ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE
International Journal of Advanced Research in Computer and Communication Engineering

ICACTRP 2017
International Conference on Advances in Computational Techniques and Research Practices
Noida Institute of Engineering & Technology, Greater Noida
Vol. 6, Special Issue 2, February 2017

Opinion Mining and Sentiment Analysis on Customer Review Documents- A Survey

Surya Prakash Sharma[1], Dr Rajdev Tiwari[2], Dr Rajesh Prasad[3]

Assistant Professor, Department of Computer Science & Engineering, NIET Gr. Noida, India[1]

Professor, Department of Computer Science & Engineering, GNIOT Gr. Noida, India[2]

Professor, Department of Computer Science, Yobe State University, Damaturu, Nigeria[3]

Abstract: People beings express themselves by giving opinions, feedback, suggestions or ideas about any object. Opinions can be expressed in many ways such as it can be expressed on twitter, facebook, reviews, blogs etc. World and technology is growing the faster rate so for taking decisions such as buying a product, voting for a politician etc. people are using opinion present on blogs, social networking sites or shopping websites. Nowadays, if someone has to buy a online product then he/she can first view other buyer's reviews on the product site and take the right decision accordingly. There are many opinions on the web about an product so taking decisions might be difficult. Therefore, opinion mining is used for classifying the reviews according to their polarity. Opinion mining (OM) is a process of the mining opinion from the review. Analyzing customers review is more important for any user in making right purchasing decision product and organization. Opinion mining is also known as sentiment analysis. In this paper we survey on Opinion mining with respect to their different levels, architecture, techniques applied, tools used, comparative study of techniques and challenges.

Keywords: Opinions, Polarity, Opinion Mining, Sentimental Analysis, Data Mining, Web Mining.

I. INTRODUCTION

Opinions are statements that reflect people's perception or sentiment. These statements also provides opinion on objects or events. Opinion Mining or Sentiment analysis is a task under natural language processing for finding the mood of the customers about a purchasing of a particular product or topic. It involves building a system to collect and examine opinions about the product made in many online purchasing sites. Opinion mining is a sub field of web content mining. Web content mining is branch of Data mining.

This paper is organized as follows: Section I Introduction, Section II Related Work, Section III Opinion mining and Sentimental analysis, Section IV Sentimental Classification, Section V Opinion Mining Techniques, Section VI Tools Used in opinion mining, Section VII Applications, Section VIII Research challenges and in the last Section Research Scope.

II. RELATED WORK

This paper proposed a framework for online opinion summarization. The inputs to the framework are product name, date and time and review of that product. The output is the summary of the review in concise manner. The system performs the summarization in three steps: (1) Product feature based, which is given by customer; (2)In each review, Identify frequent features in each opinion sentence (3) Finding out whether feature is opinionated and also identify an orientation of the words and finally summary will be created [1].

This paper, propose a novel method to identify opinion features from online reviews by exploiting the difference in opinion feature statistics across two corpora, one domain-specific corpus and one domain-independent corpus[2]. This review paper discusses existing techniques and approaches for feature extraction in sentiment analysis and opinion mining[3]. This paper gives an overview of opinion mining that has been done on Hindi language[4].

This paper present feature based opinion mining and their efficiency in term of precision, recall and accuracy[5].

In this paper first extracts the feature, modifier and opinion from the dataset and then using clustering mechanism divides them into discrete clusters by user's opinion[6]. This paper proposes a feature wise opinion mining system to determines the polarity of the opinions in reviews documents using Senti-WordNet[7].

III. OPINION MINING AND SENTIMENTAL ANALYSIS

Opinion mining is a system which is used to identify and extract subjective information in text documents. Opinion mining is also called Sentimental analysis. The sentiment

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE
International Journal of Advanced Research in Computer and Communication Engineering
ICACTRP 2017
International Conference on Advances in Computational Techniques and Research Practices
Noida Institute of Engineering & Technology, Greater Noida
Vol. 6, Special Issue 2, February 2017

may be his or her judgment, evaluation of product customer reviews.

The prediction of sentiment can be done by two methods:
1. Direct opinions: Text documents that give positive or negative opinion about the product directly. Example:"Batteries backup of this mobile is bad.
2. Comparison Opinions: Opinions in text document that are meant to compare the object with some other objects.

