

Opinion Mining Techniques and Tools - A Study

A. Al Firthous¹ Dr.P.Arul²

¹Research Scholar ²Assistant Professor

^{1,2}Department of Computer Science

¹Bharathiyar University Coimbatore – 641 046 ²Government Arts College Trichy

Abstract— Online reviews taking a vital role in the public, people like to post their views, ideas and opinion about product, movies, and social events in different way in the internet like blogs, Social websites and groups. Online user's opinion increase day by day, people share their experience on shopping and commenting their opinion, so mining the user's opinion is interesting and growing research area, In E-Business product reviews taken as a important area in both side seller and buyer to improve the quality of products and idea to purchase a new product respectively. In this paper we given a detailed study about opinion mining or sentiment analysis and various techniques, tools related to opinion mining, also Preprocessing is important to data before mining, so that so that we explain various data preprocessing techniques too.

Key words: Sentiment Analysis, Opinion Mining, Machine Learning

I. INTRODUCTION

Sentiment Analysis is a rising area of research which widely uses information retrieval and computational linguistics. Opinions of people always take a major role in decision making process and are become an essential part of day today life. Opinions are widely shared on many online resources like review sites, corporate websites, discussion groups, blogs, social media etc. in this people express their ideas and thoughts it also allow business organizations to enterprise and improve their capabilities and products by providing best of their services according to the reasonable needs of the people. User opinion is measured as a important resource which is useful for both clients and product producer. In web pages, the opinions are written in NL language (Natural Language) and are in the form of unstructured. Taking all the user reviews manually and taking in to value of review one by one is computational trouble and is not practically possible with respect to client and manufacture point of view. Therefore it is more competent to automatically process the large reviews and provide the needed information in a suitable form. Sentiment analysis giving the solution to above mentioned problem, there are few main fields of research prevail in Opinion Mining: Opinion classification, feature based Opinion classification and opinion summarization. Sentiment (Opinion) classification deals with classifying entire documents according to the opinions towards certain objects. Feature-based Sentiment classification on the other hand considers the opinions as features of certain objects. Opinion summarization task is totally different from traditional text summarization because only the features of the product are mined on which the user have expressed their opinions [1]. Opinion summarization does not summarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as in the classic text summarization. Sentiment analysis has different applications in different field like

media, marketing and decision making [2]. For example, voters can make their decision according to the reviews of the politicians. In sentiment analysis, there are two types of information, namely, facts and opinions. Facts are objective in nature, this are the statements which describe the nature of a product or event. Opinions are the attitude, appraisals and emotions regarding to that entity [3]. The majority of researches done on objective nature of the product but recent trend are to focus on opinions.

II. SENTIMENT CLASSIFICATION TECHNIQUES

Sentiment classification is nothing but to find common theme among a document or set of text. We have various types of classification techniques in this chapter some machine learning techniques for classification is discussed.

A. Machine Learning Approach

Machine learning techniques in which machine trained to learn from past occurrence and make the machine able to do some pre-defined task. In sentiment analysis, machine learning techniques play an important role [9]. Machine learning techniques typically need two different sets of documents one for the training set and other for the test set. Training set has the functionality to differentiate the features of dataset. Whereas test set is used for validation of the performance, done by the classifier. For classification there are number of machine learning techniques, some of them are SVM, Naive Bayes and maximum entropy[1]. There are number of machine learning techniques available like Naive Bayes, maximum entropy and support vector machine (SVM) [4].

1) Naive Bayes

Naive Bayes is a simple technique based on the probabilistic model. It is generally used for the document classification. The Naive Bayes classifier is based on Bayesian probability and assumed that feature probabilities are independent of one another.

2) Maximum Entropy

Maximum entropy classification (MaxEnt, or ME, for short) is an alternative technique which has proven effective in a number of natural language processing applications (Berger et al., 1996). Nigam et al. (1999) show that it sometimes, but not always, out performs Naive Bayes at standard text classification. Importantly, unlike Naive Bayes, MaxEnt makes no assumptions about the relationships between features, and so might potentially perform better when conditional independence assumptions are not met.

3) Support Vector Machine (SVM)

SVM, a deterministic classification technique, it is considered the best text classification method. Support vector machine provides a discriminative platform to distinguish the points of data in to the pre defined classes and then make the decision that relies on the support vectors. In the starting phase, centric vector has been calculated and then on the basis of the similarities found

document is classified to the class from which the calculated value match [6]. SVM gives the maximum marginal difference between the two classes, so it is considered as the most appropriate technique for the classification.

