Aspect Based Opinion Mining For Identifying Customer Preferences

Ms. Aarati Mahadik\textsuperscript{1} and Mrs. Asha Bharambe\textsuperscript{2}
\textsuperscript{1, 2}Dept. of I.T., V.E.S. Institute of Technology, Mumbai India

Abstract- In today's digital world almost all work can be done with internet, as use of web is rapidly increase the data generated by the users is in huge amount in the form of customers reviews, blogs, social networking etc. In our system we propose feature/aspect wise opinion mining for feature/aspect ranking with considering product reviews. Generally products have number of feature/aspects. Some features/aspect may have more importance or some may have less than other as per different opinion from different users. Ranking features/aspect will help firms and users to know more easily about particular product feature/aspect pros and cons. Proposed system used Senti-WordNet for calculating opinion values of aspects with C4.5 classifier to improve the accuracy of opinion polarity. To get more accuracy for feature/aspect summarization semantically similar word approach is used in proposed system.

Keywords-Opinion mining, Aspect ranking, Semantic approach, Sentiment analysis.

I. INTRODUCTION

In today's world most of the people go for online services like purchasing any product or booking hotels, travels etc. But before going for any purchase for product or any services customer/user wants to go for various reviews for same product/services given by the other customers/users previously. As number of users for e-commerce websites are increased day by day so as increase in reviews written by them is huge. It is not possible for customers/users to get exact information as per their requirements from huge amount of reviews. Information gain from all reviews is very useful for customers/users and manufacturer so that they will get to know about pros and cons of product with respect to their aspects/features. Thus mining this review data and knowing about the customer/users opinions and classify them is an important task. Sentiment analysis is a task of natural language processing that deals with opinion mining for text with respect to topic. Using aspect ranking information customers/users can pay more attention to the aspect which are most liked by others, and manufacturer can also pay attention to aspect with low ranking which help them to take more accurate decision about product buying and improvement in product aspect respectively. It is not an easy task to know about individual feature/aspect from huge amount review data, so knowing about aspect importance from customer/user point of view automatically is an important task. Although document level or sentence level sentiment analysis will classify reviews as positive or negative and get fails to know about customer/users likes and dislikes about any particulars. Having positive review and negative review doesn't mean that opinion holder like or dislike all features/aspects about particulars respectively. Motivated from above observation an aspect based opinion mining system is proposed.

II. RELATED WORK

Previously done work in the aspect based opinion mining are mentioned below in the section:

Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, Member [1] investigate Product Aspect Rnaking. Product aspect ranking is beneficial to a wide range of real-world applications. In this paper, they investigate its usefulness in two applications, i.e. document-level sentiment classification that aims to conclude a review document as expressing a positive or negative overall view, and extractive review summarization which aims to summarize consumer reviews by selecting informative review sentences.
Justin Martineau, and Tim Finin[3] have they work with the delta TFIDF, which is general purpose technique to efficiently weight word scores. Using this technique calculation of aspect value is done but does not take into consideration the frequency of words related with aspect with it.

By using association rule mining Hu and Liu [4] extract aspects from customer reviews and distance based approach is used to extract opinion words and phrases after extracting aspects. To compute the polarization of each extracted opinion word WordNet was used. The negative words were also handled by them.

Richa Sharma, Shweta Nigam and Rekha Jain[5] authors of proposed system used customer reviews of mobile phones are collect together from web site like Amazon.com. Proposed system is based on unsupervised technique. Dictionary based approach of the unsupervised method is use to decide the orientation of sentences. WordNet is used as a dictionary to decide the opinion words and their synonyms and antonyms. After collecting the reviews, they are sent to the POS tagging component where POS tagger label all the words of the sentences to their appropriate part of speech tag. POS tagging is an important segment of opinion mining, it is essential to determine the features and opinion words from the reviews. POS tagging can be completed manually or with the help out of POS tagger. Manual POS tagging of the reviews obtain a lot of time. Here, POS tagger is used to tag all the words of reviews. With the help of seed list record, the polarity of the sentences is determined for each feature/aspects. polarization is determined on the basis of bulk of opinion words, if the number of positive words are more, then the polarity of the sentence is positive or else the polarity is negative and if the number of positive and negative words are equal then the sentence show the neutral polarity.

III. PROPOSED SYSTEM

The proposed aspect ranking system work as follows. Data collection for proposed system is done in two forms, we have collected large number of reviews from web in data set. Reviews are collected from the websites on the internet such as ammazo.com flipkart.com etc. This collected data set will be load by admin authority for further processing and get stored in database. Secondly users can also register, login and give their reviews about products which will get stored in database for further processing.

The pipeline for the aspect ranking opinion mining system is define as follows(fig 1):
1) Extracting opinion words for given list of aspects
2) Sentiment Analysis
3) Ranking of Aspect

For given reviews we determine whether opinion given by the users for the given list of aspects are positive or negative. After sentiment analysis with the help of c4.5 classifier we rank the aspects with the help of aspect ranking algorithm.

