Techniques and Applications of Sentiment Analysis - Review

Sujata Laxman Sonawane and Pallavi V. Kulkarni
Government Engineering College, Aurangabad ,Maharashtra ,India.

ABSTRACT

People opinions and experiences are very valuable information in decision making process. The necessary content of reviews should be extracted and analyzed properly. This information is useful for manufacturer and consumers. Sentiment analysis is used to analyze each opinionated phrase or word and determine the sentiment orientation. And finally it gives the summarized opinion of a document or sentence.

© 2017 Elixir All rights reserved.

Introduction

With the advantage of Internet, a major change has been introduced in the life of common people and their decision. Before buying any product people like to check the reviews of different brands of a product of people who posted on site. So that they understand how the product is before buying it. During the decision making process in every field, it is very necessary to consider the customer's opinion. When customer's opinion is considered then organization get the important information such as what other people think. In the past, person asked opinion about any product to friends or family members, colleagues. Organizations conducted surveys and they analyze the document manually to understand the opinions of consumer. These public opinions are very important for businesses for marketing, customer relations.

As anyone can post their comments, reviews on site. These reviews are available publicly. A review is collected in the form of star or score rating then people can't understand the proper score. But if the review is presented in the form of open text then it is easy to expresses their opinion. But it is very difficult to read a number of reviews and take decision based on review. So it becomes very necessary to employ a method to analyze the feedback of users and overall polarity of reviews. System is needed here to extract and summarize the opinions. Therefore automated sentiment analysis systems are needed.

Sentiment Analysis

Sentiment means a view or judgment, approach, thought or emotion about something. Sentiment analysis term is referred to the extraction of speakers' opinion in given source material by using natural language preprocessing, linguistic compilation and text mining [5]. Sentiment analysis or opinion mining is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes.

Difference between Sentiment Analysis and Opinion Mining

Dave et al [13] has introduced the term opinion mining. According to them an ideal opinion mining tool should execute the major tasks as

i) To generate a list of product attributes in terms of quality features etc. by processing a set of search result for a given item.

ii) To aggregate the opinions about each of them like good, poor or mixed.

The term "sentiment" is the automatic analysis of text and summarizes the decision. Opinion mining is for finding the opinion of the text. For e.g. Jay felt the product was below expectation. So, the opinion of Jay about the product is that it is below expectation.

The rest of the paper is organized as follows. In section 2, related work on sentiment analysis gathered. In section 3, we describe the different techniques and methods of sentiment analysis. In section 4 we discuss the applications of sentiment analysis. In section 5, we present our conclusion.

Literature Review

Prabhu has presented a lexicon based approach for discovering sentiments. They have done the sentiment analysis for twitter data. They provide a method for preprocessing, negation, stemming, and removal of stop words, identify emoticons and hash tags, and convert non-grammatical words to grammatical words. They conclude that a lexicon based approach is as good as the lexicon it uses [1].

Khan used the Sentiwordnet to extract the semantic scores of sentences. They used the rule based domain independent method. The proposed method outperforms the machine learning methods. It achieves an average accuracy of 87% at the feedback level and 83% at the sentence level for comments and 97% at feedback and 86% at the sentences for customer reviews [2].

Anil kumar present a lexical approach that will identify opinion expressed using short words or SMS words. They use semantic based approach to find users' opinion from short words or SMS words apart of the regular opinionated phrase It efficiently detects opinion using lexical analysis [3].

Sunidhi dwivedi proposed an unsupervised dictionary based approach. Document level opinion mining system

performs 1) data preprocessing- remove unwanted data, sentence split, tokenization, POS tagging and stemming.
2) Extraction of opinion words- adjectives are extracted using Apache Open NLP tool and extracted adjective stored in text file. 3) Seed list preparation- extracted adjectives are matched with the seed list words to find their polarity. 4) Polarity detection 5) calculation of sentimental score. Paper compares the sentimental score of different mobiles [4].

Shailendra kumar singh shows which sentiment classification techniques used on what type of data set [5].

