



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

A Survey on Opinion Mining

Akanksha Dubey¹, Nikhil Chitre², Vasundhara Ghate³

B.E, Department of Information Technology, MAEER's MIT College of Engineering, Kothrud, Pune, Savitribai Phule Pune University, Pune, India^{1,2}

Assistant Professor, Department of Information Technology, MAEER's MIT College of Engineering, Kothrud, Pune, Savitribai Phule Pune University, Pune, India³

ABSTRACT: Over the last few years, due to the increase in growth and use of Internet, share of user's opinion has increased which has led to the increase in inspiration towards opinion mining. Opinion Mining is crucial for both individuals and companies. Individuals like to see the opinions provided by other customers about a particular product or service. Companies need to analyze the feedback of its customers in order to improve its future decisions. Reviews on the Internet about a particular topic can be in millions which makes it tedious to understand a user's opinion and wants. Sentiment analysis allows us to extract reviews and present the summary which could be beneficial for market research and product enhancement. In this paper, we do a survey of Opinion Mining, covering various methods in sentiment analysis and challenges that appear in this field.

KEYWORDS: Opinion, Sentiment Classification, Opinion Mining, Sentiment Analysis.

I. INTRODUCTION

Opinions are statements that reflect people's sentiment or perception or opinion on objects or events. These are subjective statements. Opinion Mining or Sentiment analysis is a type of natural language processing for tracking the mood of the customers about a particular product or topic. It involves building a system to collect and examine opinions about the product made in blog posts, forums, tweets, or comments. It is widely studied in Data Mining, Text Mining and Web Mining. **Text mining** refers to the process of deriving high-quality information from **text** and is used in different fields like machine learning, computational linguistic, information retrieval, statistics, and data mining to form mining algorithms. **Web mining** is a sub part of text mining that is used to mine the semi structured web data in form of web structure mining, web content mining and web usage mining. Opinion mining is very useful to know to get the answers for a variety of questions such as **Which camera should I buy? Which film should I watch? Which mobile should I buy? Which book can I purchase? In which University should I take admission?** As a result of which, a person is no longer dependent on their relatives or friends. In general, opinion mining helps to gather information about the positive and negative viewpoints of a particular topic. Finally, only the positive and highly scored opinions obtained about a particular item are recommended to the customer. In order to enhance their marketing, large companies are making use of opinion mining.[2]

II. RELATED WORK

Our work is partially based on and closely related to opinion mining and sentiment classification. Extensive research has already been done on analysis of sentiment of review text and the analysis of subjectivity (deciding whether a statement is objective or subjective). Another area is related to it is feature based analysis of sentiment, in which sentiments on particular features of a product are decided. The work basically concentrates on discovering the sentiments associated with a statement. Opinion summarization consists of three basic tasks. The first task is to extract the attributes of an item and to recognize opinions that are associated with product attributes in each statement and then recognize the opinion polarization. Finally produce list according to the feature-opinion pairs as its summary [10]. Various [7], [11], [12] research work on feature-based sentiment analysis have exploited numerous ways for extraction and refinement of features, which NLP and rule-based methods, ontology-based methods and statistical methods. Yet another proposed method is to mine features from review data using association rule mining method.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

III. OPINION MINING and SENTIMENT ANALYSIS

Opinion mining is a system which is used to identify and extract subjective information in text documents. Basically, opinion mining or sentiment analysis tries to analyze the opinion of a reviewer about some aspect and also the overall contextual polarization of a text document. The sentiment may be his or her judgment, mood or evaluation. A key problem in this area is in sentiment classification, where a text is labeled as a positive or negative evaluation of a target object (movie, book, item etc.). The evaluation of sentiment can be done in two methods:

1. **Direct opinions** are the opinions that gives positive or negative opinion about the product directly. “The features of this mobile are bad”, can be called as an example of direct opinion.
2. **Comparison Opinions** are the opinions that are meant to compare the object with some other alike objects. For example, “The features of mobile A is better than that of mobile B.” expresses a comparison.

