

Opinion Mining: An Overview

Ananta Arora¹, Chinmay Patil², Stevina Correia³

Student, Information Technology, Dwarkadas J Sanghvi College of Engineering, Mumbai, India ^{1,2}

Assistant Professor, Information Technology, Dwarkadas J Sanghvi College of Engineering, Mumbai, India ³

Abstract: Now-a-days due to the rapid growth of internet, people are expressing their views and opinions regarding, products, services and policies on the web in large numbers. This huge amount of feedback is very crucial for both organizations as well as individuals. The task of analysing these reviews is done by Opinion Mining (also known as Semantic Analysis). It aims for distinguishing the emotions expressed within the reviews, classifying them into positive or negative opinions and summarizing it into a form that is easily understood by users. Opinion Mining can be used by organizations to help improve their products and services. Also, it can be used by individuals in the process of decision making. This paper presents a review that covers the different techniques and approaches that are used in opinion mining systems. Also, this paper highlights various application areas and challenges related to the Opinion Mining.

Keywords: Opinion Mining, Sentiment Analysis, Supervised Learning, Unsupervised Learning, Semi-Supervised Learning, Challenges, Application Areas.

I. INTRODUCTION

In today's world, the social media, blogs, forums, e-commerce web sites, etc. encourage people to share their opinion, emotions and feelings publicly. Since these sources have become very popular among the general public, plethora of opinions and reviews are generated every day. These opinions, reviews and experiences are a valuable source of information for the decision making process. However, to generate profits or to benefit from these opinions and experiences, the accumulated content should be extracted and analysed properly. [Opinion mining is a type of natural language processing that studies the attitude, feelings or sentiments of the public about a particular topic, product or services. It involves an area of NLP, computational linguistics and text mining, and refers to a set of techniques that deals with data about opinions and tries to gain valuable information from them.] [6] Textual information can be broadly categorized into two main types: facts and opinions. Facts are objective expressions about entities, events and their properties. Opinions are subjective expressions that describe an individual's sentiments, appraisals or feelings toward entities, events and their properties. The concept of opinion is very broad. [2] Opinion mining is important for both individuals as well as companies. A customer might want to know the opinion of other customers regarding a product so as to analyse it before buying. Companies want to analyse the feedback of customers about their products to make future decisions. Opinion mining can be applied to almost any domain, from customer products, services, financial services, and healthcare to political elections and social events.

II. NEED FOR OPINION MINING

There has been a drastic change in the way people express their views and opinions over the Web. Nowadays, people can post reviews of any and every product on merchant sites. They can also express their views about anything on Internet forums, discussion groups, and blogs, which

are collectively called the user-generated content. This behaviour generates new and measurable sources of information with many practical applications. Consider an example, if at all anyone wants to purchase a commodity today, one does not have to rely solely on his/her friends and family as there are multiple reviews on the Web that gives opinions about that specific product. As customer reviews are easily available, an organization need not conduct surveys or employ consultants to search for consumer opinions regarding their products as the user generated information on the Web can give them such details. [2] However, finding opinion sources and monitoring them on the Web can still be a difficult task as there are a variety of sources, and each source may also have a huge volume of opinionated text (text with opinions or sentiments). In many cases, opinions are hidden in long forum posts and blogs. [2] Finding relevant sources, extracting related statements with opinions, reading, summarizing and organizing them into usable forms is strenuous for a human reader. Therefore, automated opinion discovery and summarization systems are required. Sentiment analysis, also known as opinion mining, grows out of this need. It is a challenging natural language processing or text mining problem. [2]

Opinion mining is a recent subclass of computational linguistics. It is not concerned with the topic of a document, but with the opinion it states. [5] In order to extract opinions from texts, recent work has tackled the issue of determining the orientation of "subjective" terms contained in texts. Subjective orientation decides whether a term that carries opinionated content has a positive or a negative connotation.

