

Review on Techniques and Tools used for Opinion Mining

Asmita Dhokrat
Dept. of CS & IT
Dr. Babasaheb Ambedkar
Marathwada University
Aurangabad, India

Sunil Khillare
Dept. of CS & IT
Dr. Babasaheb Ambedkar
Marathwada University
Aurangabad, India

C. Namrata Mahender
Dept. of CS & IT
Dr. Babasaheb Ambedkar
Marathwada University
Aurangabad, India

Abstract Humans communication is generally under the control of emotions and full of opinions. Emotions and their opinions plays an important role in thinking process of mind, influences the human actions too. Sentiment analysis is one of the ways to explore user's opinion made on any social media and networking site for various commercial applications in number of fields. This paper takes into account the basis requirements of opinion mining to explore the present techniques used to developed an full fledge system. Is highlights the opportunities or deployment and research of such systems. The available tools used for building such applications have even presented with their merits and limitations.

Keywords: Opinion Mining, Emotion, Sentiment Analysis, EM Algorithm, SVM algorithm

1. INTRODUCTION

Emotions are the complex state of feelings that results in physical and emotional changes that influences our behavior. Emotion is a subjective, conscious experience characterized mainly by psycho-physiological expressions, biological reactions, and mental states. Emotion is often associated and considered commonly significant with mood, nature, personality, disposition, and motivation. It is also influenced by hormones and neurotransmitters such as dopamine, noradrenaline, serotonin, oxytocin, cortisol and GABA [1]. Emotion is a positive or negative experience that is associated with a particular pattern of physiological activity. Humans carry lot of emotions like happiness, sadness, angry, disgust, surprise, fear, panic, scared etc. identifying these emotions are very easy in face to face communication compare to written communication. But now a day's use of social media has increased rapidly and the huge amount of textual data became available on web, mining and managing this vast data has become a crucial task. As the growth of E-facilities have increased lots of people got encouraged to write their emotions, views, opinions about a person, product, place or anything they want.

Opinion Mining or Sentiment analysis involves building a system to explore user's opinions made in blog posts, comments, reviews or tweets, about the product, policy or a topic [2]. Opinion mining is nothing but finding the opinion of person from sentences and classify them on the basis of polarity. As the world changed into E-World the way of expression has dramatically changed for example wide use of smiley's and symbols can be seen as expression while texting. Social communication can be observed on internet and new term has been coined for various ways of communication like texting, twitting, posting etc. people like to communicate with others through internet, they want to share their feelings, likes, dislikes, opinions, views, reviews, emotions etc. people are happy to share their personal life via social media, the use of social media has increased so much and so rapidly that even no body worries about what they are sharing and is this good to

share our personal life with unknown persons? Is there any need to share our photos, videos or our daily activities on internet? So finding the sentiment, emotion behind this activity is also an important task for understanding the psycho-socio status. So from that text, mining the opinions of people and finding their views, reaction, sentiments and emotions have become challenging task.

Opinion Mining is the field of study that analyzes people's opinion, sentiments, evaluations, attitudes and emotions from written text. Opinion Mining is one of the most active research areas in Natural Language Processing and is also widely studied in data mining, web mining and text mining this research has spread outside of computer science to the management science and social science due to its importance to business and society. The growing importance of sentiment analysis coincides with the growth of social media such as Reviews, Forums, discussion groups, chatting, blogs, micro-blogs, twitter and social networks.

1.1 Categorization of Text

Sentiment analysis is also called as opinion mining; as it mines the information from various text forms such as reviews, news & blogs and classifies them on the basis of their polarity as positive, negative or neutral [3]. It focuses on categorizing the text at the level of subjective and objective nature. Subjectivity indicates that the text contains/bears opinion content for e.g. Battery life of Samsung mobiles is good. (This sentence has an opinion, it talks about the Samsung mobile phones and showing positive (good) opinion hence it is Subjective). Samsung mobiles are having long battery life. (This sentence is a fact, general information rather than an opinion or a view of some individual and hence its objective) [4]

1.2 Components of Opinion Mining

There are mainly three components of Opinion Mining [3]:

- **Opinion Holder:** Opinion holder is the holder of a particular opinion; it may be a person or an organization that holds the opinion. In the case of blogs and reviews, opinion holders are those persons who write these reviews or blogs.
- **Opinion Object:** Opinion object is an object on which the opinion holder is expressing the opinion.
- **Opinion Orientation:** Opinion orientation of an opinion on an object determines whether the opinion of an opinion holder about an object is positive, negative or neutral.

