

Product Ranking using Opinion Mining

Ashutosh Tiwari, Aditya Vishwekar, Karishma Sawant, Prof. Kavita Bathe

Department of Computer Engineering

K.J. Somaiya Institute of Engineering and Information Technology,

Ayurvihar Complex, Everard Nagar, Sion, Mumbai-400022

Maharashtra, India

Email:{ashutosh.t, a.vishweshwar, karishma.sawant, kavitag}@somaiya.edu

Abstract- Due to the high increment in the usage of internet and growth of web, there has been explosive exponential increase in the user-generated content in the form of blogs, discussions, customer reviews, discussion forums, social networks etc. it is a challenging task to extract the knowledge from these data, which are stored in unstructured or semi-structured manner. Online shopping is becoming more and more important because of internet and the amount of user-generated content it contains in the form of reviews. Therefore, it is necessary to design effective system to summarize the result of review in the form of knowledge. The summarization of pros and cons of product characteristics helps consumers to quickly find their product. In this paper, we present a review of product ranking system using opinion-mining techniques. The user can specify product and desired feature to get the targeted ranking result of products.

Index Terms- Opinion Mining, Product Ranking, POS Tagging, Information Retrieval

1. INTRODUCTION

The growth of World Wide Web is exponential in terms of size and the variety of contents created. Due to which the e-commerce is growing as well. The most of the data generated in e-commerce comes from reviews provided by the users of the platform. These users are the buyers who give the reviews on these products. These reviews represent user's experience, level of satisfaction with that product. These reviews influence the decisions of other customer's interest in that product; importantly these feedbacks have an important role in developing marketing strategy and decision making of product commercial lifecycle. It is difficult to process large amount of review manually and take decisions about selecting the right product. The extraction of knowledge from these long and huge numbers of reviews is difficult. The user may not get the right product of his choice because of, not processing the reviews properly or getting a biased feedback by just reading small number of reviews. These reviews also help the manufacturers to rule out the possibility of product being rejected or affecting the sales by focusing on the feedback of the products. Secondly, a large number of reviews and textual information are difficult for product manufacturers to understand the drawbacks of products and customer's opinion.

2. REVIEW OF LITERATURE

Past works on opinion mining with respect to sentiment classification refers to the use of standard machine learning algorithms such as [1]. In [1] Pong and Shivakumar examined the effectiveness of applying machine-learning techniques to the sentiment classification problem. Naïve Bayes, SVM and Maximum Entropy were used and their results were compared. In [2] features, candidates were ranked by feature relevance and feature frequency. The problem was formulated as bipartite graph and page ranking algorithm HITS was used to find the importance features and rank them. Double propagation is based on the observation that there are neutral relations between opinion words and features. The association between feature (noun) and opinion words (adjectives) helps in determining the overall sentiment towards that feature.

In [3] sentiment aware feature set were created based on appraisal words lexicons. The orientation of the comment made and the feature were extracted to give sentiment towards that product. Review categorization was done by aggregating the sentiments on the features. Given the Sentence *S*, which contains product feature *f*, the adjective words in *S* are identified, and the polarity of these are identified using appraisal lexicons. Authors in [4] make use of POS as a baseline to extract the features from the datasets. New POS based pattern are identified for sentiment-rich phrase extraction. These patterns are

able to extract sentiment- rich phrases, which incorporate contextual information, and dependency relation based phrase. After extraction of Sentiment-rich phrase, semantic orientation of these entire phrases is computed using PMI method.

The semantic analysis of sentences can also be used to find the key information form the document. These information components are centered on product, associated opinions as well as features, which can be extracted using natural language processing techniques. In [5] Stanford parser is used to generate dependency tree also known as word-word relationship, encoding grammatical relations between every pair of words. Polarity of extracted opinions for each feature are classified using senti-wordNet, for each opinion sentences are examined and mapped into classes with sentimental values.

In [6], Authors extracted product-feature object and computed context polarity. The word present in sentence are searched in product ontology and feature ontology and if there is hit, syntactic parser is used to map the ontology word to ontology word.

Yin-Fu Huang and Heng Lin [7] proposed a framework for ranking the products using 1) product review 2) product popularity and 3) product release month. Average polarities of all reviews of a product are considered to calculate the score along with Popularity weight and Weight of product release month.

3. PROBLEM STATEMENT

Product reviews are generally in large number, which is difficult to read and comprehend the total sentiment. In order to identify the total sentiment of the user towards specific feature set F from total set of features say S, we identify the sentiment of reviews given towards specific feature among given features and present the list of products having reviews in favor of given features to not in favor of given feature. Polarities of reviews are analyzed towards the targeted feature and result is collected.