III. TYPES OF OPINION

Opinions are of two types:

A. Comparison opinion

It is basically a comparison statement that compares an object with some other similar objects. For example,

“Movie-x is not as great as movie-y.” expresses a comparison. [1]

B. Direct opinion

It specifies positive or negative opinion about an entity directly. For example, “The movie is amazing” conveys a direct opinion. [1]

IV. LEVELS OF OPINION MINING

There are three levels in opinion mining, namely [1]:

- A. Opinion Mining at Document Level
- B. Opinion Mining at Sentence Level
- C. Opinion Mining at Feature Level

A. Opinion Mining at Document level

Document level opinion mining classifies the opinion about a certain entity, stated in the entire document into three types: positive, negative and neutral. In this case, each document focuses on a single object and contains opinion from a single opinion holder. Turney presented a work based on distance measure of adjectives found in an entire document with known polarity i.e. excellent or poor. The author presents a three step algorithm [1]:

1. The adjectives are extracted along with a word that provides appropriate information.
2. The semantic orientation is captured by measuring the distance from words of known polarity.
3. The algorithm counts the average semantic orientation for all word pairs and classifies a review as recommended or not.

In contrast, Pang et al. presented a work which is based on classic topic classification techniques. The proposed technique aims to test whether a selected group of machine learning algorithms can produce good results when opinion mining is implemented at document level which is associated with two topics: positive and negative. He presented the results using Naive Bayes, Maximum Entropy and Support Vector Machine (SVM) algorithms. The resultant output that depends on the method and test data sets was better as compared to other techniques. Apart from the document-level opinion mining, the next sub-section describes the classification at the sentence-level, which classifies each sentence as a subjective or objective sentence and determines whether the opinion stated in the sentence is positive or negative.

B. Opinion mining at Sentence level

The sentence level opinion mining involves two steps:

1. To identify whether the given sentence is subjective (opinionated) or objective. To determine whether the opinion stated in the sentence is positive, negative or neutral.
2. Mining at Sentence Level assumes that a sentence contains only one opinion for e.g., “The output of this speaker is poor.” However, it is not true in many scenarios. For example, consider a compound sentence, “The picture quality of this camera is amazing and so is the battery life, but the viewfinder is too small for such a great camera”, this statement expresses both, positive and negative opinions, i.e. we say it is a mixed opinion. For “picture quality” and “battery life”, the sentence is positive, but for

“viewfinder”, it is negative. It is also positive for the camera as a whole. Riloff and Wiebe [1] used a technique called Bootstrap Approach to identify the subjective sentences and the results obtained from these tests were approximately 90% accurate. Yu and Hatzivassiloglou on the other hand, [1] talk about sentence classification i.e whether a sentence is subjective or objective, and its orientation i.e whether that sentence is positive or negative or neutral. For this, they have presented three algorithms: (1) Sentence Similarity Detection, (2) Naïve Bayes classification and (3) Multiple Naïve Bayes classification. For opinion orientation, they have used a technique that is similar to the one used by Turney [1] for document level opinion mining. Wilson et al. [1] pointed out that a single sentence may not only contain multiple opinions, but also subjective and factual clauses. It is useful to pinpoint such clauses. It is also important to identify the strength of opinions. Document-level opinion mining and sentence-level opinion mining do not consider the features of objects that have been specified in the sentence.

C. Opinion mining at Feature level

Opinion mining at feature level extracts the features of the object from the comment and determines whether the stated opinion is positive or negative. It then groups the feature synonyms and produces a summary report. Liu [1] utilized supervised pattern learning methods to extract the object features for identification of opinion orientation. To identify the orientation of opinion, he used lexicon based approach. This approach basically uses opinion words and phrases in a sentence to determine the opinion. The working of lexicon based approach is described in following steps.

1. Identification of opinion words
2. Role of Negation words
3. But-clause

V. TECHNIQUES OF OPINION MINING

There are three approaches for performing Opinion Mining [11]. They are:

A. Machine Learning Based Approach

This includes:

- 1) Supervised Learning
- 2) Unsupervised Learning
- 3) Semi-Supervised Learning

B. Lexicon Based Approach

Techniques based on Natural Language Processing (NLP) and Lexicon Based Approach utilise Parts Of Speech (POS) information and WordNet.

C. Hybrid Approach

It is basically a combination of Machine Learning Based Approach and Lexicon Based Approach.