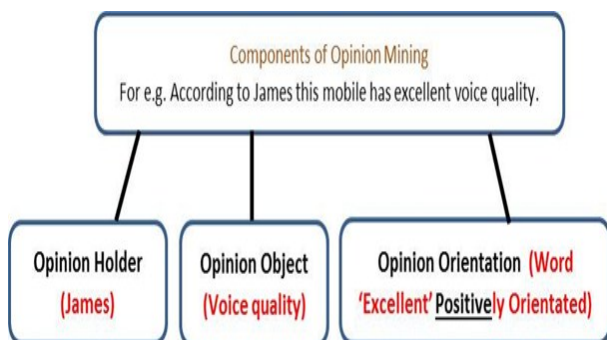


Figure 1 Components of opinion Mining

2. DIFFERENT LEVELS OF SENTIMENT ANALYSIS

In general, sentiment analysis has been investigated mainly at three levels [4].

- **Document level:** The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment. For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product. This task is commonly known as document level sentiment classification.
- **Sentence level:** The task at this level goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion. This level of analysis is closely related to subjectivity classification which distinguishes sentences (called objective sentences) that express factual information from sentences (called subjective sentences) that express subjective views and opinions.
- **Entity and Aspect level:** Both the document-level and sentence-level analyses do not discover what exactly people liked and did not like. Aspect level performs fine-grained analysis. Aspect level was earlier called feature level (feature-based opinion mining and summarization).

3. CHALLENGES IN OPINION MINING

There are several challenges in Opinion Mining as follows,

- **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[5].

- **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[6].

- **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[5].

- **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[6].

- **Mixed Sentences:** Suppose the word is positive in one situation may be negative in another situation. For e.g. Word LONG, suppose if customer says the battery life of Samsung mobile is too long so that would be a positive opinion. But suppose if customer says that Samsung mobile take too long time to start or to charge so it would be a negative opinion.

- **Way of Expressing the Opinion:** The people don't always express opinions in the same way. The opinion of every individual is different because the way of thinking, the way of expressing is vary from person to person.

- **Use of Abbreviations and shortforms:** People using social media more and that to for chatting, expressing their views using shortcuts or abbreviations so the use of colloquial words is increased. Uses of abbreviation, synonyms, special symbols is also increase day by day so finding opinion from that is too difficult. For e.g. F9 for fine, thnx for thanks, u for you, b4 for before, b'coz for because, h r u for how are you etc.

- **Typographical Errors:** Sometimes typographical errors cause problems while extracting opinions.

- **Orthographics Words:** People use orthographic words for expressing their excitement, happiness for e.g. Word Sooo..... Sweetttt....., I am toooo Haappy or if they in hurry they stress the words for e.g. comeeeee fasssssst I am waittttnggg.

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

4. DATA SOURCES AND TOOLS OF OPINION MINING

While doing research the collection of data is the biggest issue and for the task like opinion mining, sentiment analysis its too difficult because lots of information is available on internet and collection of that data and extraction of opinion from huge amount of data is too hard. So here we discussed about some available data sources and the tools which is used for extraction the sentiments and opinion of the given text.

4.1 Data Sources available for Opinion mining

There are various data sources available on web, i.e. Blogs, Micro blogs, online posts, News feeds, Forums, review sites etc.

