Aspect Based Analysis for Rating Prediction of the Restaurant Reviews

Namita Mittal, Basant Agarwal, Shalini Laddha, Manish Sharma

Department of Computer Science & Engineering
Malaviya National Institute of Technology,
Jaipur, Rajasthan, India.

Abstract

With the increased accessibility of internet everywhere, merchants selling products on the Web often ask their customers to review the products that they have purchased and the associated features and services. For a popular restaurant/product, there are a large number of reviews available on the websites. Although the websites allows us to read reviews and understand the qualities of a restaurant/product ourselves, it is really hard to get an overall idea of the same without reading all the reviews. This makes it difficult for a potential customer to read all the available reviews and take an informed decision on whether to purchase the product or to choose the restaurant for dining or lunch. This paper proposes the method to generate ratings for features of a restaurant/product from its opinionated reviews, i.e., given a set of reviews about a restaurant/product, we can obtain a set of features and its ratings. The customer can easily compare two restaurants or products on the basis of feature based rating and can quickly take the decision.

Keywords: Opinion polarity, summarization, reviews, Feature based Sentiment classification.

I. INTRODUCTION

With the rapid expansion of e-commerce, the customers are asked to give online reviews which the other people (customers or merchants) can read to get the overall idea about the product [8, 9, 10]. But there is no way to compare the products or restaurants at their feature level [5]. There is no way to distinguish between a restaurant with great food and below average ambience with the restaurant with the exact opposite features with such a measure. This lack of feature wise information about the products and restaurants reviews and ratings in individual categories (e.g. food, ambience, and service for restaurants) hinders easy comparison between restaurants and quick decision making [5].

On the other hand, customer reviews, particularly the text describing the features, comparisons and experiences of particular product/restaurant provide a rich source of information to compare products/restaurants and make the decisions. Online retailers like amazon.com and flipkart.com allow customers to add reviews of products they have purchased [16, 17]. These reviews have become a diverse and reliable source to aid other customers. Traditionally, many customers have used expert rankings which rate limited a number of products/restaurants. Existing automated ranking mechanisms typically rank products/restaurants based on their overall quality. However, a product/restaurant usually has multiple features, each of which plays a different role. Different customers may be interested in different features of a product or restaurant, and their preferences may vary accordingly [3] [4].

Reading all the reviews and getting an overall idea about the product or restaurant is really time consuming. While reading reviews, the user looks for the describing words i.e. adjectives. As the no of reviews increases, it becomes difficult for the customer to take an informed decision. It is better to have a summary of rating of any restaurant /product for each of the features like Food, Ambience and Service for restaurant on the scale of 0-5 in spite of reading the lots of long reviews. These summaries can be used to efficiently compare two restaurants/products in a very small time.

In this paper, we aim to generate significant attributes of a restaurant in the 4 categories: Food, Price, Ambience and Service. The goal of this paper is to develop a system which can rate the features of a restaurant/product along a scaled range from zero to five. So far, most of the major research on sentiment analysis has been done to predict the polarity of text: positive or negative sentiment, but not subjective opinions along a multi-class continuum.

The main objective of this paper is to develop a system which can rate the features of a restaurant/product along a scaled range from zero to five. Rest of this paper is organized as follows, In section 2 related work is presented. Section 3 describes proposed approach. Section 4 discusses the experiments and results, and finally section 5 presents the conclusion.

II. RELATED WORK

Sentiment analysis research has been increased in recent times due to its importance in industry importance [11,12, 13, 14, 15, 18]. The work done in this research is closely related to Minqing Hu and Bing Liu in [1] on Mining and Summarizing Customer Reviews. They have studied the problem of generating summaries of the customer reviews. In their model they have generated the summary in the format like every feature is written with its opinion sentence. As the number of reviews increases the size of the summary also in-creases so it becomes inconvenient for the customer to read all the opinion

sentences of all the features of particular product and take an informed decision.

