

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

A Survey on Opinion Mining problem and levels of Analysis

Dhanashree Kulkarni¹, Prof. S.F.Rodd²

Dept. of Information Technology, DYPCOE, Talegaon, Ambi, Pune, India

Department of Information Science and Engineering, KLS, GIT, Belgaum, India

ABSTRACT: Now-a-days several websites encourage users to exchange their views, give suggestions and express opinions related to products, services and polices, Popularity of such sites has increased resulting in enormous collection of people's opinions on the web which is mainly in unstructured manner. There is a need to extract useful and informative content from these opinions. This has created a new challenging area of research called opinion mining and sentiment analysis. Opinion mining and sentiment analysis is a field of study which deals with extracting and classifying people's opinion in the form of positive, negative or neutral comments. This survey paper gives an insight of sentiment analysis problems and introduces the levels of analysis used in opinion mining.

KEYWORDS: Sentiment Analysis, Opinion Mining, Natural language processing, Information Extraction.

I. INTRODUCTION

Opinion mining is the field of study that analyses people's opinions, evaluations, appraisals, and emotions towards different entities such as products, organizations, individuals, issues, services, topics and their attributes. It is an extension of data mining which utilizes natural language processing techniques to extract people's opinion from World Wide Web and thus represents a large problem space [4].

Whenever there is a need to make a decision, people want to know other people's opinions. In today's world, businesses and organizations are very much interested in finding consumer or public opinions about their products and services. With the enormous growth of social media such as reviews, forum discussions, blogs, twitter, Facebook comments and postings in social network sites on the Web, individuals and organizations are increasingly using the content in these media for decision making. It has thus become a necessity to collect and study opinions on the Web [1].

The Web is mainly a huge repository of structured and unstructured data. The analysis of this data to extract underlying user's opinions and sentiments is a challenging task which is termed as Opinion Mining and forms the base of this paper. Opinion Mining analyses each opinionated word and determines its sentiment polarity orientation, whether it is positive or negative or neutral [2]. It gives the summarized opinion of a writer or speaker. Sentiment analysis can be done at three levels namely word level, sentence level and document level.

The organization of the paper includes Sentiment Analysis and its related challenges in section II. The levels of Analysis are explained under section III and section IV explains applications of opinion mining.

II. OPINION MINING – A CHALLENGING PROBLEM

A. The Problem of Opinion Mining

In the past, when anyone needed an opinion he/she would ask his/her friends or family. When an organization or a business needed public opinions, they would conduct surveys, opinion polls etc. But now-a-days reviews and

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

discussions on social media, blogs, comments, and postings in social network sites are considered for opinions [6]. Nowadays, if one wants to buy a consumer product like mobile or camera, one is no longer limited to asking one's friends or family for opinions because there are many user reviews and discussions in public forums on the Web.

To introduce the problem the following example of a review can be considered.

"I bought a Samsung camera two months ago. I love it. The picture quality is great. The battery life is also long. However, my friend thinks it is too heavy."

The review has a number of opinions, both positive and negative. Sentence 2, 3 and 4 expresses positive opinion whereas sentence 5 expresses negative opinion. Positive, negative and neutral are called opinion orientations or polarities. This review has opinions from two persons, which are called opinion sources or opinion holders [3]. The holder of the opinions in sentences 2, 3, and 4 is the author of the review but for sentence 5, it is author's friend.

According to Bing Liu [1], an *opinion* is a quadruple, (g, s, h, t) , where g is the opinion target, s is the sentiment about the target, h is the opinion holder and t is the time when the opinion was expressed. Given a set of opinion documents, sentiment analysis consists of the following main tasks.

Task 1 is entity extraction and categorization where all entity expressions are extracted and synonymous entity expressions are grouped into entity clusters (or categories).

Task 2 deals with aspect extraction and categorization which includes extracting all aspect expressions of the entities, and categorizing these aspect expressions into clusters.

Task 3 includes opinion holder extraction and categorization wherein opinion holders are extracted for opinions from text or structured data. The task is analogous to above two tasks.

Task 4 deals with time extraction and standardization where we extract the times when opinions are given.

Task 5 is aspect sentiment classification where it is determined whether an opinion on an aspect is positive, negative or neutral, and a numeric sentiment rating to the aspect is assigned.

Task 6 includes opinion quintuple generation which produces all opinion quintuples based on the results of the above tasks.