- **Blogs:** Blogs are nothing but the user own space or diary on internet where they can share their views, opinions about topics they want.
- **Online Reviews:** on Internet various review sites are available through that you can check online reviews of any product before purchasing that.
- **Micro blogging:** Micro blogs allow users to exchange small elements of content such as short sentences, individual images, or video links", which may be the major reason for their popularity.
- **Online Posts:** people share their own ideas, opinions, photos, videos, views, likes, dislikes, comments on specific topics etc.
- **Forums:** An Internet forum, or message board, is an online discussion site where people can hold conversations in the form of posted messages.

This table gives you an idea about the available data sources along with the address of sites from which you can download the posts, tweets, reviews for products etc.

Table 1. Available Data sources with web address

D a t a S o u r c e s	Respective Sites/ Source
Bl o g s	http://indianbloggers.org/ , http://www.bloggersideas.com/ , http://www.digitaltrends.com/ , http://thoughts.com/free-blog , http://blog.com/ , http://blog.hubspot.com/ https://wordpress.com/ ,
R e v i e w S i t e s	http://www.sitejabber.com/ , http://www.toptenreviews.com/ , http://www.trustedreviews.com/ , https://in.pinterest.com/ , http://www.business-edge.com/ , http://www.websitemagazine.com/ , http://www.yellowpages.com
M i c r o - B l o g g i n g	https://tumblr.com/(Tumblr) , http://friendfeed.com/(Friendfeed) , http://www.plurk.com/top/(Plurk) , https://twitter.com/(Twitter) , http://www.jaiku.com/(Jaiku) , http://www.qaiku.com/(Quiku) , https://www.identi.ca/(Identica) , http://www.spotjots.com/(Spotjots) , http://www.meetme.com/(Meetme)
O n l i n e P o s t s	https://www.facebook.com/(Facebook) , https://myspace.com/(MySpace) , http://www.skype.com/en/(Skype) , https://www.linkedin.com/(LinkedIn) , https://diasporafoundation.org/(Diaspora) , https://plus.google.com/(GooglePlus) , https://www.whatsapp.com/(Whatsapp) , https://www.snapchat.com/

	(Snapchat), https://telegram.org/(Telegram) , https://www.flickr.com/(Flickr)
F o r u m s	http://www.forums.mysql.com , http://www.forums.net.com , http://www.forum.joomla.org , https://forums.digitalpoint.com , http://www.bookforum.com , http://www.myspace.com/forums , http://tsrmatters.com/(The Student Room) , http://ubuntuforums.org/ , https://stackoverflow.com/ ,

4.2 Tools available for Opinion mining

As we discussed in 4.1 there are various data sources are available on web and mining those data is difficult task. Main difficulty is extraction of emotions, structure of text, form of data i.e. image or text, the language used on internet for communication is vary from person to person or state to state. So here are some ready to use tools for opinion mining for various purposes like data preprocessing, classification of text, clustering, opinion mining, sentiment analysis etc.

The table no. 2 shows the name of particular tool as well as uses of these tools.

Table 2. List of available tools

Name of Tools	Uses
STANFORD CORENLP [7]	POS tagging, Named entity recognizer, Parsing, Coreference resolution system, Sentiment analysis, Bootstrapped pattern learning
WEKA [8]	Machine learning algorithm for Data Mining, Data pre-processing, Classification, Regression, Clustering, Association rules, Visualization.
NLTK [9]	Classification, Tokenization, Stemming, Tagging, Parsing, Semantic reasoning, Provides lexical resources such as WordNet
APACHE OPENNLP [10]	Tokenization, Sentence segmentation, Part-of-speech tagging, Named entity extraction, Chunking, Parsing, Coreference resolution
LingPipe [11]	Entity extraction, POS tagging, Clustering, Classification.
GATE [12]	Tokenizer, Gazetteer, Sentence splitter, POS tagging, Named entities transducer, Coreference tagger
Pattern [13]	Data mining, POS tagging, N-gram search, Sentiment analysis, WordNet, Machine learning, Network analysis, Visualization
Robust Accurate Statistical Parsing [14]	Statistical Parser, Tokenization, Tagging, Lemmatization and Parsing

5. EXISTING WORK IN OPINION MINING

As we know the beginning of opinion mining has marked in late 90's but this paper discusses the advances carried out from the year 2002 to 2014. In this section brief tabulated information about the major contribution in the field of opinion mining is shown. The table no. 3 shows details about the author, their work, different techniques used while working on Opinion

Mining and brief introduction of that paper as conclusion of that paper.