A. Machine Learning-Based approach

This approach is comparatively more practical as compared to other approaches. This research is supported by the Fundamental Research Grant Scheme because of its automated implementation and ability to handle large collections of data on the Web.

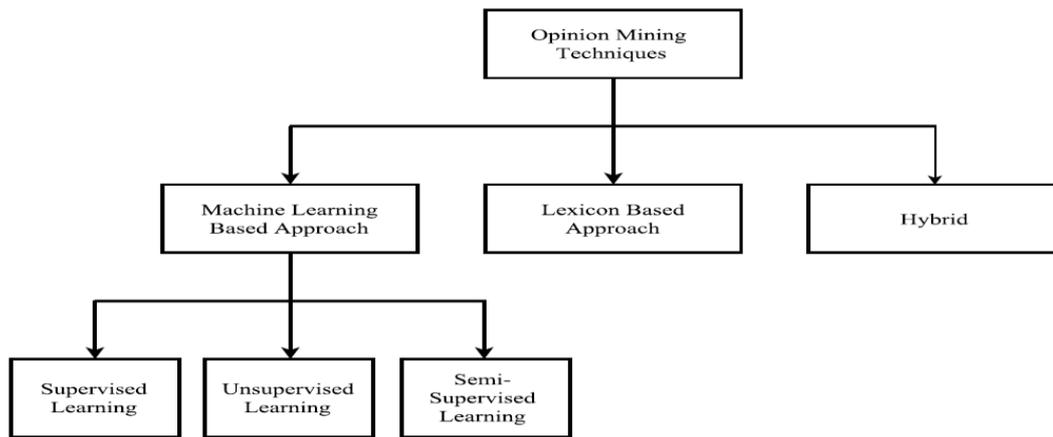


Fig1. Opinion Mining Techniques

Machine Learning-Based methods can be classified into three types: supervised, unsupervised and semi-supervised learning methods.

1) Supervised Learning:

Supervised learning is a mature and successful solution in traditional topical classification. It has been implemented to mine opinions and the results obtained are quite satisfactory. Important supervised classification algorithms include: Naïve-Bayes, a generative classifier that estimates prior probabilities of $P(X|Y)$ and $P(Y)$ from the training data and generates posterior probability of $P(Y|X)$ based on these prior probabilities; Support Vector Machine (SVM), a discriminative classifier that does not make any prior assumptions as per the training data and directly estimates $P(Y|X)$; and the lazy learning algorithm K-Nearest Neighbor (KNN), which does not require prior construction of a classification model. In both topical and opinion classification, Naïve Bayes and SVM are the most common and effective supervised learning algorithms. The biggest disadvantage of supervised learning is that it is sensitive to the quantity and quality of training data and may fail when training data is biased or insufficient. Opinion mining at the sub-document level causes more problems in case of supervised learning based approaches because the amount of information available for the classifier is very less.

2) Unsupervised learning:

In the process of text classification, many a times it becomes troublesome to construct labelled training documents. On the contrary, it is easy to construct unlabelled documents. Unsupervised learning methods serve as a solution to this problem. LDA and pLSA are examples unsupervised methods used to elicit latent topics in textual documents. These topics are nothing but features, and each and every feature is basically a distribution over terms. Accurately training a large amount of data is the prerequisite of unsupervised techniques. This is the biggest disadvantage in using unsupervised learning. Fully unsupervised models often produce incoherent topics because the objective functions of topic models do not always correlate well with human judgements. In spite of

this disadvantage, unsupervised learning helps in providing a solution to acquire knowledge about the given data.

3) Semi-Supervised learning (SSL):

SSL is a relatively new machine learning approach. SSL models tries to overcome the drawbacks of both supervised and unsupervised methods. SSL learns from both labelled and unlabelled data, hence overcoming the major drawback of supervised learning which is learning only from labelled data. Hence, using SSL with unlabelled data specially when there is fixed amount of labelled data, can help to achieve improvement over supervised learning. Also, the constraints of unsupervised learning approaches are not present with SSL if we include some form of prior knowledge to unsupervised models.