III. DATA SOURCES

User's opinion is a major criterion for the improvement of the quality of services rendered and enhancement of the deliverables. Blogs, review sites, data and micro blogs provide a good understanding of the reception level of the products and services.

A. Blogs

With an increasing usage of the internet, blogging and blog pages are growing rapidly. Blog pages have become the most popular means to express one's personal opinions. Bloggers record the daily events in their lives and express their opinions, feelings, and emotions in a blog (Chau & Xu, 2007). Many of these blogs contain reviews on many products, issues, etc. Blogs are used as a source of opinion in many of the studies related to sentiment analysis (Martin, 2005; Murphy, 2006; Tang et al., 2009).

B. Review Sites

For any user in making a purchasing decision, the opinions of others can be an important factor. A large and growing body of user-generated reviews is available on the Internet. The reviews for products or services are usually based on opinions expressed in much unstructured format. The reviewer's data used in most of the sentiment classification studies are collected from the e-commerce websites like www.amazon.com (product reviews), www.yelp.com (restaurant reviews), www.CNET download.com (product reviews) and www.reviewcentre.com, which hosts millions of product reviews by consumers.

C. Micro Blogs

Twitter is the most popular example of the microblogging. In twitter users create messages called "tweets". These tweets express the sentiments and emotions of users about different topics [2].

IV. PREPROCESSING TECHNIQUES

Some preprocessing task (1) stop word removing (2) stemming (3) POS tagging

A. Stop Word Removing

Stop word is a commonly used word (such as "the") that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query. When building the index, most engines are programmed to remove certain words from any index entry. The list of words that are not to be added is called a stop list. Stop words are deemed irrelevant for searching purposes because they occur frequently in the language for which the indexing engine has been tuned.

1) Example Tool: Voyant Tools

In Voyant Tools you can select from pre-existing stopword lists in various languages or create your own. For tools that support stopword lists, you can begin by clicking on the options icon availed in the tool.

B. Stemming

Stemming attempts to remove the differences between inflected forms of a word, in order to reduce each word to its root form. For instance hears may be reduced to the root hear, to remove the difference between singular and plural in the same way that we removed the difference between lowercase and uppercase.

1) Example Tool

Many Online tools also available for stemming purpose. Like,

http://9ol.es/porter_js_demo.html



Fig. 1: Highlighted with green words as stemming
In Screen words highlighted with green indicate the words as stemming.

C. POS Tagging

A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like 'noun-plural'.

1) Example

<http://parts-of-speech.info/>

The online POS tagger is availed to our text.

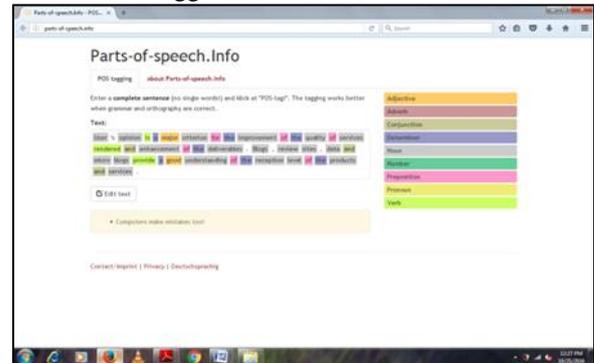


Fig. 2: Online POS tagger is availed to our text
In this, all the words shown in different color to differentiate the words such as noun, verb, adjective, etc

V. DIFFERENT LEVELS OF OPINION MINING [5]

1) Opinion Mining at Document level

The whole document is written about only one product and contains opinion posted by a single opinion holder. The results are presented using naive bayes, maximum entropy and support vector machine algorithms and good results are shown comparable to other ranging from 71 to 85% depending on the method and test data sets.

2) *Opinion Mining at Sentence level*

Different people, who have already used product, have written their opinions for product. A sentence contains only one opinion posted by single opinion holder; this could not be true in many cases e.g. there could be multiple opinions in compound. Secondly the sentence boundary is defined in the given document.

3) *Opinion Mining at Feature level*

The data source focuses on features of a single object posted by single opinion holder. All the features or attributes are separated and for particular feature the opinions are extracted.

