



A Perspective Survey on Opinion Mining Tools and Techniques

M. K. Prakash¹, G.Dinesh²
Assistant Professor^{1, 2}

Department of Computer Applications
Department of Information Technology
Sri Ramakrishna Mission Vidyalaya College of Arts and Science, India

Abstract:

Opinion Mining is a useful tool for gathering information about customers' opinions about products, brands or companies. Opinion mining is a type of natural language processing which could track the mood of the people about any particular product by review. Opinion Mining is a process of automatic extraction of knowledge by means of opinion of others about some particular product, topic or problem. The idea of Opinion mining and Sentiment Analysis tool is to process a set of search results for a given item based on the quality and features. Opinion mining is also called sentiment analysis due to large volume of opinion which is rich in web resources available online. Analyzing customer review is most important, by doing that we tend to rate the product and provide opinions for it which is been a challenging problem today. This paper focus on the opinion mining challenges, tools and techniques for a particular product, from the customer review.

Keywords: opinion mining, statement analysis, text mining, web content mining.

I. INTRODUCTION

The evolution of automated systems and digital information in every field of life is evolving rapidly which tends to generate data. As a result huge volumes of data are produced in field of science, engineering, medical, marketing, finance, demographic etc. Automated systems are meant to automate analysis, summarization and classification of data and number of efficient ways is available to store huge volumes of data. Data Mining is defined as extracting information from huge sets of data. Data mining is the procedure of mining knowledge from data. Text mining is an interdisciplinary method used in different fields like machine learning, information retrieval, statistics, and computational linguistics.[1] Web mining is the process of using data mining techniques and algorithms to extract information directly from the Web by extracting it from Web documents and services, Web content, hyperlinks and server logs. The goal of Web mining is to look for patterns in Web data by collecting and analyzing information in order to gain insight into trends, the industry and users in general. Web mining is a branch of data mining concentrating on the World Wide Web as the primary data source, including all of its components from Web content, server logs to everything in between. The contents of data mined from the Web may be a collection of facts that Web pages are meant to contain, and these may consist of text, structured data such as lists and tables, and even images, video and audio.

Web content mining: This is the process of mining useful information from the contents of Web pages and Web documents, which are mostly text, images and audio/video files. Techniques used in this discipline have been heavily drawn from natural language processing (NLP) and information retrieval.

Web structure mining: This is the process of analyzing the nodes and connection structure of a website through the use of graph theory. There are two things that can be obtained from this: the

structure of a website in terms of how it is connected to other sites and the document structure of the website itself, as to how each page is connected.

Web Usage Mining: Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site.

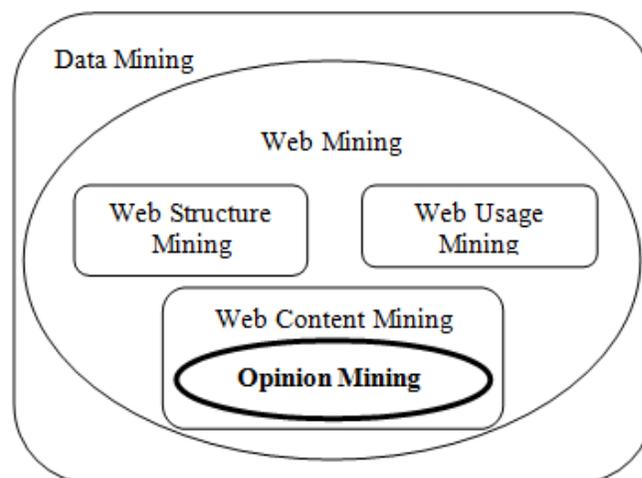


Figure.1. a. Opinion Mining Taxonomy

Opinion mining is a topic in Text mining, Natural Language Processing (NLP), and Web mining discipline[3]. The goal of Opinion Mining is to make computer able to recognize and express emotions. A thought, view, or attitude based on emotion instead of reason is called sentiment. [4]Sentiment Analysis or Opinion Mining is the study of people's opinion, attitude or emotion towards an event or product. It is a process of finding user opinion about particular topic or product. Topic can be anything, like,

news, event, movie, etc.[4]

II. DATA SOURCE

People and companies across good trainings exploit the high and unique sources of data for various purposes. User Opinion is an important criterion for the progress of the quality services. Blogs, review sites, Data set and micro-blogs furnish a good understanding for the deliverable view of the products and services provided to the customers / viewers [5].