There are generally two types of opinions, regular opinions and comparative opinions [4]. A regular opinion expresses a sentiment only on a particular entity or an aspect of the entity, e.g., "Maaza tastes very good," which expresses a positive sentiment on the aspect *taste* of Maaza. A comparative opinion compares two or more entities based on some of their shared aspects, e.g., "Slice tastes better than Maaza" which compares Slice and Maaza based on their tastes and expresses a preference for Slice

B. Challenges in Opinion Mining

The most important words that indicate sentiments are opinion words. These words commonly express positive or negative sentiments. For example, *good*, interesting, marvellous, wonderful, *amazing* are positive sentiment words, and *bad*, terrible, ugly, poor, are negative sentiment words [12]. Although sentiment words and phrases are important for sentiment analysis there are several issues related to usage of these opinion words.

A positive or negative opinion word may have different orientations in different application domains. For example; "This movie has an unpredictable plot" is a positive phrase, while "Your vehicle has an unpredictable steering wheel" is a negative one. The opinion word unpredictable is used in different ways. Also it may happen that a sentence containing sentiment words may not express any sentiment. e.g., "Can you tell me which Hyundai car is good?" The above sentence contains the opinion word "good", but neither expresses a positive or negative opinion on any specific Hyundai car.

Sarcastic statements are much harder to deal with [10], e.g., "What a wonderful camera! It stopped working in two days." Here the word wonderful means a positive opinion but actually the statement made is negative. Sarcastic statements are not so common in consumer reviews about products and services, but are very much common in

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

political discussions. It also has to be taken into consideration that many sentences without opinion words can also imply opinions.

Opinion spamming has also become a major problem in sentiment analysis. People post fake opinions to promote or to discredit products, services, organizations etc. Such people are called opinion spammers. It is very important to detect spamming activities and ensure that the opinions on the Web are a trusted source of valuable information. Sentiment analysis is a NLP problem [6]. Hence the natural language overheads like word sense ambiguity, co-reference resolution, Implicitness, inference etc. form major challenges in tasks of opinion mining.

III. LEVELS OF ANALYSIS

Sentiment analysis can be done at word level, sentence level and document level [2].

A. Document Level

Document level task is mainly classification of the whole opinion document into a positive or negative sentiment. For example, if a product review is given, the document level task is to determine whether the review gives an overall positive or negative opinion about the product [4]. This task is commonly called as document-level sentiment classification. This analysis level assumes that each document expresses opinions on a single entity such as a single product (eg: mobile). Thus, it may not be applicable to documents which compare multiple entities.

B. Sentence Level

Sentence Level analysis takes into consideration every sentence and determines whether each sentence expresses a positive, negative, or neutral opinion. Neutral opinion usually means no opinion. This level of analysis is concerned with subjectivity classification [7]. Subjectivity classification distinguishes objective sentences from subjective sentences. Objective sentences are the ones that express factual information. Subjective sentences are the ones that express subjective views and opinions.

C. Entity and Aspect Level

The previous two levels namely document level and the sentence level analysis cannot find out what exactly people liked and did not like. Aspect level performs an analysis which is finer-grained [10]. Aspect level is sometimes called feature-based opinion mining. Aspect level mining does not look at language constructs such as documents, paragraphs, sentences etc. It directly looks at the opinion itself. An opinion not only consists of a sentiment such as positive or negative, it also consists of a target. Opinion targets are very important as they help us understand the opinion mining problem better. For example, although the sentence “although the battery is not that great, I still love this mobile” clearly has a positive tone, but it is not possible to say that this sentence is entirely positive. In fact, the sentence is positive about the mobile, but negative about its battery. In many applications, opinion targets are explained with the help of entities or their different aspects. For example, the sentence “The Samsung mobile’s call quality is good, but its battery life is short” evaluates two different aspects. One is call quality and another is battery life. The sentiment on Samsung mobile’s call quality is positive, but the sentiment on its battery life is negative. The call quality and battery life are the two opinion targets. With the help of aspect level analysis, a structured summary of opinions about their aspects can be produced. This means unstructured text is converted to structured data which can be used for all kinds of qualitative and quantitative analysis [10].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

A common form of summary is based on aspects and is called *aspect-based opinion summary* [4]. An example which summarizes a set of reviews of a camera, is shown below:

Product: Camera
Aspect: **General**
Positive: 70 <Individual review sentences>
Negative: 16 <Individual review sentences>
Aspect: **Picture quality**
Positive: 75 <Individual review sentences>
Negative: 10 <Individual review sentences>
Aspect: **Battery life**
Positive: 60 <Individual review sentences>
Negative: 12 <Individual review sentences>
...

IV. APPLICATIONS OF OPINION MINING

Opinion mining and sentiment analysis have been considered by many companies as part of their mission [5]. Opinion Mining has become an important part of search engines. Sentiment-analysis and opinion-mining systems also have a very important role in imparting sub-component technology for other systems.