Figure 2. Detailed study of existing research work.

Ref. No.	Author's	Title of the Paper	Techniques used	Conclusion	Year
[15]	B. Pang, L. Lee, S. Vaithyanathan	Thumbs up? Sentiment Classification using Machine Learning Techniques	Naive Bayes, Maximum entropy classification and SVM	In this paper, discussion on sentiment classification of movie reviews on the basis of positive and negative is given in length. They used three different machine learning algorithms for text classification for increasing accuracy in classification.	2002
[16]	Turney P.D	Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews	Pointwise mutual information (PMI) and Information Retrieval (IR)	In this paper author classified reviews on the basis of thumbs up(recommended) and thumbs down (Not recommended) and classification is predicted by semantic orientation, for this purpose they used unsupervised learning algorithm and PMI-IR uses to measure the similarity of pairs of words or phrases.	2002
[17]	Michael G. Madden	A New Bayesian Network Structure for Classification Tasks	Partial Bayesian Network (PBN) K2algorithm	Author proposed a methodology for induction of a Bayesian network structure for classification and this structure is called Partial Bayesian Network (PBN). It is implemented using the K2 framework. Learning the Partial Bayesian Network essentially reduces to a Bayesian Network learning problem using the K2 algorithm. The complexity of K2 algorithm is exponential to the number of variables; hence, PBN is also feasible only for small data sets.	2002
[18]	Hai Leong Chieu, Hwee Tou Ng	Named Entity Recognition: A Maximum Entropy Approach Using Global Information	Maximum Entropy	In this paper authors presented maximum entropy based named entity recognizer for global information. It uses information from the whole document to classify each word with just one classifiers.	2003
[19]	Turney P.D., Littman M.L	Measuring praise and criticism: inference of semantic orientation from association.	Pointwise mutual information (PMI) and Latent semantic analysis (LSA)	This paper introduced a method for inferring the semantic orientation of a word from its statistical association with a set of positive and negative paradigm words.. They use pointwise mutual information (PMI) and latent semantic analysis (LSA) to measure the relation between a word and a set of positive or negative words and according to this paper LSA gives better results than PMI.	2003
[20]	S.-M. Kim, E. Hovy	Determining the sentiment of opinions	SVM (Support Vector Machine)	Here author discusses about identifying sentiments. Here classification and combination of sentiment at word and sentence levels for identifying opinion holder they used learning techniques like SVM.	2004
[21]	Soo-Min Kim, Eduard Hovy	Identifying Opinion Holders for Question Answering in Opinion Texts	Maximum Entropy	In this paper they used Maximum Entropy for learning the opinion holders automatically with the help of two ways i.e. Classification and Ranking.	2005
[22]	Jack G. Conrad, Frank Schilder	Opinion Mining in Legal Blogs	Language Model and Naive Bayes classifiers.	Here in this paper authors discuss about the scope and opinion mining of blogs which is increased in legal domain. Here they first construct a Weblog test collection containing blog entries that discuss legal search tools. Then they subsequently examine the performance of a language modeling approach deployed for both subjectivity analysis and polarity analysis.	2007