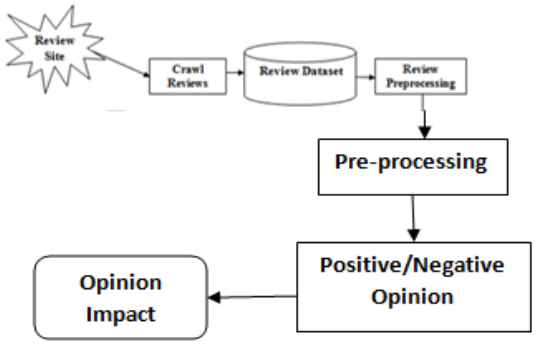

Figure1: Opinion Mining Work Flow

For example, "The features of object A is better than that of object B."

Data Mining is the analysis step of the KDD process and the entire process is dependent on it.

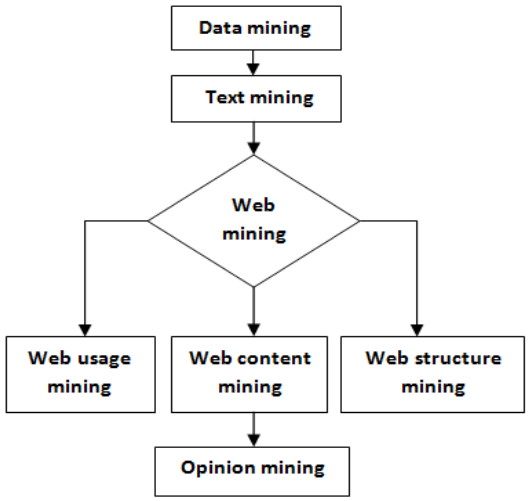

figure2: Data Mining and Its Classification

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

2. Web Structure Mining
It is the process that is used to identify the relation between Web pages that are linked by direct link connection or information.

3. Web Content Mining
Web content mining is the technique which retrieves useful information from contents of the web pages. It involves examining of all the contents on a web page and by using search query discover its significance.

Opinion Mining is part of web content mining. The figure2 shows this categorization clearly.

IV. SEMENTAL CLASSIFICATION

There are three types of level of opinion mining [2].
1). Document level Opinion Mining: They classified document according to sentiment instead of the topic. They contain in the form of forums or blogs, the document level analysis is not desirable. So subjectivity/objectivity classification is used in this type of classification.
2). Sentence level Opinion Mining:
In sentence level Opinion Mining, opinion word is extracted from sentence level and makes them as positive and negative.
3).Feature level Opinion Mining: The features that contain opinion words are found out, and a phrase level classification is done. It is also known as feature-level opinion mining.

V. OPINION MINING TECHNIQUES

There are mainly three types of techniques (fig. 3)**:**
Supervised Learning Techniques : The most widely used supervised learning techniques are Support Vector Machines(SVM), Neural network, Multi-Layer Perceptron (MLP), Decision tree, Naïve Bayes(NB) Classification, Maximum Entropy(MaxEnt).
2. **Unsupervised Learning Techniques:** Mostly used techniqueare Clustering algorithm, expectation-maximization algorithm, matrix factorization, principal component analysis.
3. **Case-Based Reasoning:** It is an emerging artificial techniques. CBR is an intelligent tool of computer reasoning and solves the problem in such a real time scenario. Solution is stored in CBR repository also known as case base.

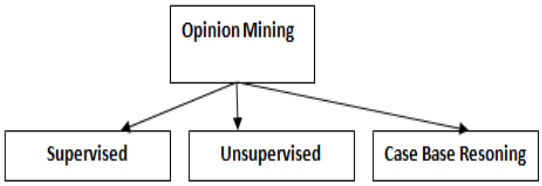

Figure 3: Types of Opinion Mining Techniques

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICACTRP 2017

International Conference on Advances in Computational Techniques and Research Practices