VI. TOOLS USED IN SENTIMENT ANALYSIS

The tools which are used to track the opinion or polarity from the user generated contents are Web Fountain, Review Seer tool, Opinion observer, Red Opal.

- Web Fountain - It uses the beginning definite Base Noun Phrase (bBNP) heuristic approach for extracting the product features. It is possible to develop a simple web interface.
- Review Seer tool – This tool is used to automate the work done by aggregation sites. The Naive Bayes classifier approach is used to collect positive and negative opinions for assigning a score to the extracted feature terms. The results are shown as simple opinion sentence [2].
- Opinion observer-This is an opinion mining system for analyzing and comparing opinions [7] on the Internet using user generated contents. This system shows the results in a graph format showing opinion of the product feature by feature. It uses WordNet Exploring method to assign prior polarity.
- Red Opal –It is a tool that enables the users to determine the opinion orientations of products based on their features. It assigns the scores to each product based on features extracted from the customer reviews. The results to be shown with a web based interface [8].

VII. CONCLUSION

Opinion Mining has a wide range of application in summarizing reviews, classifying reviews, information system, market analysis and decision making. Sentiment analysis is a broad range of fields of natural language processing and text mining. It is found that different types of features and classification techniques are combined in an efficient way to enhance the sentiment classification. This survey is based upon the work supported by various authors but there are many improvements which can be made to the opinion mining application in terms of making use of further linguistic and contextual clues. So, the survey conducted by us is hopefully of significant use to help researchers in knowing about the past and recent trends in same area.

REFERENCES

- [1] Souza, Ellen, et al. "Characterizing Opinion Mining: A Systematic Mapping Study of the Portuguese Language." International Conference on Computational Processing of the Portuguese Language. Springer International Publishing, 2016.
- [2] G. Vinodhini and R. M. Chandrasekaran, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering , Volume 2, Issue 6, June 2012 ISSN: 2277 128X.
- [3] Balazs, Jorge A., and Juan D. Velásquez. "Opinion mining and information fusion: A survey." Information Fusion 27 (2016): 95-110.
- [4] K. Xu, S. S. Liao , J. Li, Y. Song, "Mining comparative opinions from customer reviews for Competitive Intelligence", Decision Support Systems 50,pp. 743–754, 2011.
- [5] Nidhi Mishra, C.K.Jha, PhD, Classification of Opinion Mining Techniques, International Journal of Computer Applications (0975 – 8887) Volume 56– No.13, October 2012.
- [6] Ravi, Kumar, and Vadlamani Ravi. "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications." Knowledge-Based Systems 89 (2015): 14-46.
- [7] Goyal, Ms Jayanti, and Ms Anjali Dadhich. "Procedure of Opinion Mining and sentiment analysis techniques: A survey." Expansion, Impact and Challenges of IT & CS (2015): 57.
- [8] ArtiBuche, Dr.M.B.Chandak, AkshayZadgoanakar "Opinion Mining and Analysis: A Survey", International Journal on Natural Language Computing (IJNLC) Vol 2 No 3 June 2013Pg No 39-48.
- [9] Bing Liu, "Sentiment Analysis: A Multi-Faceted Problem", IEEE Intelligent Systems, 2010.
- [10] Goyal, Ms Jayanti, and Ms Anjali Dadhich. "Procedure of Opinion Mining and sentiment analysis techniques: A survey." Expansion, Impact and Challenges of IT & CS (2015): 57.
- [11] Ganeshbhai, Solanki Yogesh, and Bhumika K. Shah. "Feature based opinion mining: A survey." Advance Computing Conference (IACC), 2015 IEEE International. IEEE, 2015.
- [12] Dave, Saurin, and Hiteishi Diwanji. "Trend Analysis in Social Networking using Opinion Mining A Survey." (2015).
- [13] Bhatia, Surbhi, Manisha Sharma, and Komal Kumar Bhatia. "Strategies for mining opinions: A survey." Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on. IEEE, 2015.
- [14] Martin, V. Maria Antoniate, K. David, and R. Bhuvaneswari. "A Survey on Various Techniques for Sentiment Analysis and Opinion Mining." Data Mining and Knowledge Engineering 8.3 (2016): 78-82.
- [15] Das, Sanjoy, et al. "Opinion based on Polarity and Clustering for Product Feature Extraction." International Journal of Information Engineering & Electronic Business 8.5 (2016).