[23]	Xiaowen Ding, Bing Liu	A Holistic Lexicon-Based Approach to Opinion Mining	Holistic lexicon-based approach	In this paper author discussed about customer reviews of products, they also discussed about the opinion words that show desirable and undesirable states. They used a holistic lexicon-based approach to solving the problem by exploiting external evidences and linguistic conventions of natural language expressions.	2008
[24]	Alec Go, Richa Bhayani, Lei Huang	Twitter Sentiment Classification using Distant Supervision	Naive Bayes, Maximum Entropy, and SVM	Author introduced a novel approach for automatically classifying the sentiment of Twitter messages as either positive or negative with respect to a query term. Author shows that machine learning algorithms like Naive Bayes, Maximum Entropy, and SVM for classifying sentiment, have accuracy above 80% when trained with emoticon data. This paper also describes the preprocessing steps needed in order to achieve high accuracy	2009
[25]	Zhao Yan-Yan Qin Bing, Liu Ting	Integrating Intra- and Inter-document Evidences for Improving Sentence Sentiment Classification	Graph-Based approach, NB (Naive Bayes), SVM(Support Vector Machine)	Here author classify the sentence into positive, negative and objective. Here author propose two such outside sentence features: intra-document evidence and inter-document evidence. Then in order to improve the sentence sentiment classification performance, a graph-based propagation approach is presented to incorporate these inside and outside sentence features.	2010
[26]	Zhongwu Zhai, Bing Liu, Hua Xu, Peifa Jias	Clustering Product Features for Opinion Mining	Expectation–Maximization (EM) algorithm	In this paper author discussed about problem in clustering of product reviews, for the same features people can express their views in different words that are domain synonyms these words need to be grouped under same feature group. For solving this problem author used EM algorithm.	2011
[27]	S. Chandra Kala,C. Sindhu	Opinion Mining and Sentiment Classification: A Survey	Naive Bayes, Maximum Entropy, Support Vector Machine	In this paper author discussed about opinion Mining, Sentiment Analysis its approaches and various Machine Learning tools like Naive Bayes, Maximum Entropy, Support Vector Machine has been discussed.	2012
[28]	Anand Mahendran, et. al.	Opinion Mining For Text Classification	Naive Bayes, Frequency distribution.	Micro blogging services for posting views, images, videos, audios, links etc. author collected this data as raw data. They used Naive bayes classifier and frequency distribution for classifying the raw data.	2013
[29]	Maqbool Al-Maimani	Semantic and Fuzzy Aspects of Opinion Mining	Fuzzy based logic techniques SVM BN classifiers PMI-IR classifier Rule-based techniques	According to Author Opinions are fuzzy in nature and dealing with the semantic part of the expressed sentiments possesses many challenges and require effective techniques to properly extract and summarize people's views. This paper presents a review covering the semantic and Fuzzy based logic techniques and methods in sentiment analysis.	2014
[30]	Nidhi R. Sharma, Prof. Vidya D. Chitre	Opinion Mining, Analysis and its Challenges	Naive Bayesian, Expectation Maximization.	In this paper author discussed about reviews of products and services, opinion mining and also its challenges. They discussed about Expectation Maximization and Naive Bayesian algorithm.	2014