The work done in this paper differs from theirs in three main aspects:

(1) The problem of coreference resolution has been resolved which was not handled by them.

(2) The rating of reviews is done and the results are summarized in the form of bar graphs.

(3) Proposed method is generic and can be used for any domain.

III.   PROPOSED APPROACH

In this section, the overview of the proposed approach for feature based sentiment analysis is presented as shown in Figure 1. The system performs the summarization in three main stages:

(1) Mining product/restaurant features that have been mentioned by the customers in their reviews.

(2) Identifying opinion words about the mined features and analyzing their polarity scores.

(3) Summarizing the results in the form of bar graph.

A. Coreference Resolution

Coreference resolution is one of the common problems of natural language processing to determine multiple expressions in a sentence or document referring to the same thing, i.e., they have the same "referent". For coreference resolution, we have used the StandfordCoreNLP in which the coreference resolution system is integrated.

For example, in the review "I bought Samsung Galaxy phone one week ago. It looks very nice. I made many calls in the past three days. They were great, "It" in the second sentence refers to Sam-sung Galaxy Phone, which is an entity, and "they" in the fourth sentence refers to "calls", which is an aspect. Recognizing these coreference relationships is clearly very important for aspect-based sentiment analysis.
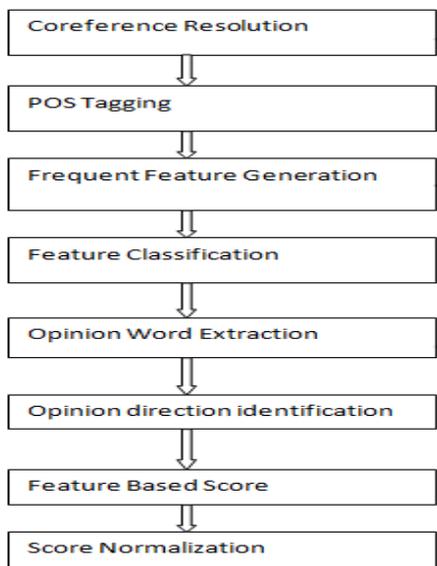

Fig 1. Proposed Feature Based Rating System

B. Feature Generation

Features are the words that represent a particular quality of a restaurant/product. We assume that features can only be singular, plural or common nouns (NN, NNS, and NNP). Thus, all the words with POS tags NN, NNS or NNP are considered in the review sentences.

1. Frequent Feature Extraction

To find features that people are more interested in from several words appearing in the review sentences as nouns, apriori algorithm is used. The features have support value greater than the threshold value are frequent features. There is a chance that some unwanted features are also extracted using this which are not relevant. So feature pruning is required .The infrequent features will be removed in the subsequent steps.

C. Feature Classification

The problem of classifying the features in the main classes is not so easy as we can't simply considering the word by itself and try to assign a particular class to it. There are two methods that can be used in this situation to classify the features:

- The context in which the word is used in the sentence,
- Hypernyms, hyponyms, synonyms, description and other linguistic forms associated with the feature word.

In this paper, the second approach is used in which Wordnet is used to find the synonyms and hypernyms of a particular query word. For example: the hypernyms of certain feature word:

- rice: grain, foodstuff, food, product, food, nutrient, substance, matter, physical entity, entity.

So this feature is classified under the Food class for restaurant reviews.

D. Opinion Words Extraction

For each sentence in the review database: if it contains a frequent feature, then extract all the adjective words close to it as words opinion using the typed dependencies obtained from the Stanford parser[5]. The dependencies that are extracted are: amod, nsubj, and neg.

For example in the sentence: "The food was delicious" we get the nsubj (delicious, food) dependency.

1. Negation Handling

Generally, negation words reverse the polarity of the opinion words in the sentences. Therefore, it is very important to handle negation indicators in the sentences for sentiment analysis. To handle the negation in the review sentences, the negation modifier dependency is used.

