

Feature Extraction from Reviews Using Opinion Mining

Aditya Vishwekar

Department of Computer Engineering, kjsieit
Ayurvihar Complex, Everard Nagar, Sion, Mumbai 40 0042
Maharashtra, India
Avishwekar10893@gmail.com

Abstract-The online shopping is becoming more convenient and prominent. Accordingly, the numbers of online customer reviews are increasing very fast. It has become difficult for the organizations and consumers to analyze these reviews for decision making process. Mining the essential features from these reviews is beneficial to consumers and organizations. Feature mining can be very helpful for customer relation management, marketing and finding the popularity of the product. The proposed framework uses N-gram, POS tagging and no-pattern.

Keywords- Feature mining, N-gram, POS Tagging, no-pattern.

1. INTRODUCTION

For the past few years, feature analysis or opinion mining (Liu, 2010; Pang and Lee, 2008) has been an active research area in NLP. Online reviews have become an important and a useful resource for the product manufacturers and the service providers to understand the voice of the customers on their product or service which helps them to adjust their marketing strategies in order to capture the market. There are several studies on feature extraction (e.g., Hu and Liu, 2004, Popescu and Etzioni, 2005, Kobayashi et al., 2007, Scaffidi et al., 2008, Qui et al., 2009). There are a number of potential applications which includes business, politics, movies, tourism, sentiment analysis, social networking sites etc. which has motivated the need for automatic feature extraction. [1] Discusses the main research problems and some techniques used to solve different levels of feature extraction problems.

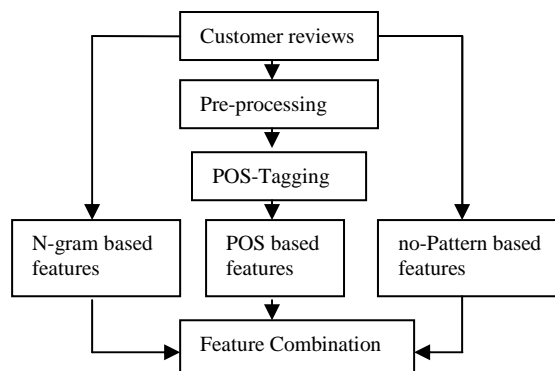


Fig. 1.1 Block Diagram

2. BACKGROUND AND MOTIVATION

The Various research groups are exploring the ways to use feature extraction and sentiment analysis as next generation paradigm shift. Review classification is most promising topic in Feature extraction. Hu and Liu (2004) proposed a technique based on association rule mining to extract product features. Due to many problems related to the accuracy, precision and recall of feature extraction exact solution is yet to be found. Lie, Bing Suk and Eamonn proposed a Double Propagation method which extracts noun features, and works well for medium-size review dataset.

3. PROBLEM STATEMENT

Online shopping is becoming increasingly important as more and more manufacturers sell products on the Internet, and many users are using the Internet to express and share their opinions. However, it is impossible for consumers to read all product reviews. Therefore, it is necessary to design effective systems to extract features to perform sentimental analysis, so that consumers can quickly find their favorable products. The main aim of this paper is to propose a technique to increase the accuracy of the feature extraction from particular review.

This paper explores the different combination of sentence classification methods to extract features from customer reviews.

4. REVIEW LITERATURE

In recent years, the focus was on ranking the online products based text mining and sentiment analysis. The existing technology is not efficient to extract features from consumer reviews. Double propagation

is an unsupervised technique which mainly extracts noun features [2]. Following the initial work in (Hu and Liu) [5], several researcher have further explored the idea of using opinion words in product feature extraction. A dependency based Method was proposed in for a movie review analysis application [3][4].

There is another important concept discussed in [2], wherein features can be extracted from a sentence based on one single word “no”. Here “no” represents word no. The basic form of the pattern is “no” word followed by noun/noun phrase. The pattern can be very helpful in feature extraction. In everyday language, people are often not strictly grammatical. They write their opinions or comments on features by this short pattern. For example in mattress domain, people always say that “no noise” and “no indentation”. Here “noise” and “indentation” are all features for the mattress. But, there are some issues with the use of this pattern. Some fixed “no” expression, like “no problem”, “no offense”. In these cases, “problem” and “offense” should not be mistaken for features. To minimize these errors N-gram technique can implemented. As discussed in [1], an N-gram is a contiguous sequence of n-items from a given sequence of text. An N-gram of size 1 is termed as “unigram”; size 2 is a “bigram”; size 3 is a “trigram”. For example, consider a sentence “I like the camera”, the bigrams are “I like”, “like the”, “the camera”. The trigrams are “ I like the” and “like the camera”. In proposed work, bigrams and trigrams occurring more than 6 times are extracted as features for the classifier.

To increase the efficiency of feature extraction, it becomes necessary to analyze and classify each word in the customer review. This analyzing and classification of the text can be done by Part-of-Speech tagging. The POS tagging plays a basic role in various Natural Language Processing (NLP) applications. Most of feature analysis systems require the POS tagging, such as information extraction, information retrieval, word sense disambiguation, machine translation, etc. Various sentiments about different features are expressed on e-commerce websites. For example, consider the sentence, “The camera is good and awesome”. The POS tagged sentence is “The_article camera_noun is_verb good_adjective and_conjunction awesome_adjective”. Only the required nouns can be classified as features. Further feature combinations and performance evaluation of a system can be carried out.

5. CONCLUSION

The field of opinion and text mining needs to be explored in order to identify important features related to various products. Research in the field of feature extraction from text can be of great help in developing marketing strategies and analyze large amount of data. Thus, the project plans on utilizing opinion mining

techniques for feature extraction from review datasets to benefit both consumers and organizations.

Acknowledgments

I would like to express my sincere thanks to **Prof. Kavita Bathe** and **Ashutosh Tiwari** for their encouragement, co-operation and suggestions given for the development of the project.

REFERENCES

- [1] Priyanka C. and Deepa Gupta, “Identifying the Best Feature Combination for Sentiment Analysis of Customer Reviews,” International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2013 .
- [2] Zhang, L.; Liu, B.; Lim, S.; O’Brien-Strain, E, “Extracting and Ranking Product Features in Opinion Documents,” Poster Volume, pages 1462-1470, August 2010.
- [3] Deptii D. Chaudhari, Prof. R.A. Deshmukh, Dr. A.B. Bagwan, Dr. P.K.Deshmukh, “Feature Based Approach for Review Mining Using Appraisal Words”, pages 178-182, 2011.
- [4] Zhuang, Li.; Feng Jing, Xiao-yan Zhu, “Movie Review Mining and Summarization”. In Proceedings of CIKM, 2006.
- [5] Hu, Mingqin and Bing Liu., “Mining and Summarizing Customer Reviews.” In Proceedings of KDD, 2004.

□