The above table information suggest that statistical techniques have been used mostly used by the researchers for extracting or mining the opinions.

6. CONCLUSION

Emotions are often associated and considered commonly significant with mood, nature, personality, disposition, and motivation. Opinion Mining or Sentiment analysis refers to extraction of opinion from given text and classify them on the basis of polarity i.e. positive, negative and neutral. In this paper, we discussed about various levels of sentiment analysis and technique used to identify and extract opinions. Here we gave some challenges used while working on opinion mining like some orthographic errors, typographical mistakes, abbreviations, colloquial words etc. are the major challenges. This paper provides a brief review to cover the major challenges, stages, application and advantages of opinion mining. In our study, we find some techniques like Naive Bayes, Maximum Entropy, and SVM etc. are very often used in opinion mining and sentiment analysis.

7. ACKNOWLEDGMENTS

We are thankful to the Computational and Psycho-linguistic Research Lab, Dept. of Computer Science & Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) for providing the facility for carrying out the research. Our thanks to the experts who have contributed towards development of the template.

8. REFERENCES

- [1] Cacioppo, 27 Nov 2014 “Studyguide for Discovering Psychology: The Science of Mind,” John Publication.
- [2] S. ChandraKala and C. Sindhu, October 2012 “Opinion Mining and Sentiment Classification: A Survey,” ICTACT Journal On Soft Computing, Volume: 03, Issue: 01.
- [3] Sharma .R. et al., March 2014 “Opinion Mining in Hindi Language: A Survey,” International Journal in Computer Science & Technology (IJFCST), Vol 4, No. 2.
- [4] Bing L., 2012 “Sentiment Analysis and Opinion Mining,” Morgan & Claypool Publishers.
- [5] Liu, B., 2010 “Sentiment Analysis and Subjectivity,” Appeared in Handbook of Natural Language Processing, Indurkha, N. & Damerau, F.J. [Eds.].
- [6] Haseena Rahmath, 2014 “Opinion Mining and Sentiment Analysis - Challenges and Applications,” International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 5.
- [7] <http://nlp.stanford.edu/software/corenlp.html>
- [8] <http://www.cs.waikato.ac.nz/ml/weka/>
- [9] <http://www.nltk.org/>
- [10] <https://opennlp.apache.org/>
- [11] <http://alias-i.com/lingpipe/>
- [12] <https://gate.ac.uk/>
- [13] <http://www.clips.ua.ac.be/pattern>
- [14] <http://www.sussex.ac.uk/Users/johnca/rasp/offline-demo.html>
- [15] B. Pang, L. Lee, and S. Vaithyanathan, 2002 “Thumbs up? Sentiment classification using machine learning techniques,” Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 79-86.
- [16] Turney P.D., 2002 “Thumbs up or down? Semantic orientation applied to unsupervised classification of reviews”, ACL, p. 417-424.
- [17] M.G. Madden, 2002 “A new Bayesian network structure for classification tasks,” Proceedings of 13th Irish Conference on Artificial Intelligence & Cognitive Science, vol. 2464.
- [18] Hai Leong Chieu . et al., “Named Entity Recognition: A Maximum Entropy Approach Using Global Information,” http://www.cnts.ua.ac.be/conll2003/pdf/16063_chi.pdf
- [19] Turney P.D., Littman M.L., 2003 “Measuring praise and criticism: Inference of semantic orientation from association,” ACM TOIS 21(4) , p. 315-346
- [20] S.-M. Kim and E. Hovy, 2004 “Determining the sentiment of opinions,” Proceedings COLING-04, the Conference on Computational Linguistics, Geneva, Switzerland.
- [21] S.-M. Kim and E. Hovy, 2005 “Identifying opinion holders for question answering in opinion texts,” Proceedings of AAAI Workshop on Question Answering in Restricted Domains, pp.20-26.
- [22] Jack G. Conrad and Frank Schilder, June 4-8, 2007 “Opinion Mining in Legal Blogs,” ACM, ICAIL '07.
- [23] Xiaowen Ding, Bing Liu, 2008 “A Holistic Lexicon-Based Approach to Opinion Mining,” ACM WSDM'08, February 11-12.
- [24] Alec Go, Richa Bhayani, Lei Huang, 2009 “Twitter sentiment classification using distant supervision,” CS224N Project Report, Stanford.
- [25] Zhao Yan-Yan, Qin Bing, Liu Ting, 2010 “Integrating Intra- and Inter-document Evidences for Improving Sentence Sentiment Classification,” ACTA AUTOMATICA SINICA, Elsevier.
- [26] Zhongwu Zhai, . et al., February 9–12, 2011 “Clustering Product Features for Opinion Mining”, WSDM'11 Hong Kong, China,.
- [27] S. Chandra Kala,C. Sindhu, October 2012 “Opinion Mining and Sentiment Classification: A Survey,” ICTACT Journal on Soft Computing, Volume: 03, Issue: 01.
- [28] Anand Mahendran, 1 June 2013 “Opinion Mining For Text Classification”, International Journal of Scientific Engineering and Technology (ISSN: 2277-1581) Volume No.2, Issue No.6, pp: 589-594.
- [29] Maqbool Al-Maimani et. al., 20th May 2014 “Semantic and Fuzzy Aspects of Opinion Mining,” Journal of Theoretical and Applied Information Technology, Vol. 63 No.2.
- [30] Nidhi R. Sharma et. al., April 2014 “Opinion Mining, Analysis and its Challenges,” International Journal of Innovations & Advancement in Computer Science IJIACS ISSN 2347 – 8616 Volume 3, Issue 1.