For example: "The food is not good." The dependency is neg (good, not).So the sentiment polarity of the word "good" need to be reversed because it is used in negative sense.

2. Automatic Feature Pruning

For the words that occur frequently in our reviews, there is a chance that they may not be the feature words. So these features are removed at this stage as we will not get any opinion word for them.

E. Assigning Polarity to Opinion Words

To find the polarity of the opinion words, the SentiWordNet is used. SentiWordNet [7] is a tool used for finding the polarity score of any opinion word. It is based on the wordnet and the synsets of the wordnet are assigned three scores on the scale of 0.0 to 1.0.

F. Finding Feature wise scores and rating

Once the individual scores of the opinion words are calculated, the total score of the product or restaurant feature wise is calculated. Then the concept of weighted average is used to find the total score.

Here the value of the weight is the number of times the opinion word occurs for the particular feature word. The score that we get is in the range [-1, 1] which we normalize to get the value in the range [0, 5].

IV. EXPERIMENT AND RESULTS

A support threshold value 0.1 is used for the apriori algorithm; which means that if an itemset appears more than 10% of the time, it will be considered as frequent. With a threshold lower than 0.1 for the frequent itemset algorithm, we got many unwanted features. The features were almost correctly classified in the respective classes. We applied our algorithm on different restaurant reviews and mobile reviews. To test our system we applied our algorithm to the set of positive and negative reviews and the results were almost accurate. The rating value greater than 2.5 out of 5 represents positive views of the customer about the restaurant. For the positive restaurant 800 reviews, results are shown in Table 1.

TABLE 1: Results for positive restaurant reviews

Features	Polarity Score	Rating
Food	0.1	2.75
Service	0.176	2.94
Ambience	0.152	2.88
Price	-0.252	1.87

TABLE 2: Results for Negative restaurant reviews

Features	Polarity Score	Rating
Food	-0.448	1.38
Service	-0.436	1.41
Ambience	-0.012	2.47
Price	-0.252	1.87

The price is having polarity less than 2.5 because the restaurant have high price. The results are satisfactory. For the negative restaurant reviews, we got the following results:

V. CONCLUSIONS

Recent trends have indicated that large numbers of customers are switching to online shopping. Online customer reviews are an unbiased indicator of the quality of a product. However, it is difficult for users to read all reviews and perform a fair comparison. This paper describes the methodology and algorithm to give rating to products based on their features using customer reviews. Experimental results show the effectiveness of the proposed approach.

In this paper, only explicit features mentioned in the customer reviews are taken care of. But the customer may have also mention implicit features in the reviews which are very hard to find. For example, "While light, it will not easily fit in pockets." Here the customer has indirectly mentioned his/her reviews about the size of the camera but the word "size" is not explicitly mentioned in the review. To extract such implicit features from the customer reviews, knowledge of semantics are required which is needed to be handled to tackle this problem [22]. We wish to tackle this problem in future. In future we would also try to implement the proposed approach for Hindi Language reviews and tweets [19, 20, 21].

VI. REFERENCES

[1] Minqing Hu and Bing Liu, Mining and summarizing customer reviews, in KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168 177, 2004.

[2] Minqing Hu and Bing Liu, Mining Opinion Features in Customer Reviews, 2004 at American Association for Artificial Intelligence.

[3] Abhishek Gupta, Tejaswi Tenneti, Ankit Gupta, Sentiment based Summarization of Restaurant Reviews, 2009.

[4] The Stanford Tagger: http://www-nlp.stanford.edu/software/tagger.shtml

[5] The Stanford parser: A statistical parser, http://nlp.stanford.edu/software/lexparser.shtml

[6] B. Liu, "Sentiment Analysis and Opinion Mining", Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 2012

[7] Andrea Esuli, Fabrizio Sebastiani, , "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining", Proceedings of Language Resources and Evaluation (LREC), Genoa, Italy, 2006.