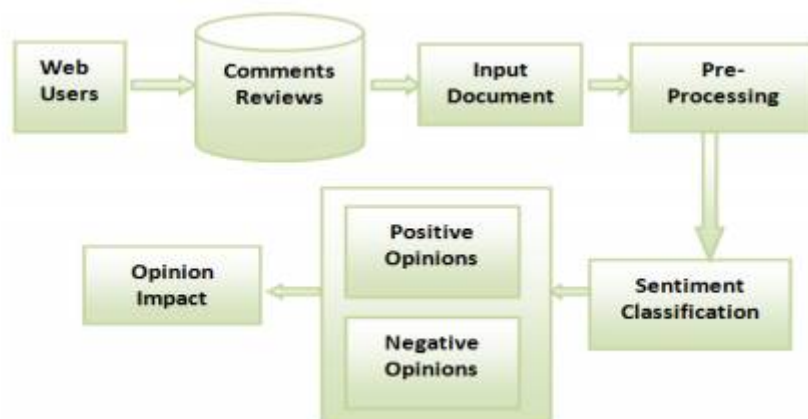


Fig. 1 Workflow of Opinion Mining

The above figure shows the workflow of Opinion Mining about how the opinions are being retrieved from people review over the comment provided by them. Opinion feature extraction is a sub discipline of opinion mining with the vast majority of existing work done in the product review domain.

Data Mining can be seen as the analysis step of the KDD process and the entire process is dependent on it. Its aim is to retrieve knowledge from huge amount of data in an understandable form that is useful for various industries and individuals.

Web mining is the method of applying data mining procedures for analyzing patterns from the Web. Web usage mining, web content mining and web structure mining are three different types of web mining.

1. Web Usage Mining

Web usage mining is the method of determining what users want to view on the Internet. Some users show interest in textual data whereas others in multimedia data. This is mainly done by making use of logs of the user.

2. Web Structure Mining

Web structure mining is the process that is used to identify the relation between Web pages that are linked by information or direct link connection.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

3. Web Content Mining

Web content mining aims to retrieve useful information from contents of the pages of the web. It involves examining of all the contents on a web page to discover its significance with the search query.

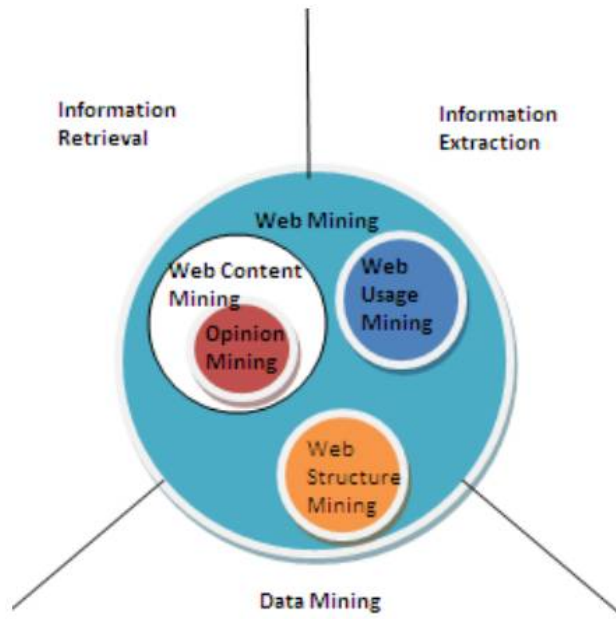


Fig. 2 Data Mining

Opinion Mining is part of web content mining. Figure above (Fig. 2) shows this categorization clearly.

DEFINITION: “If a collection of text documents (T) are given, that have opinions on an item, opinion mining intends to recognize attributes of these items on which opinion have been given, in each of the document $t \in T$ and to find polarization of the reviews i.e. whether the reviews are negative or positive”. [5]

IV. SENTIMENT CLASSIFICATION

Sentiment analysis of user data usually judges the difference of opinion of the user reviews. In such studies, sentiment analysis is mostly done on three levels – phrase level, sentence level and document level. The methods used are machine learning as well as semantic orientation.

1. Document Level

Document level sentiment classification is performed on the overall sentiments expressed by writer. Documents are categorized according to the opinions instead of topic. It is to sum up the entire document as positive or negative polarity about any object (electronic, vehicle, movie, and books etc).

