Enhanced Opinion Mining Approach for Product Reviews

Yugandhara Gawand
Department of Computer Engineering
Pillai College of Engineering
New Panvel, Navi Mumbai, India

Satishkumar Varma
Department of Computer Engineering
Pillai College of Engineering
New Panvel, Navi Mumbai, India

Abstract:- Advanced technology development and high speed internet connection have connected people across the world. Every business marks its presence on web and directly gets connected with the interested customers. On the other hand, consumers express their opinion about the services and products on companies' website, forum or any other social platform. Therefore, as data is available in bulk, it is very difficult for manufacturers and buyers to make decision due to vagueness. This paper proposes an enhanced opinion mining approach that filters ambiguity and highlights general sentiments of a particular product. This aspect based opinion mining approach has three important phases, prepossessing followed by feature extraction and then ranking.

Keywords:- Sentiment Analysis; Opinion Mining; Polarity; Machine Learning.

I. INTRODUCTION

Analysis of available data is essential for efficient decision making. Opinion mining techniques that come under Natural Language Processing (NLP) or Information Extraction (IE) are analyzing approaches that aims to understand writers' views expressed in the form of text. These sentiments can be either positive, negative or objective. Opinion mining focuses on determination of experience and insights of a consumer or a review writer for a particular product and then assign correct polarity to the text reviews.

Opinion mining identifies the phrases of a text that represents the real sentiment. People can express their views as objective facts or subjective opinions. It is necessary and important to understand both the types. Sentiments have different classes as objective (describe about facts), and subjective (views with polarity). Subjective type is again classified as positive (represents joy and satisfaction) and negative (represents disappointment).

The Sentiments have specific degree of positivity, negativity or objectivity. Therefore, the sentiments of the given text can be calculated in terms of different levels.

Document Level Analysis - entire document is assigned a single polarity, either positive, negative or objective. Sentence Level - document is differentiated into sentences and every sentence is assigned separate polarity based on its type. Hence, sentences will have their own polarity. Phrase Level – makes segments at phrase level and then assign each phrase a separate polarity. It is also known as aspect based analysis.

II. RELATED WORK

Major concern with opinion mining approaches are their improper identification of characteristic of given reviews. If the system fails to identify exact feature then classification of reviews becomes extremely difficult and challenging task. In [1] the system tries to find out opinion mining from the patient review. This system uses probabilistic aspect mining model. It focuses on finding aspects related to one class labels. Advantages of this system again it is not domain-specific. But if the aspect value is too small to find.

The aspect-based opinion mining [2] process uses syntactic dependency, aggregate score of opinion words, SentiWordNet and aspect table together. It also attempts to detect the overall polarity of a sentence. It is not suitable to find out multiple aspect, implicit aspect, and comparative sentence.

The multiple aspects from different domains is given in [3]. For example user gives review about movie then there is possibility user may give review about actors, direction and many other things. It is complex task to find multiple aspects but this system is not considering short review. And it is domain-specific.

In [4] feature identification and opinion mining has been achieved by using ontology based intelligent systems. Main goal of this system use knowledge based system. To find aspect-based sentiment analysis. Which can be used for multiple domains. But one of the limitation of this system is if aspects are not clearly defined then it hard to find correct aspect. Another one is very small data set is used.

New complex NLP-Based 27 rules are introduced in [5] for the tasks of subjective and sentiment classification at the aspect-level. This system focuses on creating prototype to analyze opinions. Which is mainly from TripAdvisor in the context of the tourism industry. The issue of multiple aspect finding is not addressed. There are systems that require starting questionnaire.

In [6] they targeted on finding customer view without using any question nary. Also detecting multiple aspect review is main aim. For doing so it requires bootstrapping based ART algorithm. It requires initial data in bootstrapping algorithm.

In [7] uses users answers for aspect mining [8]. It can be used for online forums/discussions/blogs. When users purchase any product it has to answer some question. System includes fives phases Question analysis, Question expansion, high quality review retrieval, subject sentence extraction and answer grouping.