The major applications of Opinion mining and sentiment analysis are given below:

- 1) **To buy a Product or Service:** While purchasing a product or service, there is a need to take a right decision about which is better and which product or service to buy. By the use of opinion mining technique, an individual can easily evaluate a product using other's opinion and experience and can easily compare different competing brands.
- 2) **Recommendation Systems:** Sentiment analysis system is said to be an augmentation to recommendation systems because by classifying the people's opinions into positive and negative, the system can say which one should get recommended and which one should not get recommend [8]. Such a system would not suggest items that receive a lot of negative feedback.
- 3) **Quality Improvement:** Companies or manufacturers can collect the critic's opinion as well as the favorable opinion about their product or service. Depending on the feedback they can improve the quality of their product or service. Online product reviews from websites can be taken into consideration.
- 4) **Detection of "flames":** Opinion mining can automatically detect arrogant words [6], over heated words or hatred language used in forum entries or tweets on various internet sources. The monitoring of newsgroup and forums, blogs and social media is easily possible by Opinion mining and sentiment analysis.
- 5) **Marketing research:** The result of sentiment analysis techniques can be helpful in marketing research so that the recent trends of consumers about some product or services can be analyzed. This can also be helpful to find out the recent attitude of general public towards some new government policy.
- 6) **Opinion spam detection:** Internet can be used by anyone and hence one can put anything on internet. Due to this there is an increase in possibility of spam content on the web. With the help of Opinion mining and sentiment analysis techniques, it will be possible to classify the internet content into 'spam' content and 'not spam' content [1].
- 7) **Policy Making and Decision Making:** Policy makers can take people's point of view towards some policy which can then be utilized in creating new citizen friendly policy. Also people's opinion and experience are considered to be very useful in decision making process.

ACKNOWLEDGMENT

It gives me great pleasure to express my feelings of gratitude to my guide Dr.S.F.Rodd, for valuable support and encouragement. I am thankful to Prof.S.D.Shirbahadurkar and Prof. Ravi Patki for their suggestions. I owe my sincere feelings of gratitude to Mr Sagar Kulkarni for his guidance which helped me a lot to write this paper.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

REFERENCES

- [1] Liu, B. (2012), "Sentiment Analysis and Opinion mining". Morgan & Claypool Publishers, May 2012.
- [2] Liu, B. (2010), "Sentiment Analysis and Subjectivity". Appeared in Handbook of Natural Language Processing, Pang, B. & Lee, L. (2008), "Opinion Mining and Sentiment Analysis". In Foundations and Trends in Information Retrieval 2 (1-2), p. 1-135.
- [2] B. Liu and L. Zhang. A Survey of Opinion Mining and Sentiment Analysis. In C. C. Aggarwal and C. Zhai, editors, Mining Text Data, pages 415-463. Springer US, 2012.
- [3] Hu, M. & Lui, B. (2004), "Mining and Summarizing Customer Reviews". In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2004 (KDD-2004), p. 168-177.
- [3] Shitanshu Verma and Pushpak Bhattacharyya, "Incorporating Semantic Knowledge for Sentiment Analysis", *Proceedings of International Conference on Natural Language Processing*, 2009.
- [4] Richa Sharma, Shweta Nigam, and Rekha Jain. Supervised opinion mining techniques: A survey.
[5] Dave K., Lawrence, S. & Pennock, D.M. (2003), "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews". In Proceedings of the 12th International Conference on World Wide Web, p. 519-528.
- [6] Dr.M S Vijaya, V Pream Sudha, "Research Directions in Social Network Mining with Empirical Study on Opinion Mining", CSI Communication Dec 2013 pp 23-26.
- [6] Jingbo Zhu, Huizhen Wang, Muhua Zhu, Benjamin K. Tsou, Matthew Ma, "Aspect-Based Opinion Polling from Customer Reviews", IEEE Transactions on affective Computing, VOL 2 No
- [7] Zhongwu Zhai, Bing Liu, Jingyuan Wang, Hua Xu, and Peifa Jia, "Product Feature Grouping for Opinion Mining", IEEE 2012.
- [7] E. Cambria, B. Schuller, Y. Xia, and C. Havasi. New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intelligent Systems, 28(2):15-21, 2013.
- [8] R. Feldman. Techniques and Applications for Sentiment Analysis. Communications of the ACM, 56(4):82-89, 2013
- [8] N. Jindal and B. Liu. Opinion Spam and Analysis. In Proceedings of the 1st ACM International Conference on Web Search and Data Mining, pages 219-230, 2008.