

Sentiment Analysis and Opinion Mining: A Survey

Suad Alhojely
Clark Atlanta University
288 Hunnicutt ST NW, Atlanta, GA

ABSTRACT

In this age, in this nation, public sentiment is everything. With it, nothing can fail; against it, nothing can succeed. Whoever molds public sentiment goes deeper than he who enacts statutes, or pronounces judicial decisions (Abraham Lincoln, 1858) [1]. It is apparent from President Lincoln's well known quote that legislators understood the force of open assumption quite a while prior. In today world, the Internet is the main source of information. An enormous amount of information and opinion online is scattered and unstructured with no machine to arrange it. Because of demand the public to know opinions about exact product and services, political issues, or social scientists. That's led us to study of field Opining Mining and Sentiment Analysis. Opining Mining and Sentiment Analysis have recently played a significant role for researchers because analysis of online text is beneficial for the market research political issue, business intelligence, online shopping, and scientific survey from psychological. Sentiment Analysis identifies the polarity of extracted public opinions. This paper presents a survey which covers Opining Mining, Sentiment Analysis, techniques, tools and classification.

Keywords

Opining Mining, Sentiment Analysis, Classification, aspect ranking, techniques

1. INTRODUCTION

Sentiment Analysis and Opining Mining represent the same field. The first use of sentiment analysis by Nasukawa and Yi, 2003[2], and the first use of opinion mining by Dave, Lawrence, and Pennock, 2003 [3]. Previously people used to obtain the information from news-paper, magazines, or web articles. Today, the people share and get information from World Wide Web to capturing and sharing their opinion. Therefore, organizations need to study public opinions for decision-making. Due to evolution and advancement of informatics enable a lot of ecommerce websites where people discuss their opinion about products and services. The consumer feedbacks contain precious and valuable knowledge for both companies and users. Thousands of products from different dealers have been offered online. For example, Amazon.com registers a total of more than 36 million products. Shoper.com records more than 5 million products from over 3,000 dealers [13]. Therefore, this will affect more and more of public to buy products online, and it will inspire the customer to purchase online and express opinions. To buy online product shopping and give reviews about product or services, the number of reviews grows rapidly. But this website and people reviews are not organized which cause difficulty for public. Nowadays, the customer opinion is critical for social influence. These reviews are critical for public because it will help users make the decision easily and helpful to the companies to know about their product's positive and negative attribute. In particular, sentiment (opinion) can be defined as opinion expressed by the consumers. Sentiment Analysis represents the opinion of the

consumer as positive (like) or negative (dislike) or may be a neutral viewpoint. The consumer will pay more attention to the aspect from the reviews, but the company will focus on improving the opinion about the product. Moreover, Sentiment Analysis can be divided into three level which are document level, sentence level, and aspect level [14]. Sentiment analysis techniques are used to express reviews, opinion, and political issues automatically from the web [15]. Sentiment Analysis is a branch of Natural Languages Processing to analysis, examines, and evaluates public opinion by using NLP. This paper explains some techniques for sentiment analysis and focuses on sentiment classification. Section 1 will provide literature survey. Section 2 will explain Sentiment Analysis Classification. Section 3 will illustrate an aspect product ranking by using SA. Section 4 will write about Sentiment Analysis Classification Technique. Section 5 will give tools used in Sentiment Analysis. Section 6 will be the acknowledgment. Section 7 will be the conclusion. Section 8 will be the references.

2. LITERATURE SURVEY

Nasukawa and Yi [2] clarify that instead of state the complete document into positive or negative, they express sentiments connect with positive or negative for a particular topic from a document. Also, they illustrate the fundamental issue in sentiment analysis which is knowing the sentiment expressed in texts whether the sentiment shows positive or negative opinion. Taylor et al. [4] show that a design system of opining mining for tourism which will be beneficial in a lot of industries. Also, they propose a system it is ti solve problem in Lake Distract tourism industry. Haddi, Lui and Shi [5] explain the sentiment of back movie feed on online. They try to decrease the nose in the text by using a collection of different pre-processing methods and using the chi-squared method to delete irrelevant advantage that does not impact its orientation. Haddi, Lui and Shi have explained many extensive experimental outcomes. By showing that on two data sets proper text pre-processing accuracy accomplished is similar to the sort of accuracy that could be accomplished in subject categorization, it is an easier problem. In Moraes, Valiati and Neto [6] concentrated on comparing between SVM and ANN under the condition of the requirement to achieved good classification accuracies. Also, experiments evaluated all methods as a function in bag-of-words (uni grams) approach in particular terms. Related sentiment learning literature the necessary contributions/findings are in two points. The first point is that in term of classification accuracy on a benchmark dataset of movies reviews. The second point is as a complete comparison in the context of balanced data.

