Sentiment Analysis and Opinion Mining : A review of various applications in multivaried domains)

Yatharth Garg
B.Tech, Department of Computer Science,
THE NORTHCAP UNIVERSITY, Gurgaon, Haryana
ygarg020@gmail.com

Jyotika Pruthi
Assistant Professor, Department of Computer Science,
THE NORTHCAP UNIVERSITY, Gurgaon, Haryana
jyotikapruthi@ncuindia.edu

***Abstract*— Internet has become the voice of billions of users around the globe. People have strong opinions and views about everything and with the help of internet, it connects people to share and exchange ideas and review a product or a service. Reviews on the internet about a particular thing could be in millions which makes it difficult to understand a user's opinion and want. Sentiment analysis allows us to extract reviews of the user and guess their likings which could be beneficial for market research and product enhancements.**

Keywords - Opinion Mining, Sentiment Analysis, Product Reviews

I. INTRODUCTION

OPNION MINING (also called Sentiment Analysis) means identifying and classifying opinions of users and customers about a particular product or topic and also determines their attitude. These opinions are collected and examined using text analytics and computer linguistics. In some cases these opinions are considered more valuable than the product itself. Opinion Mining becomes very useful technique to draw a conclusion about a topic or product. Social networking sites, personal blogs and online rating sites have facilitated collection of opinions and sentiments of the users which help in making a further decision.

Usually, an opinion could be a positive or a negative remark. Example "That restaurant is really good"-positive remark. "The food at the restaurant was stale" – negative remark.

These days, reviews are also distinguished by a thumbs up and thumbs down, rating them on a scale. In the above example, opinion was based on just a single feature i.e., food. Let's see the following opinion-

"I went to a restaurant yesterday. When I entered, I was mesmerized by the ambience and the welcome by the staff. The menu had plenty of dishes which suited my taste. Some of the food ordered was stale and the prices were high compared to other places. The service was fairly fast and the playlist was also good."

Such type of opinion has described all the qualities of the product. This paper describes the different ways for sentiment analysis related to product or topic reviews.

II. LITERTAURE SURVEY

Sentiment Analysis has become a major part to analyze a product based on the reviews of the people. Some of the sources to offer a good understanding of the response of the products and services are-

A. E-Commerce Websites

E-Commerce websites have become a new breeding ground for people who want to share their views. These sites store millions of product reviews by the consumers. These sites allow the users to be completely blunt about their reviews on the products and the services provided.

Flipkart which started everything in India is a major factor. They are delivering from books to electronics to clothes. There is no such thing which they aren't delivering. The opinions given about the products are completely transparent. For any person, while making decision to purchase something, the opinions and decisions of others can be a beneficial factor. So every product can be rated and be commented by the consumer. Flipkart also ensures that the consumer review is not fake.

B. Social Networking Websites

Websites like Facebook, Twitter, and Instagram have become very popular and bring people closer from around the globe. People express their views publicly which can be viewed by every person interested in viewing the post. These websites have become so popular and addictive to not only common man but also celebrities and

politicians use them to express their opinions. Each site has its own specialty to allow people to express their emotions.

Facebook, being the most popular of them all, has over a billion users and allows users to express their opinions and views on any topic. People express their opinions using photos, videos, and wall posts. Any post which a user admires, they can either like it or comment on it to express their views.

Instagram which was developed by Facebook has been a popular mobile application which is used to share photos and videos to other users. Twitter on the other hand allows its users to show their opinions in 140 characters or less. Twitter has become a worldwide trend as it allows its user to connect with high-profile personalities directly. One of the main reasons behind Twitter's success was the # (hashtag) trend. People prefixed '#' to a word which could become a worldwide trending topic and anyone could read other people opinions on that topic. For e.g.-

"Today the parade was worth watching. #RepublicDay." Republic Day will become a trending topic if tweeted multiple times and people can also check the opinions of other users.

C. Working with Data Sets

Websites are providing data about different topics – restaurants, movies, books, colleges etc. Stupidsid.com is one such website which allows prospective graduate students to research about Master's from universities located in USA. They contain data about college reputation, finance, placements, living conditions etc.

imdb.com is an online location where one can research about the movies and review them. It even contains a list of top 250 movies of all time which has been rated by the users. The 'imdb ratings' are popular phrase which is used by many to review a movie.

