

# A Review on Opinion Mining Techniques and Challenges

Nikita Bansal  
B.Tech. Student  
Department of Computer  
Science  
Mangalayatan University

Prashant Pandey  
B.Tech. Student  
Department of Computer  
Science  
Mangalayatan University

Banti  
B.Tech. Student  
Department of Computer  
Science  
Mangalayatan University

Aprna Tripathi, PhD  
B.Tech. Student  
Department of Computer  
Science  
Mangalayatan University

## ABSTRACT

With the evolution of web technology, there is a huge amount of data present in the web for the internet users who use almost every available data for one to another usage in their day to day life. These users not only use data but also give their opinion in the form of feedback, thus generating additional useful information. By these user's opinions, views, feedback and suggestions available, it becomes really useful to explore, analyse and organize their views for better decision making. Nowadays, Merchants selling product on the web often ask their customers to review the products that they have purchased. As E-Commerce is becoming more and more popular, the number of customer reviews that a product receives grows rapidly. This makes it difficult for a customer to read them to make a decision on whether to purchase the product. In this research, we aim to mine and to summarize all the customer reviews of a product. And this summarization task involves the mining of the features of the product on which the customers have expressed their opinion on whether the opinions are positive or negative. This workflow of reviews is facilitated by a technique called Opinion Mining. Opinion itself means "*what other people think*" and mining illustrates "*opinions explained in the form of positive, negative or neutral comments and quotes underlying the text*". This survey gives an overview of the efficient techniques which can be implemented, discuss about the applications, recent as well as future challenges yielding in the field of opinion mining and about tools which are used to track the opinion or polarity from user generated contents.

## Keywords

Opinion mining, Data Mining, Opinion Summarization, Sentiment Analysis, Text Mining, and Web Mining

## 1. INTRODUCTION

Opinion mining involves building a system to explore user's opinions made in the form of blogs, comments, reviews or tweets, or about the product. It aims to determine the attitude of a user about a particular topic. In the recent years, the exponential increase in the Internet usage and exchange of user's opinion is the motivation for Opinion Mining.

Opinion Mining is a type of natural language processing for tracking the attitudes, feelings or appraisal of the public about particular topic, product or services. Textual information in the entire world is of two types: facts and opinions. The facts basically are the objective expressions which describe the entities, events and properties whereas the opinion is the subjective expression which describes people's opinions, emotions and sentiments towards entities and their properties. The current search engines search for facts because they assume that the facts are true and can be expressed with keywords. But these search engines do not find the opinions because opinions are very difficult to express by keywords and that is why their ranking strategy, are not appropriate for

opinion retrieval information. The introduction of Blog track in TREC 2006, a considerable work has been done in this field which comprises of opinion mining at sentence level, passage level or document level and feature level.

There are many challenges in opinion mining. The first challenge is that opinion word is considered to be positive in one condition may be considered negative in another condition. A second challenge is that people do not always state opinions in a same way. This art of Opinion Mining is useful to recognize the subjectivity and objectivity of a text and further classify the opinion orientation of subjective text. In short we say that Opinion Mining is an automated extraction of subjective content from text and identifying the orientation such as positive or negative in that text. It aims to explore feelings of a person who writes the text. It uses Natural Language Processing and Machine Learning ethics to determine opinion in the text.

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 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.

Spam filtering refers to detection and removal of fake opinions that mislead the users by giving inappropriate positive or negative opinions to some objects in order to sponsor or spoil the objects reputations. It is also a research issue in healthy opinion mining. There are lots of Free and Open Source tools available for performing Natural Language Processing and Machine Learning tasks. Following tools like GATE, NLTK, Apache Mahout, Weka, Rapidminer, KNIME, and Open NLP etc. can be used to develop your own opinion mining automated system.