2.1 Blogs

With an growing usage of the Internet, Blog pages and blogging are mostly used [2]. The names connected to universe of all blog sites are called blogosphere [6]. Blog pages are used to express one's personal opinions about any product or topic. People like to share their opinions, ideas or suggestions with others on a blog. Blogging is an occurrence thing because of its simplicity of creating blog posts and reviews, its free form and unedited nature. Blogs are used as a source of opinion in many of the studies linked with sentiment analysis.

2.2 Review Sites

For any user in deciding a purchasing decision, others opinion is an important factor. The user generated reviews and suggestions are largely available on the Internet. The reviews for products or services are available as opinions in unstructured format. The reviewers data are used in Sentiment classification studies are collected from the ecommerce websites like www.yelp.com (restaurant reviews), www.amazon.com (product reviews), www.flipkart.com (product reviews), which hosts millions of product reviewed by customers [3].

2.3 Data Set

Many works in the field uses movie reviews data for classification. The Multi-Domain Sentiment (MDS) contains different types of product reviews taken from Amazon.com, Flipkart.com including Books, dresses, Kitchen appliances and Electronics things, with many positive and negative suggestions / reviews for each territory.

2.4 Micro-blogging:

Micro-blogging is the well-known communication tool for internet users. A large number of messages appear daily in web-sites for micro-blogging such as Twitter, Tumblr and Facebook. Twitter is very popular micro-blogging service where users express messages called "tweets". These Tweets are used to express their own opinions/suggestions about different topics. Sometimes these Twitter messages are also used as data source for Sentiment Classification.

III. OPINION MINING CHALLENGES

Since opinion mining is a relatively new filed, thus there are several challenges to be faced. According to opinion mining techniques are just primitive for opinions and comparisons identification and extraction. Mainly these challenges are related to the authenticity of the extracted data and the methods used in it. A summary of challenges of opinion mining [8] is as follows:

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.
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.

3. If a customer-One comments on mobile phone, "the voice quality is excellent" and customer-Two comments, "sound quality of phone is very good". Both are talking about same feature but with different wording. To group the synonym words is also a challenging task.

4. 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.

5. As the customer comment in free format, she can use 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.

6. 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 .

7. In Bing Liu approach opinion always classified only in two categories positive and negative but Neutral opinion also expressed sometimes. Liu considers only adjective as opinion words but opinion can also expressed as adverb, adjectives and verb. For example "like" is a verb but also an opinion word. His approach finds the implicit features because it extracts the sentences contain at least one feature word. So the features commented by customer indirectly are ignored [7].

8. Lexicon based methods use for opinion mining has not an effective method to deal with context dependent words. For example the word "small" can express the either positive or negative opinion on the product features. For a mobile phone if customer comments that "size of mobile phone is small" this sentence does not show either size is positively opinioned or negatively.

9. To finding of spam and fake reviews, mainly through the identification of duplicates.

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 behavior to validate data and provide further analysis into the data ahead of opinion expressed.

12. The continuous need for better usability and user-friendliness of the mining systems.

IV. OPINION MINING TOOLS

The tools are used to trace the opinion or polarity from the user generated texts. They are,

1. **Red Opal:** It is a tool that makes able the users to find products based on attributes / features. It assigns the points / ranks to each product based on their features, which are extracted from the customer generated reviews. The extracted features are displayed in graph format. The extracted features / attributed are assigned rank by Naive Bayes Classifier as positive and negative review. The results are displayed in the form of attributes and its score.
2. **Web Fountain:** It is used to create a simple web interface. It uses the beginning definite Base Noun Phrase (bBNP) heuristic

method to extract the features of the product.

3. **Review Seer Tool:** It is used to automate the work performed by aggregation sites. The Naive Bayes Classifier method is used to collect positive and negative opinions from customer reviews for assigning a rank to the extracted features [9].

4. **Opinion Observer:** This Opinion Mining system is used for analyzing and comparing customer generated opinions on the Internet. This system displays the results of a product feature by feature in a graph format.

5. **NLTK:** The natural language toolkit is a tool for text processing, classification, tokenization, stemming, tagging, parsing etc. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as Word Net, along with a suite of <http://www.nltk.org/>

6. **GATE:** Useful in developing a pipeline. Language analysis modules for various languages are contributed by developers. They are available to be used plugged in the pipeline.