[8] Basant Agarwal, Namita Mittal, "Prominent Feature Extraction for Review Analysis: An Empirical Study", In Journal of Experimental and theoretical Artificial Intelligence, Taylor Francis, 2014,DOI:10.1080/0952813X.2014.977830.

[9] Basant Agarwal, Namita Mittal, "Semantic Feature Clustering for Sentiment Analysis of English Reviews", In IETE Journal of Research, Taylor Francis, Vol: 60 (6), pages 414-422, 2014.

[10] Basant Agarwal, Namita Mittal, Vijay Kumar Sharma, "Semantic Orientation based Approaches for Sentiment Analysis", Book Chapter in the book titled "Case Studies in Intelligent Computing – Achievements and Trends", CRC Press, Taylor & Francis. pp: 62-75, 2014.

[11] Basant Agarwal, Namita Mittal, "Machine Learning Approaches for Sentiment Analysis" , Book Chapter in "Data Mining and Analysis in the Engineering Field", IGI Global, pp- 193-208, 2014

[12] Basant Agarwal, Namita Mittal, Vijay Kumar Sharma, "Feature Extraction Methods for Semantic Orientation based Approaches to

Sentiment analysis", In 10th International Conference on Natural Language Processing (ICON-2013). pp: 225-230, 2013.

[13] Basant Agarwal, Vijay Sharma, Namita Mittal, "Sentiment Classification of Review Documents using Phrases Patterns", In Second International Symposium on Natural Language Processing (NLP'13). pp.1577-1580, 2013

[14] Basant Agarwal, Namita Mittal, "Sentiment Classification using Rough Set based Hybrid Feature Selection", InProceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA'13), pages 115–119 , NAACL-HLT, Atlanta, USA, 2013.

[15] Basant Agarwal, Namita Mittal, "Optimal Feature Selection for Sentiment Analysis", In 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2013),Vol-7817, pages-13-24, Greece, Samos. 2013.

[16] Basant Agarwal, Namita Mittal, "Categorical Probability Proportion Difference (CPPD): A Feature Selection Method for Sentiment Classification", In Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2012), COLING 2012, IIT-Bombay, Mumbai, pages 17–26, 2012.

[17] Basant Agarwal, Namita Mittal, "Text Classification using Machine Learning Methods- A Survey", In 2nd International Conference on Soft Computing for Problem Solving (SocPros-2012), Vol:236, pp: 701-710, 2012.

[18] Basant Agarwal, Namita Mittal, Pooja Bansal, Sonal Garg, "Context-aware Sentiment Analysis using Common-sense knowledge based Ontology", In Computational Intelligence and Neuroscience, Article ID: 715730, 2015, (SCIE Indexed), Article ID 715730, 9 pages, DOI: http://dx.doi.org/10.1155/2015/715730.

[19] Namita Mittal, Basant Agarwal, Garvit Chouhan, Nitin Bania, Prateek Pareek, "Sentiment Analysis of Hindi Reviews based on Negation and Discourse Relation", In the proceedings of 11th Workshop on Asian Language Resources, (in conjunction with IJCNLP -2013). pp: 45–50, 2013.

[20] Namita Mittal, Basant Agarwal, Garvit Chouhan, Prateek Pareek, Nitin Bania, "Discourse based Sentiment Analysis for Hindi Reviews", In 5th International Conference on Pattern Recognition and Machine Intelligence (PReMI'13), pp. 720–725, 2013, Springer.

[21] Namita Mittal, Basant Agarwal, Saurabh Agarwal, Subham Agarwal, Promod Gupta, "A Hybrid Approach for Twitter Sentiment Analysis", In 10th International Conference on Natural Language Processing (ICON-2013), pp.116-120, 2013.

[22] Namita Mittal, Basant Agarwal, Ritul Aditya,"Ontology Driven Semantic Query Optimization", International Journal of Computer Systems (IJCS), Vol: 1(3) , pages: 85-90, 2014.