III.   PROPOSED SYSTEM

The proposed system involves three phrases pre-processing, feature extraction followed by aspect mining and ranking. Overview of proposed system is given in figure 1.
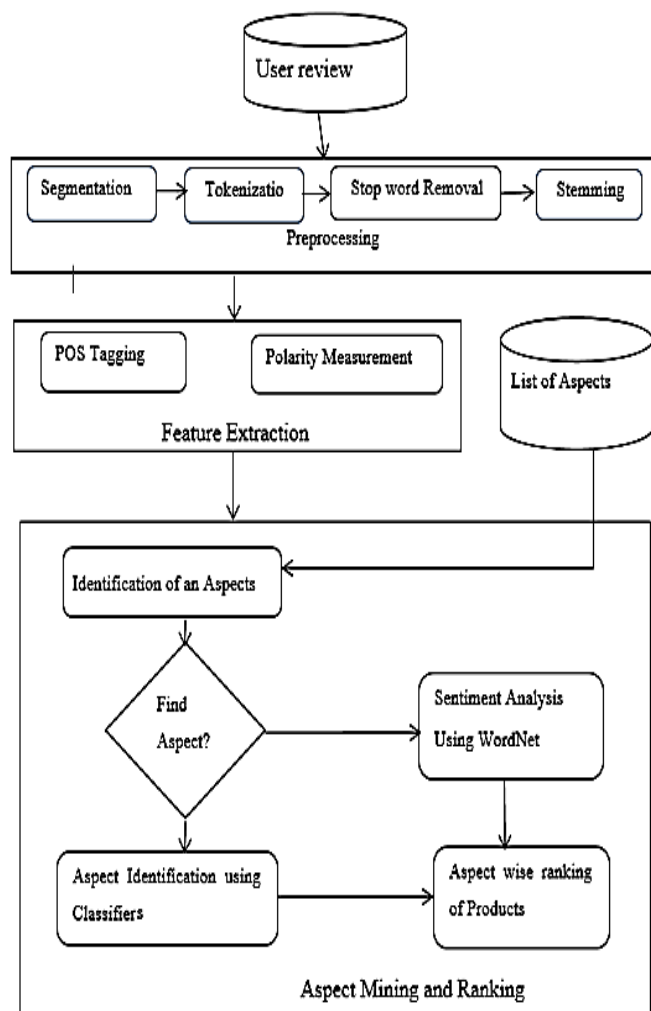

Fig 1:- Proposed Enhanced Aspect based Opinion mining System.

Algorithm describes the detailed processing of proposed system. The algorithm specifies step by step process of this enhanced opinion mining approach.

The algorithm describes internal functioning about system.

Input: Users query
Output: Opinion analysis on users query

Step 1: Recognize users query.
Step 2: Search reviews for product which entered by user.
Step 3: Pre-processing and Feature Extraction.
Step 4: Identify the implicit feature, multiple-aspect feature based on review.
Step 5: Extract the opinion word from extracted feature like adjective / verb in a review sentence. For e.g. cannon camera is low in price as compare to Nikon.
Step 6: Prepare the table for extracted feature along with opinion word
Step 7: Assign weight to each extracted opinion word.
Step 8: Read given aspect list and extract adjacent adjectives as opinion for matching aspect from the reviews
Step 9: Calculate the sentiment orientation according to their weights classified as positive, negative and neutral. It is carried out using C4.5 classifier and Naive Bayes classifier.
Step 10: Calculate aspect score in all reviews.
Step 11: Find out overall opinion about product.
Step 12: Finally calculate the ranking for product aspects.

Detailed description about important stages of this enhanced system is given below.

A.   Pre-processing

➢ Sentence segmentation - It is the process of differentiation of the given text document/ reviews into its constituent sentences along with its word count. In English, sentence is segmented by identifying the boundary of sentence which ends with full stop.
➢ Tokenization - It is the process of dividing sentences into small meaningful pieces. These pieces are called as tokens.
➢ Stop word removal - This is the process removing extra word which have less importance to effectively find out aspects.
➢ Stemming - In this process root of the word is find out. And extra suffixes are removed.

B.   Feature Extraction

Feature extraction is the process of extracting feature from the processed review. Which again follows certain phases.

➢ POS Tagging - In this process words are marked as corresponding to particular part-of-speech. The Open NLP POS Tagger uses a probability model to predict the correct POS tag out of the tag set. Polarity Measurement: Sentence polarity calculated as an integer number. Positive number associated with positive polarity, negative number associated with negative polarity. Proposes system uses SentiWordNet for calculating opinion values of aspects [9].

➢ Polarity Measurement - In next level polarity is calculated in integer number. Positive/Negative/Neutral in positive or negative number or 0.

The following equation use to calculate polarity of the sentence.

$$pol(S) = \sum_{i=1}^{n} pol(cls|i) + \sum_{j=1}^{m} pol(phrase|j1)$$

C. Feature Classification

This approach uses C4.5 algorithm of decision tree for classification of semantics (positive, negative, neutral) for product reviews given in English language. The system is capable to classify large set of data. This approach also focuses on finding info gain with respect to aspects identified using Synsets replacement provided as external file. For comparative analysis Naive Bayes classifier is also implemented. The performance C4.5 classifier is compared with Naive Bayes classifier.