7. **OpenNLP:** perform the most common NLP tasks, such as POS tagging, named entity extraction, chunking and co-reference resolution. <http://opennlp.apache.org/>.StanfordCoreNLP.

8. **Ling Pipe:** Ling Pipe is used for linguistic processing of text including, clustering classification and entity extraction etc. <http://alias-i.com/lingpipe/>

9. **Orange:** Orange is an open source data visualization and analysis tool, where data mining is done through visual programming or Python scripting. The tool has components for machine learning, add-ons for bioinformatics and text mining and it is packed with features for data analytics. <https://blog.biolab.si/tag/sentiment-analysis/>

10. **KNIME:** Knime is a java open-source, cross-platform application which name means "Konstanz Information Miner". It is actually used extensively for data mining, data analysis and optimization. It can be downloaded as the core application itself (Knime Desktop), or the whole SDK which is based on Eclipse Helios. <https://www.knime.com/>

V. Opinion Mining Techniques

Study the evolution of sentiment analysis research can be done by the analytical tokens, or building blocks and the implicit information associated with those tokens. We can group the existing approaches into four main types they are keyword spotting approach, lexical affinity method, statistical methods, and concept-based techniques.[10].

5.1 Keyword spotting:

This approach makes use of unambiguous affect words such as happy, sad, afraid, and bored to classify the text into different categories. "Today was a great day" as being affectively positive, it is likely to assign the same classification taking an example of sentence like "Today wasn't a great day at all." Keyword spotting also relies on the presence of actual effect of words and not for a sentence.

Sometimes, a sentence conveys through their meaning rather than affect adjective words. For ex the following sentence "My husband just wants to take custody of my children away from me" states strong emotions, but uses no affect keywords, and therefore is ineffective. Lexical affinity is slightly more sophisticated than keyword spotting.

5.2 Lexical affinity:

This approach not just detects obvious affect words, but also assigns arbitrary words a probable "affinity" to particular

emotions. For example, lexical affinity might assign the word "accident" a 75-percent probability of indicating a negative effect, as in "car accident" or "hurt by accident." This approach usually trains probability from linguistic corpora. Although it often outperforms pure keyword spotting, there are two main problems with this approach. First, negated sentences (I avoided an accident) and sentences with other meanings (I met my friend by accident) trick lexical affinity, because they operate solely on the word level. Second, lexical affinity is often biased towards a category, directed by the linguistic corpora's source. Due to which it becomes difficult to build and reuse the model for different systems.

5.3 Statistical methods:

This approach is popular for affect text classification. Researchers use statistical methods on projects, by using a machine-learning algorithm for a large training corpus of affectively annotated texts, the system will learn combination of things as in keyword spotting and lexical affinity. Generally, statistical methods are semantically weak, which means that individually with the exception of obvious affect keywords a statistical model's other lexical or co occurrence Elements have little predictive value. Due to which statistical text classifiers give accurate results only when they are given sufficiently large amount of data as input. So, these methods will affectively classify the reviews or opinion text on large for ex page or paragraph level, they don't fare well when it comes to smaller texts such as clauses or sentences.

5.4 Concept-based approaches:

To accomplish semantic text analysis methods such as semantic networks and web ontology are used. This helps the approach to identify the affective and conceptual data related to natural language opinions. In this approach instead of blindly using the matches of keywords and counts of word co-occurrence by using the large semantic knowledge dictionaries, and rather depend on the actual meaning associated with natural language contexts. Concept-based approaches can identify minute expressed sentiments when compared to purely syntactical techniques. Concept-based approaches can analyze multi-word expressions that don't explicitly convey emotion. The concept-based approach relies heavily on the depth and breadth of the knowledge bases it uses. Without a comprehensive resource that encompasses human knowledge, an opinion mining system will have difficulty grasping the semantics of natural language text.

VI. DIFFERENT LEVELS OF OPINION MINING

6.1 Document level Opinion Mining:

The basic data unit is a one document of opinionated text [11] where in document level classification deals with classifying to a single or one review for that entire document. One of the situation where document level opinion mining doesn't give good results is in forums blogs etc where there is a comparison made between two different topics in such a case document level mining can classify to wrong single review. Therefore classification depending on subjective and objective becomes very important.