The figure depicts and hence analysis the major important steps in order to achieve an opinion impact. The web users post their views, comments and feedback about a particular product or through various blogs, forums and social networking sites. Data is collected from such opinion sources in such a way that only the reviews related to the topic, that is searched is selected. The input document is then pre-processed. Pre-processing, in this content, is the removal of the fact based sentences, thus choosing only the opinionated sentences. Further refinements are made by removing the negotiations and by sensing the word disambiguation. Then, the irrelevant opinion fetched at the time of refining is then stored in another database for further reuse or for further opinion extraction. This reuse of irrelevant opinion data is a part of further research oriented feature. Then, the process of

extracting relevant features is done. Feature selection can potentially improve classification accuracy, narrow in a key feature subset of sentiment discriminators, and provide greater insight into important class attributes. The extracted features contribute to a document vector upon which various machine learning techniques can be applied in order to classify the polarity(positive and negative opinions)using the obtained document vector and finally the opinion impact is obtained based on the sentiment of the web users.

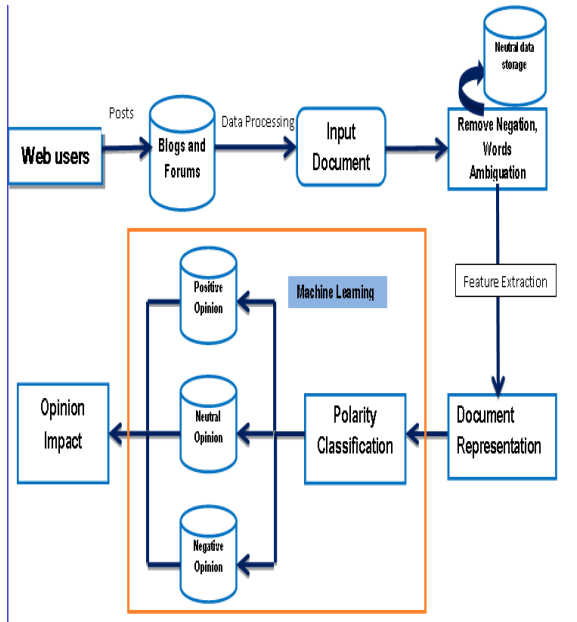


Fig 1: Work flow Diagram of Sentiment Analysis

## 2. STATE OF ART: OPINION MINING TECHNIQUES

The task of opinion mining at feature level is to extracting the features of the commented object and after that determine the opinion of the object i.e. positive or negative and then group the feature synonyms and produce the summary report. Document level opinion mining is about classifying the overall opinion presented by the authors in the entire document as positive, negative or neutral about a certain object [22, 23]. The assumption is taken at document level is that each document focus on single object and contains opinion from a single opinion holder. The most prominent work was done by Turney [17]. “Poor” and “Excellent” seed words are used by him to calculate the semantic orientation, point wise mutual information method is used to calculate the semantic orientation.

Harb et al. [16] used two sets of seed words with positive and negative semantic orientations to perform blog classification.

Meeyoung Cha et al., 2010 investigated the dynamics of public influence across different issues / topics to determine the influence author used in degree, retweets and mentions approaches on a large dataset of twitter.

The sentence level opinion mining is associated with two tasks [19, 20, 21]. First one is to identify whether the given sentence is subjective (opinionated) or objective. The second one is to find opinion of an opinionated sentence as positive, negative or neutral. Nidhi Mishra et al., has described the importance of opinion and sentiments mining. Opinion mining is distributed into four sub-categories: sentence level, document level, feature level and compound level respectively, and has

discussed the various tools and mining methods like precision, recall and F-measure on movie reviews or product reviews [3].

Table 1 summary the work done in opinion mining.

Table 1: Summary of existing opinion mining techniques

S. No	TECHNIQUES USED	YEAR	Concept	TOOLS	REF.
1.	1. Document Level 2. Sentence level 3. Task of Opinion mining at Feature level 4. Opinion Mining in Compound sentence.	2015	It classifies the opinion into different levels for better understanding and mining.	Red Opal (on feature level)[13]	[2]
2.	Normalizati on Algorithm		In this system, we rank the books based on weights assigned to them.		[4]
3.	Naïve Bayes Classifier	2013	This basically helps in deciding the polarity of data in which opinions / reviews / arguments can be classified as positive or negative	1.Web Fountain 2.Review Seer	[13]
4.	Support Vector Machine (SVM)	2013	The basic goal is to search a decision boundary between two classes that is excellently far away from any point in the training data.	Enterprise-level software	[4]
5.	Multi-Layer Perceptron (MLP)	2013	MLP is a neural network which is feed forward with one or more layers between input and output. Feed forward implies that, data flows in one direction		[4]
6.	Clustering Classifier	2013	To identifying the prominent features of human and objects and recognizing them with a type one can require the object clustering.	1.WEKA 2.LingPipe	[4], [15]
7.	Supervised Machine Learning	2014	Classification is a technique used in supervised learning which is used to predict the possible outcome from given data set on the basis of defined set of attributes.	1.Pattern 2.WEKA 3.LNTPK 4.LingPipe	[6], [15]
8.	Unsupervised Learning	2014	Clustering is a technique used in unsupervised learning. Objects in one cluster are dissimilar to the objects in other clusters.		[6]
9.	Case Based Reasoning	2014	Case based reasoning is an emerging Artificial Intelligence supervised technique. The solutions are stored in CBR repository called Knowledge base or		[6]