3. SENTIMENT CLASSIFICATION

Sentiment classifications are based on polarity, which may become positive, negative, or neutral. That's mean opinions may be classified into positive, negative, or neutral. Moreover, there is a forth type which is a constructive opinion

which obtains suggestion to make the product better [7]. Opinions are classified into three categories: the first one is direct opinions which opinion holder directly attack to target. Second one of opinion is comparative opinions which are opinion holder compare among entity. The third one is indirect opinions, which are implied as in idioms or expressed in a reverse way as in sarcasm. Researchers have studied sentiment analysis into three level:

3.1 Document Level Sentiment Classification

Document level sentiment classification aims to classify the entire document as positive or negative. There is much actual work use one of the two types of classification techniques which are a Supervised method and Unsupervised method to build level document sentiment.

3.1.1 Supervised method:

Sentiment classification is performed at document level sentiment [1]. Sentiment classification can be used as a supervised classification problem with four classes positive, negative, neutral, and constructive [8] [9] [10] [11]. Also, supervised request machine-learning algorithms like SVM Support Vector Machines to conclude the relationships between the opinions that expressed and text segment. A lot of researchers found that supervised learning techniques can perform well in SVM and Naïve Bayes (Pang et al. 2008) [8].

3.1.2 Unsupervised method:

Unsupervised classification is performed at the sentence level [1]. There are two types of unsupervised classification, which are lexicon-based, and syntactic-pattern based. Sentence and aspect level sentiment classification for the lexicon-based can be used.

3.2 Sentence Level Sentiment Classification:

In this level, the task is to determine each sentence in the document as positive or negative opinions. Sentence level sentiment analysis has classified the polarity. This level is close to document level but here it accomplished by every sentence [12]. However, there may be complex sentences in the text which make the sentence level is not helpful. There are two phases in level sentence sentiment done in every single sentence: first, each sentence classified, as subjective or objective, and the second one is the polarity of subjective sentence are concluded.

3.3 Aspect Level Sentiment Classification:

It supposes that a document has a hold opinion on many entities and their aspects. Aspect level classification needs discovery of these entities, aspects, and sentiments for each of them.

4. NEW PRODUCT ASPECT RANKING

Chetan Mate in India proposed a product aspect ranking framework by using sentiment analysis automatically to distinguish the main aspects of a special product on online consumer reviews[18]. He suggested a work that will base on consumer reviews in four steps. The first step of the product aspect ranking framework is that Reviews Extraction and Preprocessing. This step is imperative because it is based on data preprocessing. Sentiment analysis applied on unstructured reviews which cause poor performance. So, the preprocessing techniques to get a better result in sentiment analysis are needed. Therefore, the methods for preprocessing are stemming and tokenization. Stemming which is to delete

the postfix from a word such as ing, tion, etc. In tokenization, the space from each sentence as showing in figure 1 will be removed. The second step of the product aspect ranking framework is that Aspect Identification of the product. There are many consumer reviews available on ecommerce website but the reviews collected in different formats on a different website that make a problem. Also, some website is an overall grading on any product like CNet.com, another website the reviews are in paragraph in format text like Viewpoints.com. Hu and Liu proposed an approach that identifies the noun and a noun phrase. The noun and noun phrase are counted then only appropriate noun is saved as the aspect [16]. The third step of the product aspect ranking framework is Sentiment Classifier. Sentiment analysis is the type of Natural Languages Processing which used in tracking polarity of people about product reviews. Sentiment analysis classifies a text as positive, negative, or neutral. There are two types of classification techniques learning: supervised learning and unsupervised learning. The lexicon-based approaches are unsupervised which use it will obtain sentiment polarity on each aspect of word and phrase [17]. On the other hand, the supervised learning depends on the training dataset. Also, he mentions to many learning based classification model like SVM, Naïve Bayes, and Maximum Entropy (ME). The forth step of product aspect ranking framework is Aspect Ranking Algorithm. The overall opinion in the reviews of the product is a total of opinion given to specific aspects in that review. So, the sentiments on important or unimportant aspects have strong and weak impacts of the overall opinion. Their proposed framework uses AFINN dictionary approach, which consists of many English words rated valence with an integer between (-5) and (+5) as positive and negative. They attempt to create an aspect-ranking algorithm to identify the important aspects.