6.2 Sentence level Opinion Mining:

In sentence level Opinion Mining, each sentence is parsed and the polarity is been derived. The same methods of document level

classification can be applied to the sentence level classification as well but subjective and objective sentences [12] must be identified. The objective sentences contain the facts of the product or topics such as price specifications and so on where as in subjective sentences it will have user opinions which can determine the sentiments. Then the subjective sentences are classified to positive or negative classes.

6.3 Phrase level Opinion Mining:

The phrase level sentiment classification is a more sophisticated approach to opinion mining. The phrases are scanned for the opinion words and then the phrase level opinion mining is applied to them. This method also does not work well in all the cases where the negation words and sentences are far apart it can identify only when negation words occur locally [13].

VII. CONCLUSION

Opinion mining is an emerging field of data mining used to extract the pearl knowledge from huge volume of customer comments, feedback and reviews on any product or topic etc. A lot of work 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 Facebook on pictures, videos or Facebook status. In future, Opinion Mining can be carried out on a set of reviews and set of discovered feature expressions extracted from reviews. The state-of-art for current methods, useful for producing better summary based on feature based opinions as positive, negative or neutral is in the classifying in three different levels of opinion mining(Document level opinion Mining, Sentence level opinion mining, phrase level opinion mining).

VIII. REFERENCES

- [1] Bakhtawar Seerat, Farouque Azam “Opinion Mining: Issues and Challenges (A Survey),” *International Journal of Computer Applications* (0975-8887) Volume 49-No.9, July 2012
- [2] S. Kasthuri, Dr. L. Jayasimman, Dr. A. Nisha Jebaseeli, “An Opinion Mining and Sentiment Analysis Techniques: A Survey”, *International Research Journal of Engineering and Technology (IRJET)*, Volume 03 Issue: 02, Feb 2016
- [3] Neeru Mago, “Opinion mining: Applications, Techniques, Tools, Challenges and Future Trends of Sentiment Analysis”, *International Journal of Computer Engineering and Applications*, Volume X, Issue IV, April 2016
- [4] Dr. P. Perumal, M. Kasthuri, “A Survey on Opinion Mining from online Review Sentences” *International Research Journal of Engineering and Technology*.
- [5] Nidhi Mishra, Dr. C.K. Jha, “Classification of Opinion Mining Techniques”, *International Journal of Computer Applications*, Volume 56, No.13, October 2012.
- [6] Sumathi.T, Karthik.S, Marikannan.M “Performance Analysis of Classification Methods for Opinion”, *International Journal of Innovations in Engineering and Technology (IJJET)*, Volume 2, Issue 4, August 2013.
- [7] B. Liu “Sentiment Analysis and Opinion Mining”, April 22, 2012.
- [8] Nidhi R. Sharma , Prof. Vidya D. Chitre “Opinion Mining, Analysis and its Challenges”, *International Journal of Innovations & Advancement in Computer Science*
- [9] N.Mishra and C.K.Jha 2012, “An Insight into task of opinion mining”, *Second International Joint Conference on Advances in Signal Processing and Information*
- [10] Erik Cambria, Bjorn Schuller, Yunqing Xia. *New Avenues in Opinion Mining and Sentiment Analysis*, *Intelligent Systems, IEEE*, Volume:28 ,Issue:2 March-April 2013.
- [11] Raisa Varghese, Jayasree. *A Survey on Sentiment Analysis and Opinion Mining*, *International Journal of Research in Engineering and Technology (IJRET)*, Vol 2 Issue 11 Nov 2013..
- [12] Nidhi Mishra et al. *Classification of Opinion Mining Techniques*, *International Journal of Computer Applications*, Vol 56, No 13, Oct 2012 Pg No 1-6..
- [13] G.Angulakshmi, Dr.R.ManickaChezian. *An Analysis on Opinion Mining: Techniques and Tools* , *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 3, Issue 7, July 2014.



M. K. Prakash MCA, M. Phil, Assistant Professor, Department of BCA, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, Tamil Nadu.



G. Dinesh, MCA, M. Phil, Assistant Professor, Department of Information Technology, Sri Ramakrishna Mission Vidyalaya College of Arts and Science, Coimbatore, Tamil Nadu.