			Case base.		
10.	Maximum Entropy	2012	Maximum Entropy (ME) classification is used. Maximum Entropy makes no assumptions about the relationships between features		[8]
11.	Subjective Genre Classification	2004	Genre classification classifies texts into different styles, e.g., “editorial”, “novel”, “news”, “poem” etc.		[9]
12.	Text Summarization	2004	Existing text summarization techniques mainly fall in one of the two categories: template instantiation and passage extraction. They emphasize on identification and extraction of certain core entities and facts in a document.		[9]
13.	Terminology Finding	2004	In terminology finding, there are basically two techniques for discovering: symbolic approaches and statistical approaches.		[9]
14.	Spearman's rank coefficient	2015	quantitative approach is used and project the users influence against different subjects/objects		[2]
15.	Integrated mixed Methods(IMiMe) approach	2015	Mixed-methods approach is combination of different qualitative and quantitative analytical strategies to overcome the drawbacks of available ethods/tools.		[2]
16.	consumer-oriented ranking, manufacturer-oriented ranking	2015	ranking mechanism is combination of econometric analysis, text mining and subjectivity analysis		[2]

### 3. ANALYSIS OF EXISTING LITERATURE

One major concern in the analysis of user's content in online applications is to determine the polarity of opinions, by extracting the subject on which opinions are addressed and the arguments are based on.

After analysing the literature related to opinion mining, following point require attention of the researchers:

1. User opinions and sentiments are easier to extract when they are applied to an entity (e.g. product, film).
2. The detection process is done most easily at the sentence level, then by aggregating individual results, applying a certain algorithm, we get to document level.
3. There are many problems faced by systems of Opinion Mining, from which we mention language issues such as negation is an important aspect in improving systems, dictionary / lexicon to various fields and possibility of their reusability.
4. Opinion mining process is centred on a domain, so, It might possible that a same approach will not as effectively work in the other domain.

### 4. OPEN CHALLENGES IN OPINION MINING

Some most observed challenges observed in the field of opinion mining which typically trying to reduce the amount of human effort needed to classify comments are mentioned below:

- The detection of spam and fake reviews, mainly through the identification of duplicates, the detection of outliers, and the reputation of the reviewer
- A word that is considered to be positive in one situation may be considered negative in another situation. Let's take an instance of word 'long'. If a customer said that a laptop's battery life is long, that would be a positive opinion. But on the other hand, if a customer said that the laptop's start-up time is long however, that would be a negative opinion.
- People do not always express their opinion in the same way. Most of the traditional text processing relies on the fact that a small difference in opinion between two texts does not change their meaning very much. But, in opinion mining “the game was great” is very much different from “the game was not great”.
- Identifying the entity is one of the challenges in opinion mining. A text may have multiple entities associated with it. For an instance, “Ram plays better than Ravi”, here there is a positive opinion for Ram but negative for Ravi.
- Addressing the problem of sudden deviation from positive to negative polarity. For example “The movie has a great cast, superb storyline and spectacular photography; the director has managed to make a mess of the whole thing.”
- Negation is one of the challenges in opinion mining, unless handled properly can completely mislead. For example “Not only do I not approve Supernova 7200, but also hesitate to call it a phone. “has a positive polarity word approved; but its effect is negated by many negations.