4. PROPOSED SYSTEM

At human level it is quite possible to distinguish and understand the polarity of given sentence. There are set of words, which help us identifying the positivity or negativity through understanding and comprehension of meaning. Similarly, supervised learning can be used to train the classifier about the possible set of classes to be used and result of their sentiment can be extracted. But in order to understand the message conveyed with respect to particular subject in the review, we must break down the

sentence to its syntactical level and find out the sentiment towards that subject.

4.1 Classification

Naïve Bayes classifier does sentiment classification of each customer review. Each review is differentiated as positive and negative. We derive the Naïve Bayes (NB) classifier by observing that by Bayes' rule,

$$P(A/B) = \frac{P(A)P(B/A)}{P(B)}$$

4.2 Identifying Sentence weight

4.2.1 *POS tagging*: Parts of speech are important because it helps in finding the adjective and its degree, which has sentimental values. POS_v is used to represent the degree of adjective and stores numeric values.

4.2.2 *Inverse Document Frequency*: In order to reduce the impact of frequently used adjectives and adverbs, inverse document frequency is used to calculate the numerical representation of given words. We use IDF to reduce the effect of common adjectives and enhance important adjectives.

IDF is calculated as follow,

$$IDF_{CR} = \ln \left(\frac{CR}{TR} \right)$$

Where CR is the number of reviews containing given adjective P in given product review set, and TR is total number of Reviews for given product.

4.2.3 *Calculating Sentence Weight*: Sentence weight is calculated as follows

$$Sentence_w = POS_v \times IDF_{CR}$$

4.3 Calculating Product Rank:

In this section, we will present a set of product ranked in order of their score. The user request is taken into consideration to find the set of products satisfying the requested features.

The product score will be calculated as follows,

$$Score_p = R_w \times CPR \times RP_w$$

4.3.1 *Calculating Review Weight (R_w)*: Review weight is calculated as follows,

$$R_w = \sum_{k=1}^n Sentence_w$$

4.3.2 *Cumulative Polarity of Reviews (CPR)*: Cumulative Polarity of Reviews is calculated as follows,

$$CPR = \frac{\sum_{k=1}^n R_{kw}}{n}$$

Where, n is total number of reviews.

4.3.3 Review Popularity Weight(RP_w):

Review Popularity Weight is used to find the popular product among others; it is calculated as follows,

$$RP_w = \frac{V_i}{\max(V)}$$

Where, v is total number of reviews for product i, and max(v) is maximum number of reviews for a product among all.

5. CONCLUSION

Opinion mining and sentiment analysis help us in identifying the sentiment of the user and the various data mining techniques support in finding valuable data and pattern inside the data. With the help of a combination of these two techniques, we can achieve desired result.

ACKNOWLEDGEMENT

We would like to convey our sincere thanks to the many people who guided us throughout the course of this project. In addition, we would like to express our sincere thanks to Prof. Kavita Bathe for here immense guidance, encouragement, co-operation and suggestions given to us at progressing stages of the project.

Finally we would like to thanks our Head of Department Prof. Sunita Patil and all teaching, non-teaching staff of the college and friends for their moral support rendered during the course of the project work and for their direct and indirect involvement in the completion of our project work, which made our endeavor fruitful.

REFERENCES

- [1] Bo Pang, Lillian Lee and Shivkumar Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," Proceeding of EMNLP, pages 79-86, 2002.
- [2] Lei Zhang, Bing Liu, Suk Hwan Lim and Eamonn O'Brien-Strain "Extracting and Ranking Features in Opinion Documents," Coling 2010 Poster Volume, Pages 1462-1470, Beijing, August 2010.
- [3] Deptii D. Chaudhari, Prof. R.A. Deshmukh, Dr. A.B. Bagwan and Dr. P.K. Deshmukh "Feature

Based Approach for Review Mining Using Appraisal Words,"

- [4] Basant Agarwal, Vijay Kumar Sharma and Namita Mittal "Sentiment Classification of Review Documents using Phrase Patterns," ICACCI, pages 1577-1580, 2013.
- [5] Tanvir Ahmad and Mohammad Najmud Doja "Ranking System for Opinion Mining of Features from Review Documents," IJCSI, pages 440-447, 2012.
- [6] P. Tian, Y. Liu, M. Liu and S. Zhu, "Research of product Ranking technology based on opinion mining," proc. The 2nd international Conference on Intellegent Computation Technology and Automation, pages 239-243, 2009.
- [7] Yin-Fu Huang and Heng Lin, "Web Product Ranking Using Opinion Mining" Proc. 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), pages 184-190, 2013