ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

An Overview on Research Challenges in Opinion Mining and Sentiment Analysis

P.Kalarani[1], Dr.S. Selva Brunda[2]

[1]Assistant Professor, Dept. of CT and IT, Kongu Arts and Science College (Autonomous), Erode, Tamilnadu, India

[2] Professor and Head, Dept of CSE , Cheran College of Engineering, Karur, Tamilnadu, India

ABSTRACT: Public opinion is the belief or thoughts of the public regarding a particular topic, especially one regarding politics, religion or social issues. Opinions may be sensitive since they may reflect a person's perspective, understanding, particular feelings, way of life, and desires. People's opinions and experience are very valuable information in decision making process. Now a day's several websites encourage users to express and exchange their views, suggestions and opinions related to product, services, polices, etc. publically. The increased popularity of these sites resulted huge collection of people opinion on the web in much unstructured manner. Extracting the useful content from these opinion sources becomes a challenging task. This situation created a new area of research called opinion mining and sentiment analysis. Opinion mining and sentiment analysis extract and classify the people's opinion automatically from the internet. This survey paper discusses various application and challenges related to the Opinion Mining and Sentiment Analysis.

KEYWORDS: Opinion mining;Sentiment analysis ;Websites; Challenge;, Applications.

I. INTRODUCTION

This era is of automated systems and digital information every field of life is evolving rapidly and generating data. As a result huge volumes of data produce in field of science, engineering, medical, marketing, finance, demographic etc . Automated systems are needed to automate analysis, summarization, and classification of data. It also helps at enterprise level to take related decisions. Multiple research fields like statistics, machine learning, artificial intelligence and visualization are involved to develop such automated systems .A number of efficient ways are available to store the huge volumes of data, computational techniques and models are required to extract the hidden patterns and knowledge. These techniques and tools are used to transform the data into useful information, to make market analysis, fraud detection and find the customer intentions etc. Such computational tools and techniques are the subject of **Knowledge Discovery in Database and Data Mining**.

Fig. 1.Data Mining Hierarchy

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

Text mining is an interdisciplinary method used in different fields like machine learning, information retrieval, statistics, computational linguistic and data mining to form mining algorithms. Some researchers defined text mining as tool to discover the new knowledge from huge volume of natural language text using computational algorithms. Web mining is a sub discipline of text mining used to mine the semi structured web data in form of web content mining, web usage mining and web structure mining.

II. OPINION MINING AND SENTIMENT ANALYSIS

Opinion mining is a technique which is used to detect and extract subjective information in text documents . In general, sentiment analysis tries to determine the sentiment of a writer about some aspect and also the overall contextual polarity of a document. The sentiment may be his or her judgment, mood or evaluation. A key problem in this area is sentiment classification, where a document is labelled as a positive or negative evaluation of a target object (film, book, product etc.)The evaluation of opinion can be done in two ways:

Direct opinion gives positive or negative opinion about the object directly. For example, "The picture quality of this camera is poor" expresses a direct opinion.

Comparison Opinion means to compare the object with some other similar objects. For example, "The picture quality of camera-y is better than that of Camera-x." expresses a comparison.

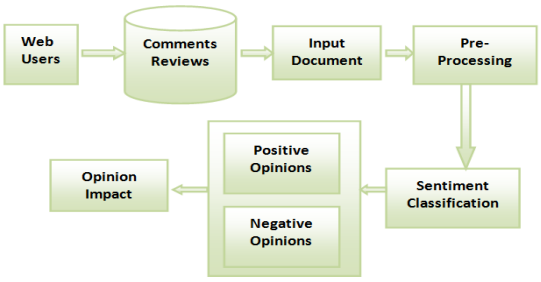

Fig. 2. Workflow of Opinion Mining

Fig. 2 have a workflow of Opinion Mining of how the opinions are being extracted from people review over their comment Opinion feature extraction is a sub problem of opinion mining with the vast majority of existing work done in the product review domain.

III. SENTIMENT CLASSIFICATION

3.1 Document level

Document level sentiment classification executed on the overall sentiments expressed by authors. Documents classified according to the sentiments instead of topic. It is to summarize the whole document as positive or negative polarity about any object (mobile, car, movie, and politician etc).

3.2 Sentence level

Sentence level sentiment classification models extract the sentences contains opinionated terms, opinion holder and opinionated object. It is one level deep to document level and just concerns to the opinionated words but not the features. Number of positive and negative words counted from sentences if positive words are maximum then opinion about object is positive and if the negative words are more than opinion is negative otherwise neutral.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Vol. 3, Issue 10, October 2015

3.3 Phrase level Opinion Mining

The phrase level sentiment classification is a much more Pinpointed approach to opinion mining. The phrases that contain opinion words are found out and a phrase level classification is done. But in some other cases, where contextual polarity also matters, the result may not be fully accurate. Negation of words can occur locally. But if there are sentences with negating words which are far apart from the opinion words, phrase level analysis is not desirable. The process is Identifying Opinion Words, the role of negation words and Clauses .