Arti buche has discussed that how the postings of social media have helped reshaping businesses. The main challenge is the prediction of stock market values to extract meaningful characteristics from news articles and decide which stock to buy, sell and hold. In this data collection is used for generating historic data, sentiment analysis techniques, prediction algorithms used to predict the stock movement. Stock prediction algorithm contains fetching numerical points which are up and down as per news sentences are fetched. Several preprocessing was done in order of how many times values went up and down in a specified range in historic data are checked. After checking it, next time it was either positive or negative polarity was assigned. Count the number of positive and negative words and result is predicted by considering the greater counter value [6].

Akanksha dubey has discussed the opinion summarization method. It consists of three tasks-1) to extract attributes of an item and to recognize opinions that are associated with the product attributes in each sentiment. 2) Recognize the opinion polarization. 3) Produce list according to the feature opinion pairs as its summary [7].

Raksha Sharma predicts the sentiment orientation of a text by analyzing the polarity of words present in the text. Sentiment lexicons are of two types: universal and domain specific. Paper focuses on finding sentiment words for movie domain with their polarity as positive, negative. In this research, finding the movie domain specific polar words is an important task. It gives input to chi-square test [8].

Chetashri et al emphasis on various methods of classifying a given piece of natural language text according to the opinions expressed in it.
It discusses the lexical and machine learning approach also variants to lexical approach. The proposed system implemented a set of technique for aspect classification and polarity identification. The SVM and Naive Bayes are the classification methods used here.

As a result it is found that prepending 'not' step, removing stopwords and stemming in preprocessing increases accuracy of system 2.44%, 4.89%, 12.2% respectively & achieved 78.05% accuracy [14].

Sentiment Analysis Techniques
Sentiment Analysis has different techniques
1) Lexicon Based
2) Machine Learning Based.
3) Hybrid Technique.

Lexicon Based
Sentiment lexicon contains a list of words which is used to express people's feelings and opinion. For example, start with positive and negative word lexicons, analyze the document. It is positive document, if it has more positive word lexicons, otherwise it is negative. The basic steps of the lexicon based techniques:
1. Pre-processing (i.e. tokenization, stemming, POS tagging, remove unwanted content)
2. Phrase extraction
3. Initially total sentiment score is 0.
4. Find the sentiment score of each word.
5. Finally summarize the score.
6. Document classification positive, negative or neutral.
Sentiment lexicon is constructed by using further three methods:
1. Manual construction of sentiment lexicon is difficult and time consuming.
2. Dictionary based approach: In this first collect a small set of opinion words manually, whose orientation is known. Then expand this set by searching in the WordNet dictionary for synonyms and antonyms. The words which found new are added to the seed list. It shares the same orientation for synonyms and opposite orientation for antonyms. Instead of WordNet, one can use SentiWordNet.
3. Corpus based approach. Corpus based method includes seed list of opinion words followed by finding other opinion words in turn helping in finding opinion words which are context specific in a larger corpus. It needs very large labeled training data. It can produce opinion words with high accuracy.

Machine Learning
It is applicable to supervised classification. It is a system capable of integrating and acquiring the knowledge automatically. To test the knowledge acquired, interpret and analyzed a corresponding knowledge.

Table 1. Applications of Sentiment Analysis.

Sr. No.	Area of Application	Description
1	Search Engine Summary Statistics	In search engine, user wants to specify the topic and rating of the desired reviews. Average semantic orientation sorts the search results. Thus the user can easily sample the most extreme reviews [9].
2	Tool for Academic Referees	When reviewing journal and conference papers.
3	E-commerce Activities	Consumer permits to submit their opinion about qualities of product and services. A summary of the product is provided by assigning ratings [7].
4	Entertainment	It helps people to choose which movie or series to watch [7].
5	Policy	Policy makers can take the point of view various policies and this information can be utilized in creating better citizen friendly policies [7].
6	Marketing	Opinion helpful to save money and time for businesses.
7	Ad campaign	Tracks and judges the success rate or launch of new product [10].
8	Voice of Market	Determine the opinion of customers about the product or services [12].
9	Education	Help students to determine which University or Institute is good for studies [7][12].
10	Expert Finding System	Used to track the literary reputations [11].