Zomato – an online restaurant search and discovery service which is used for a similar purpose but for restaurants. Users share their opinions about particular restaurant and other people can view them.

III. SENTIMENT CLASSIFICATION

Sentiment analysis of user data usually judges the divergence of the user reviews. In such studies, sentiment analysis is done on three levels – attribute level, sentence level and document level. The techniques used are machine learning and semantic orientation.

A. Machine Learning

Machine learning basically employs three methods – Naive Bayes, support vector machines and maximum entropy classification. Other machine learning methods in the natural learning processing are:

K-Nearest Neighborhood, ID3, centroid classifies and the N-Gram model.

Machine learning techniques are basically better than human classifications from sentiment analysis. However the accuracy attained is much lesser when compared to topic-based classification. For example – "How can anyone listen to this song?", the statement contains no negative words. Hence, sentiment needs more clarification than the typical topic-based classification.

According to G.Vinodhini research, Naïve Bayes, commonly used algorithm for document classification is used to calculate the probabilities by using the combined probabilities of words and topics. Support Vector Machine is a text classification which typically outperforms the Naïve Bayes method. It searches for a decision surface to separate the training data points into two classes and makes decisions based on the support vectors.

The centroid classifier algorithm is very direct and easy. Firstly, a centroid vector or prototype vector for each class is computed. Then the comparison between the testing documents to the mean is calculated. Finally, it assigns to documents the label of the class of training samples whose centroid is closest to the observation.

The K-Nearest Neighbourhood is a type of instance-based learning, or lazy learning, in which all the calculation is postponed until the classification. It mainly depends upon the topic labels attached to the training documents comparable to the test document. The drawback of the method is that it is delicate to the local structure of the data.

B. Semantic Orientation

The Semantic Orientation for sentiment classification is also called "unsupervised learning" because it does not entail any previous drill for data mining. Instead it determines how much an opinion is inclined towards positive or negative. For example, few words which are synonyms for each other could differ in meaning as one could define interest and other won't mean the same.

The semantic orientation is somewhat less precise but is very useful in real-time applications. The results confirm that it is feasible to inevitably filter opinions from unstructured data.

In G.Vinodhini paper, in a review where an opinion is made, can't offer enough related data to determine the orientation of opinion, Chunxu Wu (2009) proposed an idea which resort to other reviews discussing the same topic to mine useful information, then use semantic similarity measures to judge the orientation of opinion. They attempted to tackle this problem by getting the orientation of context distinct opinions , then consider the

context reliant on opinions using linguistic rules to close orientation of context distinct-dependent opinion ,then extract contextual information from other reviews that remark on the same product feature to judge the context indistinct-dependent opinions.

C. Negative Opinions

Negation is a very useful verbal statement that can majorly affect the polarity of the opinion and should be used for sentiment classification. Reviews can be affected not only by the words (not, neither etc.) but also verbal elements.

For example – "Although the book is widely popular, but I fail to see why.", "The movie is rated poorly by the critics but is loved by the viewers." As seen from these examples, studying the negative opinions is a difficult task but is definitely very important aspect for sentiment classification.

According to the paper, Kennedy and Inkpen (2005) assess a negation model which is fairly similar to the one proposed by Polanyi and Zaenen (2004) in document-level polarity classification. A simple scope for negation is chosen. A polar expression is thought to be negated if the negation word immediately precedes it.

IV. CONCLUSION

Sentiment analysis or opinion mining has surfeit amount of applications in information systems, including classifying reviews, summarizing review and various real time applications. It is found that sentiment classifiers is usually dependent on topics and labels. Using different algorithms and classifications, the anomalies and benefits of others can help overcome drawbacks of other individuals.

Although, more work is also needed in future to enhance the performance of different algorithms for the sentiment analysis so that it can be used for more and new applications. The techniques and theorems are progressing rapidly, but still lot of problems still persists.

More future research could be dedicated to these challenges.

ACKNOWLEDGMENT

I would like to extend a note of thanks to my guide Ms.Jyotika Pruthi for her constant support and motivation.

REFERENCES

[1] G.Vinodhini, RM.Chandrasekaran "Sentiment Analysis and Opinion Mining: A Survey", Volume 2 Issue 6, June 2012, www.ijarcsse.com.

[2] A. Neesha Jebaseeli, E.Keerubakaran, "A Survey on Sentiment Analysis of (Product) Reviews", Volume 47, 2012, International Journal of Computer Applications (0975-888).