5. SENTIMENT ANALYSIS CLASSIFICATION TECHNIQUES

Sentiment Analysis Classification Techniques divided to three techniques, which are Machine Learning (ML) approach, Lexicon-based approach, and Hybrid approach. The ML approach stratifies the ML algorithms and uses linguistic advantage. The classification methods that use ML approach can be divided into Supervised and Unsupervised learning method. The Lexicon-based approach depends on sentiment lexicon. Lexicon considers as an important indicator for sentiment, which called opinion word. Lexicon divided into dictionary-based approach and corpus approach. The dictionary-based approach finds opinion seed words, and then search the dictionary of their antonyms or synonyms. But the corpus-based approach starts with a list of seed opinion and then finds another opinion in a big corpus to try help finding opinion words in context. The corpus approach uses the statistical or semantic method to determinate the polarity of sentiment. The Hybrid approach is a combination of ML approach and Lexicon-based approach, and it is very popular in Lexicon-based approach as shown in figure 2.

6. TOOLS OF SENTIMENT ANALYSIS CLASSIFICATION

There are so many open-source text-analytics tools used for natural language processing such as information extraction, and classification can also be applied for sentiment analysis. The following are tools used for Sentiment Classification:

1. NTLK: A natural language toolkit is a tool for text processing, cataloging, tokenization, stopping, tagging, parsing, etc. It provides easy-to-use interfaces to more

than 50 corpora and lexical resources such as Word Net, along with a suite of <http://www.nltk.org/>

2. GATE: Beneficial if you want to create a pipeline. Language analysis modules for developers contribute various languages are available to be applied plugged in your pipeline.
3. 3. OpenNLP: perform the most common NLP tasks, such as POS tagging, named entity extraction, chunking, and co-reference resolution. <http://opennlp.apache.org/>StanfordCoreNLP: If you need part of speech categories, syntactic analysis (phrase structure or dependency analysis), co-reference or named entities in text.
4. OpinionFinder: It supports to identify individual sentences and to crate different parts of subjectivity in these sentences, including the conclusion holder of the subjectivity and words that are incorporated into expressions communicating positive or negative suppositions. <http://code.google.com/p/opinionfinder/>
5. Ling Pipe: Ling Pipe is used for linguistic processing of text including, clustering cataloging and entity extraction, etc. <http://alias-i.com/lingpipe/>

7. CONCLUSION

Sentiment Analysis and Opining Mining are very important research because Sentiment Analysis help in summarizing opinion and reviews of public. They consider as research filed. However, Sentiment Analysis still need to improve and progress. Moreover, there are many challenges like the polarity in a complex sentence. In addition, the vocabulary of natural languages is a lot which causes difficulty. This survey highlights the basic ideas about Sentiment Analysis and then explains in details the Sentiment Classification, Technique Classification, tools that available for Sentiment Analysis, and a new feature which is Product Aspect Ranking. In particular, knowing the consumer opinion and the overall opinion about any product have been identified the aspects ranking algorithm, which will be known the most important aspects and consumer opinions given to every aspect on their overall opinions.

8. ACKNOWLEDGEMENT

The author would like to thank Dr. Roy George for his supports and suggestions. These supports and suggestions are greatly appreciated. Saudi Arabian Cultural Mission funds research of Suad Alhojely.

9. REFERENCES

- [1] Goutam Chakraborty, Murali Pagolu, Satish Garla, "Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS" Pp 181-184, 2013.
- [2] T. Nasukawa, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing Definition of Sentiment Expressions," pp. 70–77, 2003.
- [3] K. Dave, I. Way, S. Lawrence, and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," 2003.
- [4] E. Marrese-Taylor, J. D. Velasquez, and F. Bravo-Marquez, "Opinion Zoom: A Modular Tool to Explore Tourism Opinions on the Web," 2013 IEEE/WIC/ACM

Int. Jt. Conf. Web Intell. Intell. Agent Technol., pp. 261–264, Nov. 2013.