The most commonly used features in sentiment classification are given below:

Term Presence and their frequency: It includes unigrams or n-grams and their frequency or presence. These features widely and successfully used in sentiment classification.

POS tagging: Part of speech is used to disambiguate sense which in turn is used to guide feature selection. In POS tagging, each term in sentences will be assigned a label which presents its position role in the grammatical context. For example, with POS tags, we can identify adjectives and adverbs which are usually used as sentiment indicators.

Negations: Negation is an important feature to take into account since it has the potential of reversing a sentiment.

Opinion words and phrases: It expresses the positive or negative sentiments.

Hybrid Technique

Research indicate that the combination of both the machine learning and lexicon based approaches improve sentiment classification performance. The main advantages of hybrid technique: 1) readability from a carefully designed lexicon 2) high accuracy from a supervised learning algorithm.

Applications of Sentiment Analysis

There are several applications of the sentiment analysis in the various fields. There are many applications which are not discussed here.

Conclusion

Sentiment Analysis is an emerging field of data mining. It has a wide variety of applications. Sentiment analysis helps businesses to improve the quality of product. Sentiment Analysis is dependent on domain, approach selected, the classification algorithms and summarization method. This paper discusses about an overview of Sentiment Analysis, related work of different researchers, its various approaches and applications sentiment analysis.

References

[1] Prabhu Palanisamy, Vineet Yadav and Harsha Elchuri "Serendio: Simple and Practical lexicon based approach to Sentiment Analysis"

[2] Aurangzeb Khan, Baharum Baharudin, "Sentiment Classification by Sentence Level Semantic Orientation using SentiWordNet from Online Reviews and Blogs", IJCSET, Vol.2, No. 4, August, 2011, pp. 539-552.

[3] Anil Kumar K. M, Suresha, "Lexical Analysis to Effectively Detect Users' Opinion", IJWesT, Vol. 2, No. 4, October 2011, pp. 105-113.

[4] Sunidhi Dwivedi, Shama Parveen, "Document Level Opinion Mining of Reviews of Mobile Phone Companies", IJIRSET, Vol. 5, Issue 6, June 2016, pp. 9718-9725.

[5] Shailendra Kumar Singh, Sanchita Paul, Dhananjay Kumar, "Sentiment Analysis Approaches on Different Data set Domain: Survey", IJDTA, Vol. 7, No. 5, 2014, pp. 39-50.

[6] Arti Buche, Dr. M. B. Chandak, "Stock market prediction using text opinion mining: A survey", IJARCSSE, Vol. 6, Issue 6, June 2016, pp. 566-569.

[7] Akanksha Dubey, Nikhil Chitre, Vasundhara Ghate, "A Survey on opinion mining", IJIRCCE, Vol. 4, Issue 6, June 2016, pp.11910-11915.

[8] Raksha Sharma, Pushpak Bhattacharyya, "Detecting domain dedicated polar words".

[9] Peter D. Turney, "Thumbs up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 40th ACL, Philadelphia, July 2002, pp. 417-424.

[10] Arti Buche, Dr. M. B. Chandak, Akshay Zadgaonkar, "Opinion Mining and Analysis: A Survey"

[11] Mohini Chaudhari, Sharvari Govilkar, "A Survey of Machine Learning Techniques for Sentiment Classification"

[12] Anu Maheshwari, Anjali Dadhich, Dr. Pratistha Mathur, " Opinion Mining: A Survey"

[13] Kushal Dave, Steve Lawrence, and David M. Pennock. "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", In *Proceedings of WWW*, pages 519–528, 2003.

[14] Chetashri Bhadane, Hardi Dalal, Heenal Doshi, "Sentiment Analysis: Measuring Opinions",ICACTA-2015 Elsevier, pp. 808-814.