

A Survey on Opinion Mining: Techniques, Tools and Research Challenges in Sentiment Analysis

K. Chitra¹, T. Kavitha², S. Hemalatha³
Assistant Professor, ^{1,2,3}Department of Computer Applications
Kongu Engineering College, Perundurai

Corresponding author: K. Chitra , kochitra@gmail.com

ABSTRACT

The Opinion mining is an ongoing field of research and development in web text mining domain. It is the computational treatment of opinions, sentiments and subjectivity of text. This survey paper focus on a comprehensive overview of the Opinion mining algorithms and the different classification with their field of applications. It is the field of computer science in which tools of software that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. The growing importance of sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. As a result, there is a huge volume of opinionated data recorded in digital form for analysis. The current research is focusing on the area of Opinion Mining also called as sentiment analysis. It extracts public opinion about the product via discussion forums, feedbacks, review sites and blogs available in digital form. One main issue in sentiment analysis of product reviews is to produce summary of opinions based on product features. We have surveyed and analyzed in this paper, various techniques that have been developed for the key tasks of opinion mining. This is very useful for the further new researchers in this field to acquire knowledge in opinion mining.

Keywords: Opinion mining, algorithms, social media, sentiment analysis, evaluations, attitudes

1. INTRODUCTION

Sentiment analysis, also known as opinion mining, is the analysis of the feelings that is people's opinions, sentiments, attitude, emotions, evaluations, appraisals towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes using natural language processing tools. Generally, opinions can be expressed about anything, e.g., a product, a service, an individual, an organization, an event, or a topic, by any person or organization. The entity to denote the target object that has been evaluated. Formally, we have the following:

Definition 1.1 (Entity) An entity e is a product, service, person, event, organization, or topic. It is associated with a pair, $e: (T; W)$, where T is a hierarchy of components (or parts), sub-components, and so on, and W is a set of attributes of e . Each component or sub-component also has its own set of attributes.

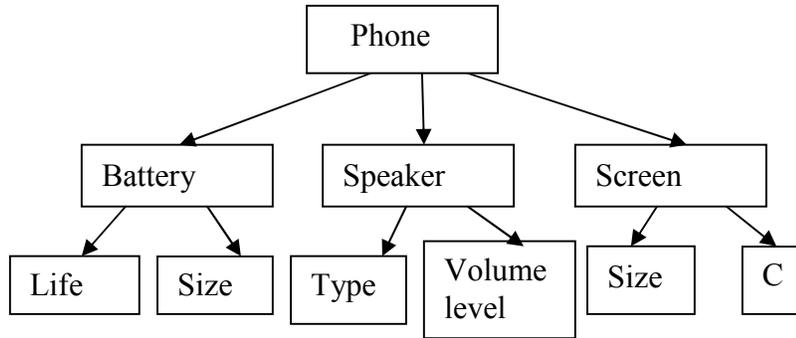


Figure. 1 Specification of a phone object in opinion mining

In above Diagram, Phone is an Object.

Battery, Speaker and Screen are the list of components in phone object.

Battery Life and Battery Size are the attributes of the battery.

Type, Volume Level is the speaker attributes.

Finally, Screen Size, Screen Color is the List of screen attributes.

Consider the example for opinion mining:

Sentence 1: I look one saree. Sentence 2: It was very beautiful. Sentence 3: Suddenly I thought to buy that saree. Sentence 4: I wish to know about the price of that saree. Sentence 5: It was high cost. Sentence 6: Therefore, I cannot able to purchase that.

In this above example, we can extract several opinions belong to positive and negative. In which, saree and I represents an entity. Petticoat and Blouse are the components. Size and color are the attributes of the Blouse and color, waist and length are the attribute of the petticoat.

In this example, sentence2, Sentence3 belongs to represents emotions explicitly. These belong to positive opinions. Nevertheless, Sentence 5 and Sentence 6 give a negative opinion about that product (saree).

Based on this definition, an entity is represented as a tree or hierarchy. The root of the tree is the name of the entity. Each non-root node is a component or sub-component of the entity. Each link is a part-of relation. Each node is associated with a set of attributes. An opinion can be expressed on any node and any attribute of the node.

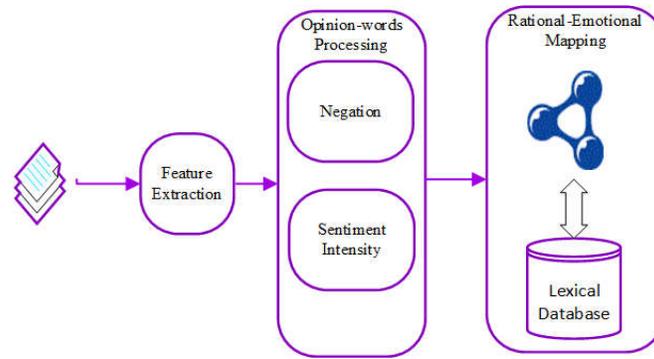


Figure.2 Emotion Detection Structure

The evaluation of opinion can be done in two ways:

- Direct opinion, gives positive or negative opinion about the object directly. For example, “The material quality of this dress is poor” expresses a direct opinion.
- Comparison means to compare the object with some other similar objects. For example, “The material quality of dress-y is better than that of dress-x.” expresses a comparison.