2. Sentence Level

Sentence level sentiment classification models retrieve the sentences that contain opinionated objects, opinion holder and opinionated terms. It is a level deeper to document level and concerns to the opinionated terms but not the features. Numbers of negative and positive terms are counted from sentences. If positive terms are highest then opinion about the given object is positive and if the negative words are highest then opinion is negative else it is neutral.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

3. Phrase Level

The phrase level sentiment classification is a more Pinpointed advance to opinion mining. The phrases that contain opinion terms are discovered and a phrase level categorization is carried out. But in other cases, where contextual polarization also matters, the outcome may not be totally accurate. Negation of terms can occur locally, but if there occurs statements with presence of the negating terms which are unlike the opinion terms, phrase level analysis is not recommendable. The process is recognizing Opinion Words, the role of negation terms and Clauses.

A. Machine Learning

- Machine learning basically makes use of three methods – Naive Bayes, maximum entropy classification and support vector machines. Some other machine learning techniques in the natural learning processing are as follows: N-Gram model, K-Nearest Neighborhood and centroid classifiers.
- Machine learning methods are basically better than human classifications for analysis of sentiments. However, the accuracy attained is lesser when contrasted to topic-based classification. Consider the example, “How can anyone read to this book?”, the statement contains no negative terms. Hence, the opinion needs more elucidation than the common topic-based classification.
- In the G. Vinodhini research paper, Naïve Bayes, a commonly used algorithm for document categorization is used to compute the probabilities by using the collective probabilities of topics and words. Support Vector Machine is a text categorization which outperforms the Naïve Bayes technique. It searches for a decision surface to split the training data points into two categories and makes decisions based on the given support vectors.
- The K-Nearest Neighborhood is a sort of instance-based learning, or lazy learning, in which all the computation is postponed until the categorization. It mostly depends upon the topic labels that are attached to the training documents in comparison to the test document. The drawback of the technique is that it is delicate to the local organization of the data. [1]
- The centroid classifier algorithm is direct and simple. Firstly, a centroid vector for each class is calculated. Then the association between the testing documents to the mean is computed. Finally, it allocates to documents the label of the class of training samples whose centroid is nearest to the observation.

B. Semantic Orientation

- Semantic Orientation for sentiment categorization is also known as “unsupervised learning” as it does not entail any previous data mining drill. Instead it establishes how much an opinion is inclined towards being positive or negative. Consider an example, few terms which are synonyms to each other could vary in meaning as one could describe interest and other describes some other emotion.
- The semantic orientation is somewhat less accurate but is very useful in real-time applications. The outcomes confirm that it is feasible to inevitably distinguish opinions from data that is unstructured.
- In G.Vinodhini paper, in a review where an opinion is given, it can't offer enough related data to establish the orientation of the given opinion. Chunxu Wu (2009) came up with an idea which resorts to other reviews discussing the same topic to extract useful information, and then use semantic similarity measures to evaluate the orientation of given opinions.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

V. ARCHITECTURE OF OPINION MINING

Opinion Mining also known as sentiment analysis is a method of finding user's opinion towards a product or a topic. Opinion mining summarizes whether user's view towards the product is positive, neutral, or negative about item, event, topic etc. Opinion mining and summarization process is made of three important steps: Opinion extraction, Opinion categorization and Opinion summarization. The review sentence is retrieved from review websites. Opinion text can be found in blogs, reviews, comments, tweets etc, which contains subjective information about the topic. Reviews can be classified as negative or positive review. Opinion summary is then generated based on features in the opinion sentences by considering common features about a topic.



Fig. 3 Opinion Mining

VI. APPLICATIONS

- 1) It is mostly used in E-commerce activities. When any customer buys any item or service from the e-commerce websites, then it permits them to submit their opinions about qualities of shopping services and products. A summary for the product and various features of the product is provided by assigning ratings.
- 2) It is used in Entertainment by helping people to choose which movie or series to watch.
- 3) Sentiment analysis can be used by policy makers who can take the point of view of the citizen's towards various policies and this information can be utilized in creating better citizen friendly policies.
- 4) It is also used in Marketing. Nowadays, each company makes available the facility to its users to provide opinions about its products and services. Hence, it is helpful for businesses to save money as well as time because there is no need any more to conduct surveys as the feedbacks related to all the products are available on their sites.
- 5) It is also used in education domain, to help students to determine which university is good for studies.

