then the synonyms are determined with the help of WordNet. Each synonym is matched with words in the seed list, if any synonym matched then extracted opinion word is stored with the same polarity in the seed list. If none of the synonym is matched then the antonym is determined form the WordNet and the same process is repeated, if any antonym matched then extract opinion word is stored with the opposite polarity in the seed list. In this way the seed list keep on increasing. It grows every time whenever the synonyms or antonyms words are found in WordNet matches with seed list.

Polarity Detection & Classification

With the help of seed list, the polarity of the sentences is determined for each feature. Polarity is determined on the basis of majority of opinion words, if the number of positive words is more, then the polarity of the sentence is positive otherwise the polarity is negative and if the number of positive and negative words is equal then the sentence shows the neutral polarity. Negation is also handled in the system, if the opinion word is preceded by negation "not" then polarity of that sentence is reversed. For example, the sentence *"The touch screen of this mobile phone is not good"* shows the negative polarity because the opinion word '**good**' is preceded by '**not**'. In the end, a features wise summary is generated. Figure shows how the summary is generated by the System.

VII. CONCLUSION

Detecting opinion relations between opinion targets and opinion words. Compared to previous methods based on nearest neighbour rules and syntactic patterns, in using a 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. The items with higher ranks are extracted out.

VIII. REFERENCES

[1]. Kang Liu, LihengXu, 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]. RichaSharma, Shweta Nigam and Rekha Jain: "**Mining of Product Reviews at Aspect Level**" International Journal in Foundations of Computer Science & Technology (IJFCST), Vol.4, No.3, May 2014.

[3]. DhanashreeKulkarni, Prof. S.F.Rodd: "**A Survey on Oproblem and levels of Analysis**"International Journal of Innovative Research in Science, Engineering and Technology Vol. 4, Issue 12, December 2015.

[4]. Padmapani P. Tribhuvan et al: "**A Peer Review of Feature Based Opinion Mining and Summarization**" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, 247-250.

[5]. SurabhiThorat: "Opinion Mining and Sentiment Analysis- Its Tools and Challenges "International Journal of IT, Engineering and Applied Sciences Research.