Designed proposed system can divided into the following modules:

A. Data Collection

We have taken initially product under three categories Mobile, Laptop and Digital camera. Reviews are collected from different review sites on web and also taken from user dynamically after registration. After collecting data POS tagging is applied to stored data as POS tagging is important process for opinion mining and this processed data is used for further sentiment analysis process of aspect list given.
B. Sentiment Analysis on given aspect list

In this module extracting opinions for particular aspect given in aspect list is inferred. In our proposed system semantically similar features list are used in order to improve the accuracy of feature based summarization. For more understanding if display is aspect/feature some people may use display word to give opinion about it but some people may use another word which is semantically similar to display i.e. screen, display size etc. in such case our proposed approach will also consider semantically similar words for display which will result in more accuracy about display aspect sentiment(Table 1). We have used Senti-word net for identifying the opinion polarity for aspect/features with C4.5 classifier. Sentiment analysis provide the pros and cons about product reviews given by the customers/users.

TABLE 1. SAMPLE TABLE FOR FEATURES, SEMANTI CALLY SIMITAR WORDS AND OPINIONS

<table>
<thead>
<tr>
<th>Feature List</th>
<th>Semantically Similar Words</th>
<th>Opinions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Display</td>
<td>Screen, Display Size etc.</td>
<td>Good, Sensitive, Nice, Bright.</td>
</tr>
<tr>
<td>Camera</td>
<td>Front camera, Rare camera</td>
<td>Awesome, great, High resolution</td>
</tr>
<tr>
<td>Battery</td>
<td>Battery life</td>
<td>Good, Long.</td>
</tr>
<tr>
<td>Network</td>
<td>Wifi, Internet, Wireless</td>
<td>Good, Fast</td>
</tr>
<tr>
<td>Platform</td>
<td>Processor, OS</td>
<td>Fast, Efficient</td>
</tr>
</tbody>
</table>

C. Aspect Ranking for product

Input: Collected review data and list of aspects of products

Output: Opinion for each product aspect given by users.

This module summarized the number of users who have reviewed aspects of different products which consider the positive and negative opinion about aspects as primary consideration. In our proposed system user can select product that he/she want to know about its aspect ranking. The
aspect with highest opinion weight will be on top position and with lowest weight at the end. User can select product as per their requirements, i.e. if he/she want to go for single product aspect ranking or they want to compare multiple product aspect ranking they can choose multiple selected product from list of product. Using frequent reviews from user/customers with overall opinion about product is used to compute the ranking.

**D. Pseudo Code for proposed system**

Step 1: Read all reviews R given by the customers/users and are stored in database.
Step 2: Apply pre-processing on review data.
   2.1: Apply sentence segmentation.
   2.2: Do tokenization and remove unwanted words.
   2.3: Stemming
   2.4: Apply POS tagging
Step 3: Read given aspect list A and extract adjacent adjectives as opinion for matching aspect from the reviews after POS tagging.
Step 4: As we are using Semantic word approach for aspects, if we didn't get aspect from the given aspect list we will get opinion for that aspect semantic meaning word.
Step 5: After opinion extraction sentiment polarity for aspect will be calculated.
Step 6: Finally calculate the ranking for product aspects.

**IV. EXPERIMENTAL RESULTS AND DISCUSSION**

In this section now we are discussing performance of the system on basis of extraction of opinion of given aspect list and aspect ranking algorithm. We have use Normalized Discounted Cumulative Gain with top k as evaluation metric for calculating performance of aspect ranking see fig. 2(for mobile phone).

\[
NDCG@k = \frac{1}{Z} \sum_{i=1}^{k} \frac{2^{l(i)} - 1}{\log(1 + i)},
\]

![Overall performance comparison Aspect-wise](image)

*Figure 2: Performance of Aspect ranking algorithm*

Overall comparison for product is also evaluated for normal sentiment method and method with C4.5 classifier with following gain formulas:

**PositiveInfo** = \( \frac{\text{No of positive reviews retrieved}}{\text{Total no. of reviews retrieved}} \) * \( \log \left( \frac{\text{No of positive reviews retrieved}}{\text{Total no. of reviews retrieved}} \right) \) (1)
GainInfo = Positive Info – Negative Info .................................................................(3)

ProductInfo = \[ \frac{\text{No. of reviews of product retrieved}}{\text{Total no. of reviews retrieved}} \right) \times (\text{Positive Info} – \text{Negative Info}) ........(4)

GainRatio(\text{Product}) = \frac{\text{GainInfo} – \text{ProductInfo}}{\text{GainInfo}} .........(5)

for sentiment method we have not used semantic approach as used with C4.5 classifier, so compare to sentiment normal method C4.5 give best method results with aspect ranking algorithm. Evaluated result for Sentiment Vs C4.5 Classifier method is shown in below figure 3 (for mobile phone).

Figure 3. Sentiment analysis and C4.5 comparison

V. CONCLUSION AND FUTURE WORK

Aspect based opinion mining gives more accurate information about product features pros and cons. In proposed system we have used semantically similar word approach which improves accuracy for identifying opinion about given aspect list as compare to single word aspect list. Using C4.5 classifier for classification we get more accurate results compare to normal sentiment analysis.

In future work we would like to use proposed system with more set of reviews and with different domain and to make system more efficient reviews can be taken dynamically from online websites.

REFERENCES