IV. ARCHITECTURE OF OPINION MINING

Opinion Mining also called sentiment analysis is a process of finding user's opinion towards a topic or a product. Opinion mining concludes whether user's view is positive, negative, or neutral about product, topic, event etc. Opinion mining and summarization process involve three main steps, first is Opinion Retrieval, Opinion Classification and Opinion Summarization. Review Text is retrieved from review websites. Opinion text in blog, reviews, comments etc. contains subjective information about topic. Reviews classified as positive or negative review.

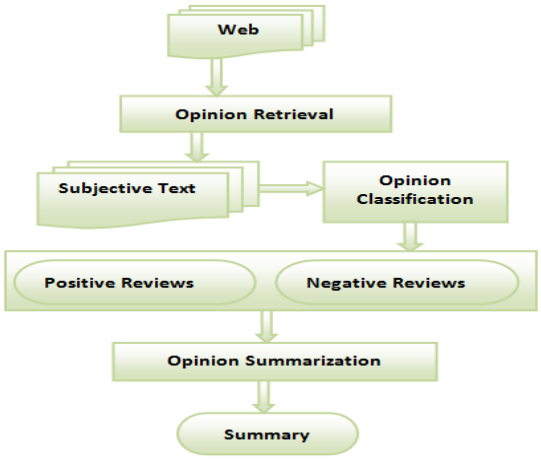

Fig.3. Architecture of Opinion Mining

Opinion summary is generated based on features opinion sentences by considering frequent features about a topic.

V. TECHNIQUES OF OPINION MINING

Major data mining techniques used to extract the knowledge and information are: generalization, classification, clustering, association rule mining, data visualization, neural networks, fuzzy logic, Bayesian networks, and genetic algorithm, decision tree. Figure 5 has the techniques of Opinion Mining.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Vol. 3, Issue 10, October 2015

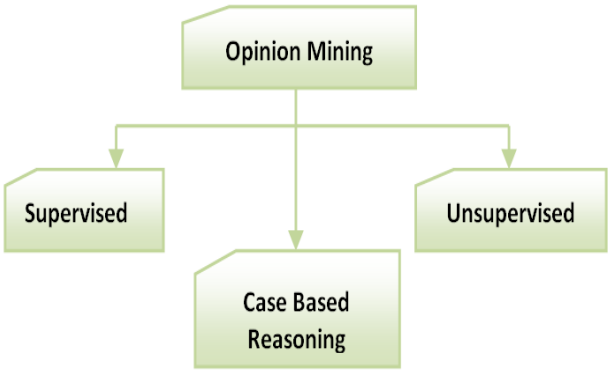

Fig 4. Techniques of Opinion Mining

5.1 Supervised Machine Learning

Classification is most frequently used popular data mining technique. Classification used to predict the possible outcome from given data set on the basis of defined set of attributes and a given predictive attributes. The given dataset is found to be the training dataset consist on independent variables (dataset related properties) and a dependent attribute (predicted attribute). A training dataset created model test on test corpus contains the same attributes but no predicted attribute. Accuracy of model checked that how accurate it is to make prediction. Product features and sentenced words are extracted using Double Propagation Algorithm.

5.2 Unsupervised Learning

In contrast of supervised learning, unsupervised learning has no explicit targeted output associated with input. Class label for any instance is unknown so un supervised learning is about to learn by observation. Clustering is techniques used in unsupervised learning. The process of gathering objects of similar characteristics into a group is called clustering. Objects in one cluster are dissimilar to the objects in other clusters.

5.3 Case Based Reasoning

Case based reasoning is an emerging Artificial Intelligence supervised technique. CBR is a powerful tool of computer reasoning and solve the problems (cases) in such a way which is closest to real time scenario. It is a problem solving technique in which knowledge is personified as past cases in library and it does not depend on classical rules. The solutions are stored in CBR repository called Knowledge base or Case base.

VI. TOOLS USED IN OPINION MINING

The tools which are used to track the opinion or polarity from the user generated contents are:

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 .
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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

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.

Opinion observer-This is an opinion mining system for analyzing and comparing opinions 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.

VII. APPLICATION AREAS OF OPINION MINING AND SENTIMENT ANALYSIS

1) **Purchasing Product or Service:** While purchasing a product or service, taking right decision is no longer a difficult task. By this technique, people can easily evaluate other's opinion and experience about any product or service and also he can easily compare the competing brands. Now people don't want to rely on external consultant. The Opinion mining and sentiment analysis extract people opinion form the huge collection of unstructured content, the internet, and analyze it and then present to them in highly structured and understandable manner.