D. Aspect Ranking for Product

Input: Collected review data and list of aspects of products
Output: Opinion for each product aspect given by users.

In our approach user can select aspect and checks its 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. To do so it uses frequency-based method and TF IDF [10] based method.

IV. APPLICATIONS

➢ *Product and Service Sectors*

As every business has both online and offline platform to connect with large number of consumers, the review of their consumers about the product and services are very important. If the review is positive business develops its good brand and also new customers come to know about the service quality. On the other hand, for negative reviews, the goods and service provider can improve the quality from next time onwards to maintain their brand and avoid losing customers, as well as consumers will take better decision for purchasing the commodity. The proposed enhanced opinion mining technique simplifies this process, instead of reading full review people can only go for specified aspect.

➢ *Medical Sectors*

Unlike goods and service sector, this sector is very sensitive, as it deals with the health of people. The curing drugs have limited aspects, such as ease of use, price, effectiveness, dosages, side effects, warnings. Similarly, other more technical aspects are chemical and molecular aspects, but usually the reviewer does not mentioned it in drug reviews. The proposed opinion mining approach helps to deal with drug reviews and the difficult words of the descriptions to find proper opinion about drugs.

➢ *Hospitality Sectors*

Travel and tourism, hotel and restaurant, clubs, parks etc. are the business where online booking is very popular in recent years. These businesses have aspects like food, location, free Wi-Fi, quality of services, pool etc. The proposed algorithm helps to identify the exact place for the user according to their budget and interest or any other aspects.

V. CONCLUSION

Feature based opinion mining for text reviews benefits users during searching and selecting best services. A robust mechanism helps to save time and provides related information. Due to social networking and e-business, reviews are available in bulk, just there is a need to analyze them according to a specific feature. Hence, aspect-based opinion mining finds related opinions for the targeted features and saves time and energy of the user. The proposed approach find outs the related aspects, classifies them and then ranks according to their value. This work aims to help people for finding opinion related to a given product using classifiers. Existing techniques have some limitations like, some systems are domain-specific, and some other are unable to find out aspects from the review. So, this enhanced algorithm provides solution to overcome previous flaws. It will efficiently find outs features from the reviews and can be used with any type of domain.

REFERENCES

[1]. Chinsha T. and S. Joseph. "A Syntactic Approach for Aspect Based Opinion Mining", In the Proceeding of IEEE 9th International Conference on Semantic Computing, pp. 24-31, 2015.

[2]. A.Jeyapriya ,C.S.Kanimozhi Selvi , "Extracting Aspects and Mining Opinions in Product Reviews using Supervised Learning Algorithm", IEEE Sponsored 2nd International Conference on Electronics and Communication Systems (Icecs 2015)

[3]. Victor C. Chenge, C.H.H. Leung, Jiming Liu, "Probabilistic Aspect Mining Model for Drug Reviews" , IEEE Transactions on Knowledge and Data Engineering, August 2014.

[4]. E. Marrese-Taylora, J. Velasqueza,F. Bravo-Marquez, "A Novel Deterministic Approach for Aspect-Based Opinion Mining in Tourism Products Reviews", Expert Systems with Applications, pp.7764-7775, 2014.

[5]. I.Penalver-Martinez, F. Garcia-Sanchez, R.Valencia-Garcia,M.Rodriguez-Garcia, V. Moreno,A. Fraga and J.Sanchez-Cervantes, "Feature-based opinion mining through ontologies", Experts systems with applications, pp. 5995-6008,2014.

[6]. Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, and Tat-Seng Chua "Product Aspect Ranking and Its Applications", IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 5, may 2014.

[7]. F.Xianghua, L. Guo, G. Yanyan and W. Zhiqiang, "Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and How Net lexicon" ,Knowledge-Based System, vol. 37, pp. 186-195,2013.

[8]. Jingbo Zhu, Huizhen Wang, Muhua Zhu, Benjamin K. Tsou, M. Ma, "Aspect-Based Opinion Polling from Customer Reviews" IEEE Transactions on Affective Computing, VOL. 2, NO. 1, pp. 37-49, 2011

[9]. Moghaddam, S. and M. Ester, "AQA: aspect-based opinion question answering", In Proceeding of the IEEE 11th International Conference on Data Mining Workshop (ICDMW'15), pp. 89-96, 2011.

[10]. XU Xueke, CHENG Xueqi, TAN Songbo, LIU Yue, SHEN Huawei, "Aspect-Level Opinion Mining of Online Customer Reviews", Key Laboratory of Web Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China