Noida Institute of Engineering & Technology, Greater Noida

Vol. 6, Special Issue 2, February 2017

VI. TOOLS USED IN OPINION MINING

Tool Name	Purpose
WEKA [10]	Data pre-processing, Classification, Regression, Clustering, Visualization, Association Rules, Machine learning algorithm for Data Mining,.
NLTK [11]	Provides lexical resources such as WordNet, Tagging, Parsing, Semantic reasoning, Classification, Tokenization, Stemming.
STANFORD CORENLP	POS tagging, Named entity recognizer, Parsing, Co reference resolution system, Sentiment analysis, Bootstrapped pattern learning
LingPipe [13]	Entity extraction, POS tagging, Clustering, Classification.
GATE[14]	Sentence Splitter, POS tagging, Tokenizer, Gazetteer, Named entities transducer, Coreference tagger
STANFORD CORENLP[[[12]	POS tagging, Named entity recognizer, Parsing, Coreference resolution system, Sentiment analysis, Bootstrapped pattern learning
Robust Accurate Statistical Parsing	Statistical Parser, Tokenization, Tagging, Lemmatization and Parsing
Red Opal	Machine learning, Network Analysis, Visualization, Data mining, POS tagging, N-gram search, Sentiment analysis, WordNet,
Pattern [15]	enables users to find products based on features
Review Seer	Automation of aggregation sites

VII. APPLICATION

The main applications of Opinion mining and sentiment analysis are as given below

1) **Opinion spam detection:**
People may write reviews about product with negative intension. Opinion mining and sentiment analysis can classify the these reviews into' spam' content and 'not spam' content [8].

2) **Purchasing Product or Service:**
By this technique, people can easily evaluate other's feedback and experience about any product or service and also he/she can easily compare the competing brands.

3) **Quality Improvement in Product or service:** In this manufactures can collect the negative opinion as well as the positive opinion about their product or service and thereby they can improve the quality of their product or service.

4) **Marketing research:**
Product or services with their new government policy can be analysed. These all result can be contributed to collective intelligent research [7].

5) **Policy Making:** Through Sentiment analysis, policy makers can take citizen's perception towards some policy and they can utilise this information in creating new improved citizen friendly policy.

6) **Decision Making:** People's opinion and experience are a very useful element in decision-making process. It gives analysed people's opinion that can be effectively used for decision making.

VIII. RESEARCH CHALLENGES

1. Product reviews, comments and feedback could be in different languages (English, Urdu, Arabic, French etc), therefore to tackle each language according to its orientation is a challenging task[17].
2. As noun words are considered as feature words but Verbs and adjectives can also be used as feature words which are difficult to identify [17].
3.To group, the synonym words is also a challenging task. Example: (a) "the picture quality is excellent" and customer-Two comments,(b) "picture quality of phone is very good". Both are talking about same feature but with different wording.
6. Orientation of opinion words could be different according to situation. For example "Camera size of mobile phone is small". Here adjective small used in positive sense but if customer parallel said that "the battery time is also small". Here small represent negative orientation to battery of phone. To identify the polarity of same adjective words in different situation is also a challenging task [17].
7. As the customer comment in free format, she/he can use an abbreviation, short words, and roman language in reviews. For example u for you, cam for camera, pic for picture, f9 for fine,b4, before, gud for good etc. To deal with such type of language need a lot of work to mine opinion [17].
8. Different people have different writing styles, same sentence may contain positive as well as negative opinion, so it is difficult to parse sentence as positive or negative in case of sentence level opinion mining[17] .
9. To finding of spam and fake reviews, mainly through the identification of duplicates[17].
10. The comparison of qualitative with summary reviews and the detection of outliers, and the reputation of the reviewer.
11. The combination of opinion with behaviour to validate data and provide further analysis into the data ahead of opinion expressed[17].
12. The natural language overhead like Implicitness, co-reference, ambiguity, inference etc. created hindrance in sentiment analysis too.
13. The biggest challenge 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

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICACTRP 2017

International Conference on Advances in Computational Techniques and Research Practices

Noida Institute of Engineering & Technology, Greater Noida

Vol. 6, Special Issue 2, February 2017

very poor in some other domain[2].
14. Word Sooo….. Sweeetttt….., I am toooo Haappy or if they in hurry they stress the words for e.g. comeeeee fasssssst I am waitttttnggg in this example People use orthographic words for expressing their excitement, happiness is also a big challange.

IX. RESEARCH SCOPE

The research scope in opinion mining and sentiment analysis are:
1) Spam Detection using Sentiment Analysis.
2) Sentiment Analysis on short Sentence that include abbreviations.
3) Improvement of existing sentiment word identification algorithm.
4) Developing fully automatic tools for analysis.
5) Effective Analysis of policy documents which containing opinion content.
6) Managing the of bi polar sentiments successfully.
7) Designing and Generation of highly content corpus database.