An opinion (or regular opinion) is simply a positive or negative sentiment, attitude, emotion or appraisal about an entity or an aspect of the entity from an opinion holder. Positive, negative and neutral are called opinion orientations (also called sentiment orientations, semantic orientations, or polarities).

2. OPINION MINING TERMINOLOGIES

This represents basic terminologies currently used in the area of opinion mining.

Fact: A fact is something that has really occurred or is actually the case.

Opinion: An opinion is a belief about matters commonly considered subjective, and is the result of emotion or interpretation of facts.

Subjective/opinionated text: A text is subjective or opinionated if it expresses personal feelings or beliefs, e.g. opinions.

Objective text: An objective text expresses some factual information about the world.

Item: An item is a concrete or abstract object such as product, service, person, event, organization. An item can be represented as a hierarchy of components, sub-components, etc.

Review: A review is a subjective text containing a sequence of words describing opinions of reviewer regarding a specific item. Review text may contain complete sentences, short comments, or both.

Short comments or pros/cons: The reviewer can describe pros and cons of the item.

3. LITERATURE REVIEW

Sentiment Analysis primarily gives a natural-language-processing point of view to help readers understand the underlying structure of the problem and the language constructs that are commonly used to express opinions and sentiments. It covers all core areas of sentiment analysis, includes many emerging

themes, such as debate analysis, intention mining, and fake-opinion detection, and presents computational methods to analyze and summarize opinions[9]. Rapid increase in internet users along with growing power of online review sites and social media has given birth to Sentiment analysis or Opinion mining, which aims at determining what other people think and comment. Sentiments or Opinions contain public generated content about products, services, policies and politics. People are usually interested to seek positive and negative opinions containing likes and dislikes, shared by users for features of particular product or service. Therefore product features or aspects have got significant role in sentiment analysis. In addition to sufficient work being performed in text analytics, feature extraction in sentiment analysis is now becoming an active area of research. This review paper discusses existing techniques and approaches for feature extraction in sentiment analysis and opinion mining. [10].The main target of the survey is to give nearly full image of SA techniques and the related fields with brief details. The main contributions of this paper include the sophisticated categorizations of a large number of recent articles and the illustration of the recent trend of research in the sentiment analysis and its related areas[11].The survey covers techniques and approaches that promise to directly enable opinion-oriented information-seeking systems. The paper aim to tackle the problem of sentiment polarity categorization, which is one of the fundamental problems of sentiment analysis. A system, called Opinion Observer, based on the proposed technique has been implemented. Experimental results using a benchmark product review data set and some additional reviews show that the proposed technique is highly effective [13].

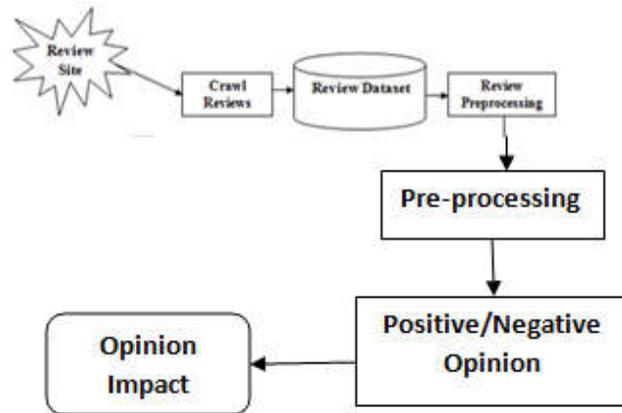


Figure.3: Opinion Mining Work Flow

4. SENTIMENT ANALYSIS PERFORMANCE

Sentiment Analysis tasks are mainly divided into the polarity of a given text at the document, sentence and feature level / attribute level / aspect level / phrase level to find whether it give positive opinion, negative opinion or neutral. This is also referred as 'Sentiment Polarity Prediction' [1]. The Sentiment Analysis performance is carried out into three levels,

- ✓ The document level
- ✓ The Sentence Level
- ✓ The Feature Level

4.1 Document Level Sentiment Analysis:

It is about classifying the overall opinionated text presented by the authors in whole document as positive, negative or neutral about a certain subject or object. Therefore, subjectivity / objectivity classification is important in this type of Sentiment Classification [2]. The main challenge in this classification is to extract informative text for deducing sentiment of the entire document.