### 5. CONCLUSION

Opinion Mining and Sentiment analysis has wide area of applications and it also facing many research challenges. Since the fast growth of internet and internet related applications, the Opinion Mining and Sentiment Analysis become a most interesting research area among natural language processing community. A more innovative and effective techniques required to be invented which should overcome the current challenges faced by Opinion Mining and Sentiment Analysis. In this paper, there are some techniques that are currently implemented in the field of opinion mining and overall challenges occurring in the retrieval of relevant opinion from the viewer's feedback. These techniques possess its advantages as well as drawbacks along with their respective tools being used to implement these techniques in the extraction of opinion from viewer's feedback in one or another way. There are some more tools that are being used in the current scenario of opining mining which make the work of current ongoing data extraction in this field more informative.

## 6. REFERENCES

- [1] Ion Smeureanu, CristianBucur 2012.Informatica Economica vol. 16, no.2, Applying Supervised Opinion Mining Techniques on Online User Reviews, Romania.
- [2] SaqibIqbal, Ali Zulqurnain, YaqoobWani, Khalid Hussain 2015.International Journal of Computer Science & Engineering Survey (IJCES) Vol.6, No.5, The Survey Of Sentiment And Opinion Mining For Behaviour Analysis Of Social Media, Pakistan, Malaysia.
- [3] Nidhi Mishra, C.K.Jha 2013. International Conference on Communication Systems and Network Technologies, Restricted Domain Opinion Mining in Compound Sentences, India.
- [4] Pravesh Kumar Singh, MohdShahid Husain 2013. Analytical Study Of Feature Extraction Techniques In Opinion Mining, India.
- [5] HaseenaRahmath P 2014. International Journal of Application or Innovation in Engineering &Management(IJAEM) Volume 3, Issue 5, Opinion Mining and Sentiment Analysis -Challenges and Applications, India.
- [6] Angulakshmi, Dr.R.ManickaChezian 2014.International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, An Analysis on Opinion Mining: Techniques and Tools, Pollachi, India.
- [7] AkshataShirishNayak, RoshanFernandes 2014.International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 12, An study and Analysis of Opinion Mining and Sentiment Analysis,India.
- [8] S ChandrakalaAnd C Sindhu 2012.Opinion Mining And Sentiment Classification: A Survey, ICTACT Journal On Soft Computing,Volume: 03,India.
- [9] MingqingHu,Bing Liu 2004.Mining and Summarizing Customer Reviews, Chicago.
- [10] Bo Pang and Lillian Lee 2008. Foundations and Trends in Information Retrieval Vol. 2, Opinion mining and sentiment analysis,Ithaca.
- [11] David Osimo, Francesco Mureddu 2008. Research Challenge on Opinion Mining and Sentiment Analysis,U.K.
- [12] Nidhi Mishra, C.K.Jha 2012. International Journal of Computer Applications (0975 – 8887) Volume 56–No.13, Classification of Opinion Mining Techniques,India.
- [13] G.Vinodhini, RM.Chandrasekaran 2012. International Journal of Advanced Research In Computer Science and Software Engineering, Sentiment Analysis and Opinion Mining: A Survey Volume 2, Issue 6,India.
- [14] PragatiVaidya 2015. Opinion Mining and Sentiment Analysis in Data Mining, India
- [15] AsmitaDhokra, Sunil Khillare,C. NamrataMahender 2015.International Journal of Computer Applications Technology and Research Volume 4– Issue 6,Review on Techniques and Tools used for Opinion Mining,India.
- [16] A. Harb, M. Planti, G. Dray, M. Roche, Fran, o. Troussel, and P. Poncelet,(2008), "Web opinion mining: how to extract opinions from blogs?," presented at the Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology.
- [17] movies.ndtv.com
- [18] Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring User Influence in Twitter: The Million Follower Fallacy. ICWSM, 10, 10-17.
- [19] B. Liu. 2010. Sentiment Analysis: A Multifaceted Problem., Invited paper, IEEE Intelligent Systems.
- [20] B. Liu. 2010 Sentiment Analysis and Subjectivity Second Edition, The Handbook of Natural Language Processing.
- [21] 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 (EMNLP), pp. 79–86.
- [22] B. Liu 2011. Opinion Mining and Sentiment Analysis, AAAI, San Francisco, USA.
- [23] B.Liu 2010. Opinion Mining and Sentiment Analysis: NLP Meets Social Sciences", STSC, Hawaii.