X. CONCLUSION

Opinion mining and sentimental analysis is an emerging field of data mining used to extract the knowledge from a huge volume of customer comments, feedback and reviews on any product or topic etc. A lot of work in opinion mining in customer reviews has been conducted to mine opinions in form of document, sentence and feature level sentiment analysis. In future, Opinion Mining can be carried out on set of discovered feature expressions extracted from reviews. The Opinion Mining and in natural language processing community, Sentiment Analysis become a most interesting research area. A more innovative and effective techniques needed to be invented which should overcome the current challenges faced by Opinion Mining and Sentiment Analysis.

REFERENCES

[1] Mugdha More and Bharat Tidke, July 2015. A Framework For Summarization Of Online Opinion Using Weighting Scheme, ACII, Vol.2, No.3.
[2] Haseena Rahmath P," Opinion Mining and Sentiment Analysis - Challenges and Applications", IJAIEM Volume 3, Issue 5, May 2014.
[3] Muhammaad Zubair, Aurangzeb Khan, Shakeel Ahmad, FazalMasudKundi and Asghar,2014.A Review of Feature Extraction in Sentiment Analysis. ISSN 2090-4304 Journal of Basic and Applied Scientific Research.
[4] RichaSharma, Shweta Nigam and Rekha Jain,2014.Opinion Mining In Hindi Language: A Survey.International Journal in foundation of Computer Science & Technology,Vol. 4,No.2.
[5] Padmapani P. Tribhuvan, S.G. Bhirhud, Amrapalli P. Tribhuvan,2014.A Peer Review of Feature Base "Opinion Mining and Summarization" ,IJCSIT, Vol. 5(1),247-250.
[6] Tanvir Ahmad,2013.Clustering Technique for Feature Segregation in Opinion Analysis, IJCA, (0975 – 8887) Volume 76– No.17.
[7] Tanvir Ahmad, Mohammad Najmud Doja,2013.Ranking System for Opinion Mining of Features from Review Documents. IJCSI, July 2012 ISSN (Online): 1694-0814
[8] Liu, B., "Sentiment Analysis and Subjectivity". Appeared in Handbook of Natural Language Processing, Indurkhya, N. & Damerau, F.J. [Eds.] (2010).
[9] Segaran, T. (2007), Programming Collective Intelligence. Sebastopol: O'Reilly Media, Inc.
[8] Pang, B. & Lee, L. (2008), "Opinion Mining and Sentiment Analysis". In Foundations and Trends in Information Retrieval 2 (1-2), p. 1–135.
[9] http://nlp.stanford.edu/software/corenlp.html
[10] http://www.cs.waikato.ac.nz/ml/weka/
[11] http://www.nltk.org/
[12] https://opennlp.apache.org/
[13] http://alias-i.com/lingpipe/
[14] https://gate.ac.uk/
[15] http://www.clips.ua.ac.be/pattern
[16] http://www.sussex.ac.uk/Users/johnca/rasp/offline-demo.html
[17] Nidhi R. Sharma, Prof. Vidya D. Chitre" Opinion Mining, Analysis and its Challenges", ence IJIACS ISSN 2347 – 8616 Volume 3, Issue 1 April 2014.

BIOGRAPHIES

Mr. Surya Prakash Sharma completed his B.Tech. in Computer Science & Engineering from Kanpur University in 2003 and M.Tech.. in Computer Science & Engineering from M.D. University Rohtak(India) in 2011.Currently he is Ph.D. research scholar in Dr APJAK Technical University Lucknow (india). His research area is Opinion Mining on sentimental analysis on customer review documents and information reterival

Dr Rajdev Tiwari received PhD in Computer Science from Dr BRA University (India) in 2012. Currently he is Professor& Head in the department of Computer Application in Greater Noida Institute of Technology (GNOIT), Greater Noida (India). He has membership of many professional bodies. He has published many research papers in national and international conference/Journals. His research interest is Data Mining & Warehousing, Soft Computing (Neural Networks, Fuzzy Logic, Genetic Algorithms), Operating System, Design & Analysis of Algorithms.

Dr Rajesh Prasad received PhD in Computer Science & Engineering from Motilal Nehru National Institute of Technology, Allahabad, India, January 2012. Currently he is the Professor & Head Department of Computer Science Yobe State University, Damaturu, Nigeria. He has membership of many professional bodies. He has published many research papers in national and international conference/Journals. His area of interest is Theory of Automata, Discrete Mathematics, Design & Analysis of Algorithms, Compiler Design, Database Management System and Cryptography.