4.2 Sentence Level Sentiment analysis:

In this type of classification, the polarity of each sentence is calculated. It is a fine-grained level than the document level sentiment classification. The sentence level sentiment classification is connected with two jobs. First one is to recognize whether the given sentence is objective or subjective opinionated. The Second one is to discover opinion of an opinionated sentence as positive, negative or neutral. Like the document classification, the sentence classification does not think about object features that have been commented in a sentence [3].

4.3 Feature Level Sentiment Analysis:

This level of sentiment classification is a much more pinpointed method to opinion mining. This type of classification considers the opinions on features of particular objects. Features of the product are defined as attributes, components and other aspects of the product, Analysis of such features are recognizing sentiment of the document is called as Feature based Sentiment Analysis [4]. The task of Feature Level sentiment classification is to extract the features of the commented object and after that conclude the opinion of the object. Positive or negative and then group the feature synonyms and make the summary report.

5. TECHNIQUES IN OPINION MINING

There are mainly three types of techniques (fig. 4):

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

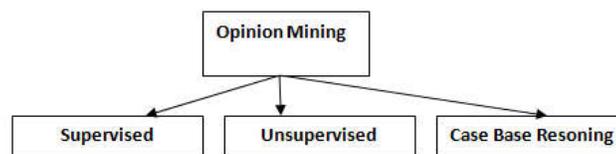


Figure. 4: Types of Opinion Mining Techniques

6. TOOLS

There are different tools are used to search the opinion or polarity from the user generated texts. These are mentioned below

i). 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.

ii). 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.

iii) 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 [5].

iv) 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.

7. OPINION MINING IN COMPOUND SENTENCE

In this section, we focus on opinion expressions in the compound sentence of a saree review that gives the opinion on the individual feature of the saree and the opinion of the saree as whole i.e. positive or negative sentiments. Apart from this, we also determine the sentiment score towards various features of a saree, such as material, color, length in meter and music. Sentiment scores are used to classify the sentiment polarity (i.e. Positive, negative or neutral) of clauses or sentences. Consider an example,

- Sentence 1: I Look one saree, it was very beautiful but length is somewhat short in my point of view
- Sentence 2: The price of the saree was too expensive, therefore I don't want to purchase it. Also that saree color is not preferable one.

The compound sentences (1 and 2) are divided into separate sentences like "*I Look one saree*" "*it was very beautiful*" "*Length is short*" "*The price of the saree was too expensive*" "*I don't want to purchase it.*" and "*saree color is not preferable one.*" After dividing the compound sentences into separate sentences or clauses, a sentiment score toward each saree features (e.g. , saree color, length and price) is calculated. (For example like the sentiment score is from 0 to 5 here 0 indicates the most negative opinion ,5 indicate the most positive opinion and between 0 to 5 we have to make the rules for positive opinion and negative opinion.) After calculating the sentiment score for each clause or sentence, the sentiment score for each review features and the overall sentiment score for the whole sentence is calculated. The following tools we use like: Sentence Delimiter [8], Part of Speech tagger [7], Named Entity recognition [7], Sentiword Net [6] for determine the opinion in compound sentence.

8. DATA SOURCE

User's opinion is the strong purpose for the improvement of the quality of services reduced and improvement of the deliverable. Blogs, review sites, data and micro blogs provide a best platform to understanding of the response level of the product and services.

8.1 Blogs: By an increasing usage of the internet, blogging and blog pages are growing rapidly. Blog pages have become the most popular means to express individual personal opinions. Bloggers record the daily events in their lives and express their opinions, feelings, and emotions in a blog.

8.2 Data Set: lots of work done in the field of movie reviews data for classification. There are many websites available for movie review data set. Additional dataset is available online is multi domain sentiment dataset available on Amazon.com, which includes like Books, DVDs, Electronics and Kitchen appliances with positive and negative reviews for each domain.

8.3 Micro-blogging: Twitter is a most popular micro-blogging service where users write messages called tweet which size is 140 characters long. These tweets sometimes express opinions about different topics. Twitter messages are also used as data sources for sentiment classifying.

8.4 Review sites: These sites help to user to purchase best product. The opinions of other users or peoples can be an important factor. A very large and growing body of user generated reviews is available on the internet. The reviews for product or services are commonly based on opinions expressed in unstructured format. The reviewers data used in most of the sentiment classification studies are collected from the ecommerce websites like Flipkart, Snapdeal, Amazon for product reviews, and yelp for restaurant reviews.

8.5 Social networking sites: Social Medias like Facebook, Twitter, and Google+ etc. are one of the biggest forums to express opinions or emotions. What is going peoples mind? They express their feeling or emotions on social media. This media helps people, company and organizations to analyses information for important decision making.