B. Lexicon-based approach

This approach searches for the opinion lexicon which is used to analyse the text. It consists of two methods. They are:

1) Dictionary-based method: It finds opinion seed words and then searches the dictionary for their synonyms and antonyms.

2) Corpus-based method: It begins with a seed list of opinion words and then searches for other opinion words in a large corpus so as to find words with context specific orientations. This can be done by using statistical or semantic methods.

V. APPLICATION AREAS

Opinion mining can be applied to almost every domain. It is used in search engines, question answering systems, recommendation systems etc. It also helps in improving human computer interactions. Opinion mining is majorly used in areas such as buying and selling of products or services, business intelligence by conducting marketing research, quality control areas, policy or decision making and so on. Sentiment analysis helps in classifying consumer feedback into 3 types: positive, negative and neutral. This provides valuable information for further analysis on market reports. [7]

The major application areas of Opinion mining are as follows [12]:

1. Buying a Commodity or Service

Now-a-days the task of buying the correct product or service from an array of options has become a child's play. Opinion mining helps in interpreting and understanding other people's experience or review about any product or service which helps in taking the right decision from a variety of options. Opinion mining takes into consideration user reviews and opinions from a large amount of data that is available on the internet, interprets it and displays it to the users in a manner that is easy to understand.

2. Market Analysis

Opinion mining can also be used as a technique to find out the current trends of consumers. Products which are liked and disliked by the end users can be found out by this technique. This is not just restricted to products. People's reaction and suggestions regarding any new government policy, rules or regulations can be found out with the help of opinion mining.

3. Suggestion Systems

Analysing people's reviews and differentiating them into positive and negative opinions can help the system to decide what should be recommended and what should not be recommended to that user.

4. Improve Products or Services

Opinion mining can serve as a boon to manufacturers. They can use opinion mining to get both, positive as well as negative feedback from the customers regarding their products or services. Based on this feedback, they can improve the lacking areas and make the necessary changes to increase their business.

5. Spam Identification

The number of internet users is increasing at a rapid rate. Since the number of users are increasing rapidly, the amount of content being uploaded is increasing at a very fast pace. This has led to high chances of spam content being posted on the web. Sometimes some users may write spam content just to confuse other users. Opinion mining can be used to differentiate between spam content and authentic content.

6. Detection of "flame"

Analysis of various social media, forums and blogs can be done with the help of opinion mining. It can be used to find out negative or arrogant words from these sources.

7. Policy Making

Since opinion mining is all about understanding people's reaction towards a particular entity, this can very well be used by the makers of different policies. These policy makers can use opinion mining to take people's feedback regarding a particular policy into consideration, and based on the approval or disapproval of the people, the policy makers decide whether any changes are to be made.

8. Government and Business Intelligence

Opinion mining, also known as sentiment analysis is well-suited for various types of intelligence applications. Business intelligence is one of the main reasons behind corporate interest in this field. [3]

For example, consider the following scenario, a car

manufacturer who is dejected due to the unexpected low sales, tries to answer the following question: "Why aren't consumers buying our cars?" Although important specifications such as the car's weight or the price of the competitor's model are quite relevant, the manufacturer needs to focus more on consumer reviews of such objective characteristics to answer this question. Also, subjective judgements regarding intangible qualities or misperceptions should also be taken into account. Like for example, "the design is tacky" or "customer service is condescending" are subjective judgements.[3]

Opinion mining techniques that are used for extracting opinions from unstructured documents are excellent tools for handling many business-intelligence tasks related to scenarios that are similar to the one we just described.[3]

With the help of public feedbacks, one can also perform trend prediction in sales or other relevant data.[3]

Government intelligence is another application that has been considered. For example, one could monitor the sources for increase in hostile or negative communications.[3]

VII. CHALLENGES

1. The primary and probably one of the most significant challenges encountered while carrying out semantic analysis on a given set of data is with respect to the credibility of the data set. For example when working on data sets containing reviews or opinions given by consumers, it is tough to judge the authenticity of these reviews or opinions. This is because some reviews or opinions might be biased as a result of brand loyalty or grudges. Also it is equally important to verify the authenticity of the users posting these reviews or opinions. [10]