2) Marketing research: The result of sentiment analysis techniques can be utilized in marketing research. By sentiment analysis techniques, the recent trend of consumers about some product or services can be analyzed. Similarly the recent attitude of general public towards some new government policy can also be easily analyzed. These all result can be contributed to collective intelligent research.

3) Opinion spam detection: Since internet is available to all, anyone can put anything on internet, this increased the possibility of spam content on the web. People may write spam content to mislead the people. Opinion mining and sentiment analysis can classify the internet content into' spam' content and 'not spam' content.

4) Policy Making: Through Sentiment analysis, policy makers can take citizen's point of view towards some policy and they can utilize this information in creating new citizen friendly policy.

5) Decision Making: People's opinion and experience are very useful element in decision making process. Opinion mining and Sentiment analysis gives analyzed people's opinion that can be effectively used for decision making.

VIII. RESEARCH CHALLENGES

1) **Detection of spam and fake reviews:** The web contains both authentic and spam contents. For effective Sentiment classification, this spam content should be eliminated before processing. This can be done by identifying duplicates, by detecting outliers and by considering reputation of reviewer.

2) **Limitation of classification filtering:** There is a limitation in classification filtering while determining most popular thought or concept. For better sentiment classification result this limitation should be reduced. The risk of filter bubble [11] gives irrelevant opinion sets and it results false summarization of sentiment.

3) **Asymmetry in availability of opinion mining software:** The opinion mining software is very expensive and currently affordable only to big organizations and government. It is beyond the common citizen's expectation. This should be available to all people, so that everyone gets benefit from it.

4) **Incorporation of opinion with implicit and behavior data:** For successful analysis of sentiment, the opinion words should integrate with implicit data. The implicit data determine the actual behavior of sentiment words.

5) **Domain-independence:** The biggest challenge faced by opinion mining and sentiment analysis is the domain dependent nature of sentiment words. One features set may give very good performance in one domain, at the same time it perform very poor in some other domain.

6) **Natural language processing overheads:** The natural language overhead like ambiguity, co-reference, Implicitness, inference etc. created hindrance in sentiment analysis too.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

IX. RESEARCH SCOPE IN OPINION MINING AND SENTIMENT ANALYSIS

The major research scope areas in sentiment analysis are:

1) Spam Detection Sentiment Analysis
2) Sentiment Analysis on short Sentence like abbreviations
3) Improving sentiment word identification algorithm
4) Developing fully automatic analyzing tool
5) Effective Analysis of policy opinionated content
6) Successful handling of bi polar sentiments
7) Generation of highly content lexicon database

X. CONCLUSION

Opinion miming is an emerging field of data mining to extract the knowledge from huge volume of data that may be customer comments, feedback and reviews on any product or topic etc. Research has been conducted to mine opinions in form of document, sentence and feature level sentiment analysis It is examined that now opinion mining trend is moving to the sentimental reviews of twitter data, comments used in Face book on pictures, videos or Face book status. Thus this paper discusses about an overview of Challenges and application area in opinion mining and the techniques and tools used for opinion mining.

REFERENCES

[1] Pang, B. & Lee, L. (2008), "Opinion Mining and Sentiment Analysis". In Foundations and Trends in Information Retrieval.
[2] Hu, M. & Lui, B. (2004), "Mining and Summarizing Customer Reviews". In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2004 (KDD-2004).
[3] Ku, L.-W., Liang, Y.-T., & Chen, H, "Opinion extraction, summarization and tracking in news and blog corpora". In AAAI-CAAW‟06.
[4] Kunpeng Zhang, Ramanathan Narayanan, "Voice of the Customers: Mining Online Customer Reviews for Product", 2010.
[5] H. Tang, S. Tan and X. Cheng, " A survey on sentiment detection of reviews", In Proceedings of the Expert Systems with Applications 36, Elsevier Ltd., Beijing, 2009.
[6] Ortigosa-Hernandez Jonathan, Rodriguez Juan Diego, Alzate Leandro, Lucania Manuel, InzaInaki, Lozano, Jose," A. Approaching sentiment analysis by using semi-supervised learning of multi-dimensional classifiers"Neurocomputing;92:98-115,2012.
[7] Kim S, Hovy E," Determining the sentiment of opinions", In: Proceedings of international conference on Computational Linguistics (COLING'04) - 2004.
[8] Read J, carol J.," Weakly supervised techniques for domain independent sentiment classification", In: Proceeding of the 1st international CIKM workshop on topic-sentiment analysis for mass opinion; P.45-52, 2009.
[9] Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the association for computational linguistics.
[10] Burns, N., Bi, Y., Wang, H., & Anderson, T. (2011). Sentiment analysis of customer reviews: Balanced versus unbalanced datasets. In Knowledge-based and intelligent information and engineering systems. Lecture notes in computer science (Vol. 6881, pp. 161–170). Berlin/ Heidelberg: Springer.