9. CHALLENGES/ISSUES

There are many issues of sentiment analysis and opinion mining following as

- The opinion word that is considered positive in one situation but may be considered negative in another situation.
- W. Jin, H. Hay Ho, and R. Srihari, 2009. Opinion Miner: A Novel Machine Learning System for Web Opinion Mining and Extraction. Proceeding of International conference on Knowledge Discovery and Data Mining Paris, France.
- To finding of spam and fake reviews, mainly through the identification of duplicates[14].
- The comparison of qualitative with summary reviews and the detection of outliers, and the reputation of the reviewer.
- The combination of opinion with behaviour to validate data and provide further analysis into the data ahead of opinion expressed [14].
- The natural language overhead like Implicitness, coreference, ambiguity, inference etc. created hindrance in sentiment analysis too.

- The biggest challenge is the domain dependent nature of sentiment words. One features set may give very good performance in one domain, at the same time it perform[15].
- The main challenging aspects exists in use of other languages, dealing with negation expressions, produce a summary of opinions based on product features or attributes, complexity of sentence or document, handling of implicit product features etc.
- Important issue is noisy texts such that with spelling or grammatical mistakes, missing or problematic punctuation and slang are still big challenges to most sentiment analysis systems.
- Each product has many names that refer to it equal within the same document and clearly across documents. This issue of automatic entity resolution is not yet solved. This is a problem for aspect extraction too, that is, how to group aspects for example, “battery life” and “power usage” refer to the same aspect of a phone.

10. CONCLUSION

This paper offers a review of research work done in several aspects of Sentiment Analysis. It describes earlier review, text mining preprocessing, data sources, sentiment analysis and its levels. It provides about an overview of opinion mining and sentiment analysis in detail with the data sources, components, methods, tools and diffenet techniques are being followed by many researchers based on domains and new applications.

The Extraction of opinion words / sentences is very challenging task. That is difficult to differentiate objective sentence or subjective informations. There are many techniques and algorithms are used to perform this task and many of the studies are remain unsolved. More future research works could be committed to these difficulties and challenges. Main contribution of this review paper is to extracting the survey of people’s opinions, emotions and sentiments in the field of opinion mining.. This review paper will be useful for researchers and beginners in the field of Opinion mining

REFERENCES

- [1] K.G. Nandakumar, Dr.T.Christopher (2015). Opinion Mining: A Survey, *International Journal of Computer Applications*, Volume 113, No.2.
- [2] Zhu Zhang, (2008). Weighing Stars, Aggregating Online Product Reviews For Intelligent E-Commerce Applications, *IEEE Intelligent Systems*, 42-49.
- [3] Ziqiong Zhang, Qiang Ye, Zili Zhang, Yijun Li, (2011). Sentiment Classification of Internet Restaurant Reviews written in Cantonese, *Expert Systems with Applications*.
- [4] B.Liu. (2010). Sentiment Analysis and Subjectivity. Second Edition, *The Handbook of Natural Lanugage Processing*.
- [5] 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*.
- [6] C. Cardie, J. Wiebe, T. Wilson, and D. Litman, (2003). Combining low-level and summary representations of opinions for multi- perspective question answering. *Proceedings of the AAAI Spring Symposium on New Directions in Question Answering*, pp. 20–27.

- [7] W. Jin, H. Hay Ho, and R. Srihari, (2009). Opinion Miner: A Novel Machine Learning System for Web Opinion Mining and Extraction. *Proceeding of International conference on Knowledge Discovery and Data Mining Paris*, France.
- [8] L. Dey, S.K. Mirajul Haque, (2009). Studying the effects of noisy text on text mining applications. *Proceedings of the Third Workshop on Analytics for Noisy Unstructured Text Data*, Barcelona, Spain.
- [9]. Bing Liu (2015), Sentiment Analysis, Mining opinions, Sentiments, and Emotions.
- [10]Muhammaad Zubair, Aurangzeb Khan, Shakeel Ahmad, FazalMasudKundi and Asghar, (2014). A Review of Feature Extraction in Sentiment Analysis. *Journal of Basic and Applied Scientific Research*.
- [11]. W. Medhat, (2014). Sentiment analysis algorithms and applications: a survey, *Ain Shams Eng. J*.
- [12]. X. Ding. S. M. Street, B. Liu, and P. S. Yu, (2008). A Holistic Lexicon based approach to opinion mining. 231-239.
- [13]. R. Feldman (2013). Techniques and applications for sentiment analysis, *Communication ACM*, 56 82-89.
- [14] Nidhi R. Sharma, Prof. Vidya D. Chitre. (2014) Opinion Mining, Analysis and its Challenges. *IJIACS* Volume 3, Issue 1.
- [15] Haseena Rahmath P (2014). Opinion Mining and Sentiment Analysis challenges and Applications, *IJAIEEM*. Volume 3, Issue 5.
- [16]. K Ravi and V Ravi (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, in Knowledge Based System.