2. Various techniques developed on the basis of supervised learning deliver good results, but their main drawback is that they need learning and also a knowledge base. Plenty of human efforts and time is needed to develop and integrate this knowledge base. Methods based on lexicons give high accuracy, but reduces recall, since lexicons are not available in all languages.[9]

3. Another major challenge in the process of semantic analysis is the evaluation of non-standardized data. Every user has a distinct style of conveying their opinions or reviews. The main difficulty in this is that the user may or may not use correct grammar, which may result in ambiguity. Also, now-a-days the use of acronyms and short forms is very common. This serves as a major hurdle when processing natural language and interpreting sentiments. [10]

4. Also there are certain linguistic issues in opinion mining. The language used may not always be English. It is tough to interpret any data which is not in English without translating it. Also, data available in English might be tricky to interpret especially when the data contains jargons (special words or expressions that are used by a particular profession or group) or words from local language. [10]

5. The softwares that are used for opinion mining are very expensive and it can only be afforded by the

government and other large organizations. It is beyond a common man's expectation. These softwares should be available to everyone so that every individual can benefit from it. [12]

6. Another challenge that is faced in opinion mining is the domain dependent nature of sentiment words. A set of features may perform well in one domain, but at the same time, it may not perform well in some other domain. [12]

7. Natural language processing overheads like ambiguity, co-reference, inference, implicitness etc create obstacles in opinion mining. [12]

VIII. CONCLUSION

Since people now have a liberty to post their views and opinions regarding any product or service on various sources online, the number of reviews being generated is soaring high day by day. Hence, opinion mining (also known as sentiment analysis) has become a need of the hour. By generating an efficient tool for opinion mining and allowing access to not just small and big organizations, but also the common man will prove to be a boon. It will help the common man analyze various aspects of a commodity or service before he/she decide to purchase it. The organizations on the other hand, can take important business decisions for the future and generate large profits. However, just like all other things, it has certain challenges that need to be overcome in order to get the best possible results.

REFERENCES

- [1] Nidhi Mishra, C.K.Jha, "Classification of Opinion Mining Techniques", *International Journal of Computer Applications* Volume 56– No.13, October 2012.
- [2] Bing Liu, "Sentiment Analysis and Subjectivity", Department of Computer Science, University of Illinois at Chicago, *Handbook of Natural Language Processing*, Second Edition, 2010,
- [3] Bo Pang and Lillian Lee, "Opinion mining and sentiment analysis".
- [4] Bing Liu, "OPINION MINING", Department of Computer Science, University of Illinois at Chicago.
- [5] Andrea Esuli and Fabrizio Sebastiani, "Determining Term Subjectivity and Term Orientation for Opinion Mining".
- [6] Solanki Yogesh Ganeshbhai, Bhumika K. Shah, "Feature Based Opinion Mining: A Survey", *IEEE International Advance Computing Conference (IACC)*, 2015.
- [7] Surbhi Bhatia, Banasthali Vidyapith, Manisha Sharma, Komal Kumar Bhatia, "Strategies for Mining Opinions: A Survey", *2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2015.
- [8] Bakhtawar Seerat, Farouque Azam, "Opinion Mining: Issues and Challenges (A survey)", *International Journal of Computer Applications*, Volume 49– No.9, July 2012.
- [9] Imene Guellil, Kamel Boukhalfa, "Social Big Data Mining: A Survey Focused on Opinion Mining and Sentiments Analysis".
- [10] Pooja Kherwa, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, Prashast Kumar, "An approach towards comprehensive sentimental data analysis and opinion mining", *IEEE International Advance Computing Conference (IACC)*, 2014.
- [11] Zohreh Madhoushi, Abdul Razak Hamdan, Suhaila Zainudin, "Sentiment Analysis Techniques in Recent Works", *Science and Information Conference*, 2015.
- [12] Haseena Rahmath P, "Opinion Mining and Sentiment Analysis Challenges and Applications", *International Journal of Application or Innovation in Engineering & Management (IJAIEM)*, Volume 3, Issue 5, May 2014.