- [5] E. Haddi, X. Liu, and Y. Shi, "The Role of Text Pre-processing in Sentiment Analysis," *Procedia Comput. Sci.*, vol. 17, pp. 26–32, Jan. 2013.
- [6] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," *Expert Syst. Appl.*, vol. 40, no. 2, pp. 621–633, Feb. 2013.
- [7] R. Arora and S. Srinivasa, "A Faceted Characterization of the Opinion Mining Landscape," pp. 1–6, 2014.
- [8] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Found. Trends® Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, 2008.
- [9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10, 2002.
- [10] D. Bessalov, B. Bai, A. Shokoufandeh, and Y. Qi, "Sentiment Classification Based on Supervised Latent n-gram Analysis," pp. 375–382, 2011.
- [11] M. Gamon, "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis."
- [12] H. Yu and V. Hatzivassiloglou. "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of the 2003 conference on Empirical methods in natural language processing, EMNLP '03, pages 129-136.
- [13] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, Tat-Seng Chua, "Product Aspect Ranking and Its Applications". IEEE2014.
- [14] Bing Liu, "Sentiment Analysis and Opinion Mining" pp.7-140,2012.
- [15] Anand Mahendran and Anjali Duraiswamy, "Opinion Mining for text classification", *International Journal of Scientific Engineering and Technology (2277-1581)*, Vol No.2,2013
- [16] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proc. SIGKDD*, Seattle, WA, USA, 2004, pp. 168–177.
- [17] Bing, Liu "sentiment analysis and opining mining" pp 7-140, 2012.
- [18] Chetan Mate "Product Aspect Ranking using Sentiment Analysis: A Survey" " Vol No 1,2015.

10. APPENDIX

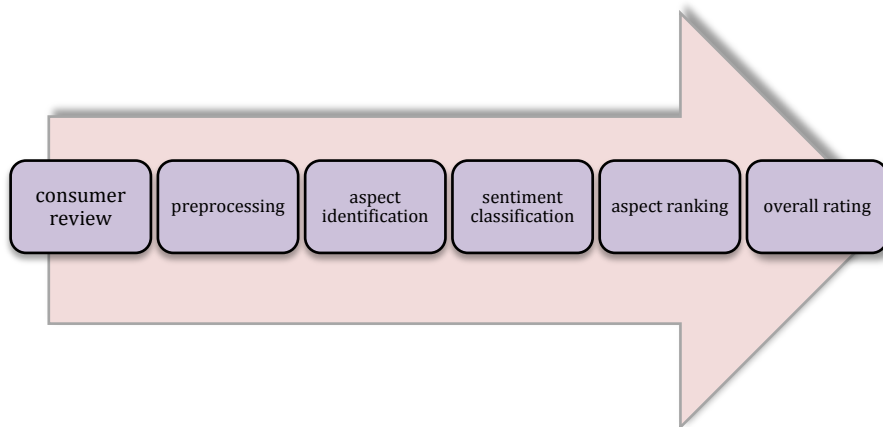


Fig 1: Architecture Diagram.

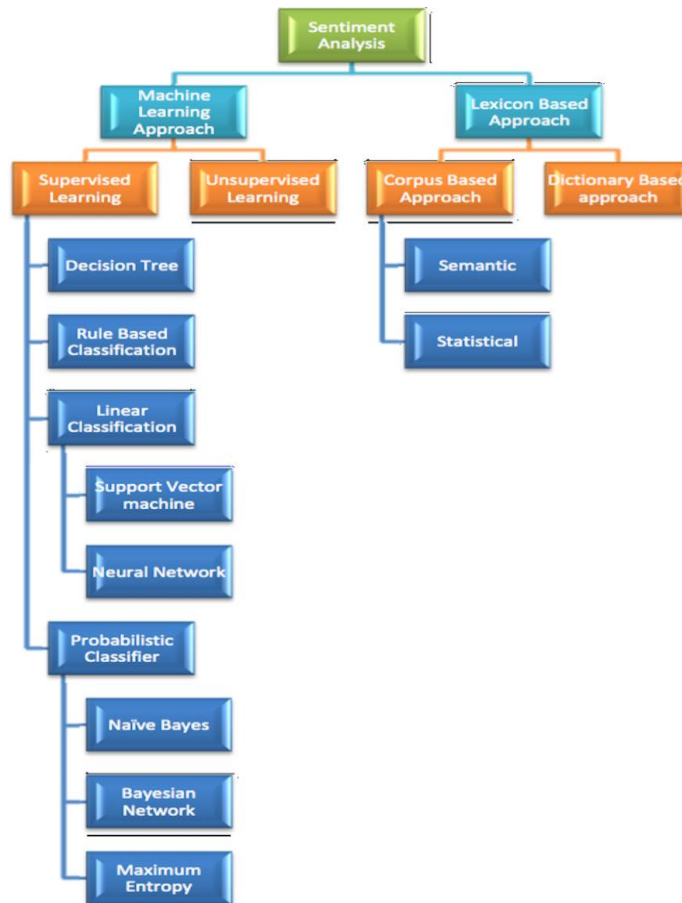


Fig. 2: Sentiment Classification Techniques.