VII. RESEARCH CHALLENGES

There are various challenges in Sentiment analysis. A few of them are discussed in this paper.

- 1) The very first challenge is "opinion word" which can be considered to be positive in one way but may be considered negative in another way.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

- 2) Second challenge is that sometimes user may convey their sentiments in an unusual way. The text in a sentence can be difficult to identify as ironic or sarcastic and this can lead to faulty polarization and misleading sentiment analysis. Reference [8] discusses this problem.
- 3) The third challenge is the language i.e, the majority of the work done in opinion mining is focused on two languages: English and Chinese and other languages needs to be explored.
- 4) Now, the fourth challenge is the sentiment given on twitter is difficult to comprehend as it consists of poor abbreviations, lack of capital letters, spelling mistakes, no proper punctuations, and grammatical errors and so on.
- 5) The fifth challenge, lies in the issue that the opinion of the reviewer changes over time. A research work is carried out to recognize how the mood of the people varies over time in Reference [9].The research done observes sites where the mood of the reviewer is clearly specified either by choosing from a given list of moods or by typing it as free text sentence.
- 6) Sixth challenge is in “detection of spam and fake comments, mainly through the recognition of duplicates, the association of qualitative with summary feedbacks, the recognition of outliers, and also the reputation of the reviewers”.

VIII. CONCLUSION

This paper mainly focuses on survey of opinion mining. Opinion mining or sentiment analysis has a surplus amount of applications in the information systems, which includes classification of reviews, their summarization and a variety of real-time applications. Researches have been conducted to mine the opinions in the form of document level, sentence level or feature level sentiment analysis. It is seen that opinion mining can now be used to extract the sentimental reviews on twitter, Facebook comments on pictures, videos or even statuses. Thus we have discussed about the overview of challenges in opinion mining and the various applications of sentiment analysis. More future research can be done on these challenges and more work can be done to overcome these challenges in future.

REFERENCES

- [1] Ytharth Garg, Jyotika Pruthi “Sentiment Analysis and Opinion Mining: A review of various applications in multivaried domains”, International Journal of Computer Science & Engineering Technology (IJCSSET), Vol. 7 No. 03 Mar 2016.
- [2] Anu Maheshwari, Anjali Dadhich, Dr. Pratistha Mathur “Opinion Mining: A Survey”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 1, January 2015
- [3] P.Kalarani, Dr.S. Selva Brunda “An Overview on Research Challenges in Opinion Mining and Sentiment Analysis”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 10, October 2015
- [4] S. ChandraKala, C. Sindhu2 “OPINION MINING AND SENTIMENT CLASSIFICATION: A SURVEY”, ICTACT Journal on soft computing, Volume: 03, Issue: 01, October 2012
- [5] Bakhtawar Seerat, Farouque Azam “Opinion Mining: Issues and Challenges (A survey)”, International Journal of Computer Applications (0975 – 8887)Volume 49– No.9, July 2012
- [6] G.Vinodhini, RM.Chandrasekaran “Sentiment Analysis and Opinion Mining: A Survey”, Volume 2 Issue 6, June 2012.
- [7] Bing Liu. 2010. Sentiment Analysis: A Multi-Faceted Problem.Neesha Jebaseeli, E.Keerubakaran, “A Survey on Sentiment Analysis of (Product) Reviews” , International Journal of Computer Applications (0975-888), Volume 47, 2012.
- [8] Zhongwu Zhai, Bing Liu, Identifying Evaluative Sentences in Online Discussions, 2011.
- [9] Krisztian Balog, Maarten de Rijke. 2006.Decomposing Bloggers' Moods - Towards a Time Series Analysis of Moods in the Blogosphere.
- [10] Khairullah Khan, Baharum B. Baharudin, Aurangzeb Khan, Fazal-e-Malik, “Mining Opinion from Text Documents: A Survey”, 3rd IEEE International Conference on Digital Ecosystems and Technologies, 2009.
- [11] M.Hu and B. Liu, “Mining and summarizing customer reviews,” in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04)*, pp. 168–177, August 2004.
- [12] B. Liu, M. Hu, and J. Cheng, “Opinion observer: analyzing and comparing opinions on the web,” in *Proceedings of the International World Wide Web Conference Committee (IW3C2 '05)*, Chiba, Japan, pp.10–14, May 2005.