



ORIGINAL RESEARCH PAPER

Computer Science

SENTIMENT ANALYSIS APPROACHES, METHODS AND APPLICATIONS: A SURVEY

KEY WORDS: Sentimental Analysis, Opinion Mining, NLP, SVM, applications, classification.

Dr. Vishal M. Pandya	Assistant Professor in Sutex Bank College of Computer Applications and Science, Surat, Gujarat, India.
Dr. Kejal C. Vadza*	Assistant Professor in Sutex Bank College of Computer Applications and Science, Surat, Gujarat, India. *Corresponding Author
Dr. Viralkumar B. Polishwala	Assistant Professor in Sutex Bank College of Computer Applications and Science, Surat, Gujarat, India.

ABSTRACT
 Web 2.0 gave regular people the power to express their feelings, opinions, views, reviews, feedbacks, to the world. Social network platforms, feedback & review forums of online shopping websites, micro-blogging platforms like twitter, feedback & review pages on hotel booking sites and many more other online places are flooding with user generated content. This content is very valuable for individuals, businesses, political parties, and to government also. This user generated content having opinions and sentiments of general people, users, customers, prospective customers, voters, citizens etc is very useful, but at the same time it is in very large amount. And that's why it requires the use of automated mining and analyzing tools and techniques. Sentiment Analysis is a process of resource creation for Sentiment Analysis, sentiment retrieval and sentiment extraction and polarity detection [1]. In layman's terms sentiment analysis is the automated extraction of attitudes, opinions, and emotions from text, using Natural Language Processing (NLP) methods. This paper presents a survey on Sentiment analysis applications and challenges with their approaches and techniques.

1. INTRODUCTION

With the advancement of technology, users of smart phones and internet connected devices are increasing exponentially, and with that e-commerce is booming. People prefer online shopping because of the advantages over traditional shopping. We plan out trip, book hotels and flights online. But it is hard to make decisions for buying products or booking a hotel room without physically experiencing them. So they mainly rely upon other users' feedback and review. What other people think has always been an important piece of information of us during the decision making process [2]. For a decade web users are providing content actively as product feedback, on social media, e-commerce sites and micro blogging sites. Huge amount of web text is available with different sentiments, which is very useful to business and general users too. But it has been a challenging task and an area of research interest for a few years. The researchers have named this area as Sentiment Analysis and Opinion Mining.

Sentiment analysis also known as opinion mining is a process that uses Machine Learning techniques to mine opinions and emotions from database, speech or text. Basic sentiment analysis does classification of opinionated text into "positive", "negative" or "neutral". It is often referred as subjectivity analysis, opinion mining, and appraisal extraction [2].

The activities or say research areas that fall under sentiment analysis are:

- i. Subjectivity Detection: It is the process of discovering if the text is opinionated or not.
- ii. Sentiment Prediction: It is the task of detecting the polarity i.e. "positive" or "negative", for the given text that may be a phrase, a sentence or a document.
- iii. Aspect based sentiment summarization: is about providing sentiment summary in the form of score points of star ratings of characteristics.
- iv. Product feature extraction: It is the process of extracting features of the product from its review or feedback text.
- v. Opinion spam detection: It is the task of recognizing bogus

or fake opinion text from reviews, posts or feedbacks.

These tasks related to sentiment analysis are done at different levels; phrase level, sentence level and document level. This paper is directed towards reviewing the idea of Sentiment Analysis through Natural Language Processing and exploring its applications and challenges. In this paper section 2 mentions most popular approaches and methods for Sentiment Analysis. Section 3 describes popular and novel applications of Sentiment Analysis. Section 4 narrates challenges faced in the process of Sentiment Analysis. And section 5 is the conclusion of the paper.

2. Approaches and methods for Sentiment Analysis

2.1 Methods

It is possible to implement sentiment analysis using both supervised and unsupervised classification techniques. Compared to unsupervised methods, supervised methods have demonstrated superior performance. Unsupervised approaches, however, are also crucial because supervised methods require a lot of expensive, labeled training data, but acquiring unlabeled data is simple. With the exception of movie reviews, the majority of domains lack labeled training data; in this situation, unsupervised approaches are particularly helpful for creating applications.

2.1.1 Supervised Methods

Building a classifier enables the implementation of supervised approaches. The examples used to train this classifier can be manually labeled examples. Support Vector Machines (SVM), Naive Bayes classifiers, and Maximum Entropy are the three most popular supervised algorithms. Supervised strategies have been shown to perform better than unsupervised techniques [3]. SVMs are better suited for sentiment classification, according to Cui et al. [4], have argued that SVMs are more appropriate for sentiment classification because they can better perform when review contains both positive and negative words. However, a Naive Bayes classifier may be better appropriate when the training data set is minimal because SVMs need a big amount of data to create a high-quality classifier. Choosing the right collection of features for sentiment classification is one of the most crucial challenges for supervised technique.

Some of the popular features used in sentiment classification are Part of Speech tagging, Negations, Term presence and their frequency, and opinion words and phrases. Part-of-speech tagging is used to disambiguate sense which in turn is used to guide feature selection [2]. It is also used to tag adjectives and adverbs in the sentences. Adjectives and adverbs are identified as opinion words and nouns are identified as features. Opinion words and phrases express positive or negative emotion, e.g. "good", "bad", "like", "love it", "not great" etc. Negative words in the sentence may change the polarity of the text, for example "The product is good" and "The product is not good" have opposite polarity.

2.1.2 Unsupervised Methods

In an unsupervised technique, classification is carried out by contrasting a text's features with word lexicons whose sentiment values have been predetermined. To determine sentiment, for instance, evaluate the material using positive and negative word lexicons first. The document is positive if it has more positive word lexicons; else, it is negative. The most prominent work done using unsupervised methods for opinion mining and sentiment detection is by Turney [5]. He chooses the seed words "bad" and "outstanding" since they are more frequently found online while calculating the semantic orientation of phrases, which is determined by the reciprocal information between the points in the phrase. The average semantic orientation of all such sentences is used to calculate the sentiment of a document. Part-of-speech (POS) patterns are employed by Ting-Chun Peng and Chia-Chun Shih [6] to extract the sentiment phrases from each review. They used an unidentified sentiment phrase as their search keyword and obtained the top-N relevant terms from a search engine. Based on the k-means clustering algorithm, Gang Li and Fei Liu [7] devised a method for grouping papers into positive and negative groups.

2.2 Approaches

There are different approaches to the task of Sentiment Analysis, and the selection depends upon the type of job on hand and the researcher's perspective. The sentiment analysis approach could be keyword-based, relationship-based, discourse-based, or knowledge based (language model based) [8].

2.2.1 Knowledge based approach

The fundamental assignment in this approach is the development of word dictionaries that show positive class or negative class. The opinion upsides of the words in the not entirely settled preceding the feeling examination work. Vocabularies can be made in various ways. It tends to be made by beginning with some seed words and afterward utilizing a few etymological heuristics to add more words to them, or beginning with some seed words and adding to these seed words different words in view of recurrence in a text. SENTIWORDNET 3.0 is a freely accessible lexical asset expressly conceived for supporting feeling order and assessment mining applications [9].

Pang et al. [3] found that term- presence gives better results than term frequency while using language based approach for sentiment analysis of movie reviews. Also their research of movie reviews shows that uni-gram presence is more suited for sentiment analysis. Dave et al. [10] found that bigrams and tri-grams worked better than uni-grams in sentiment classification they performed on product reviews.

2.2.2 Relationship-based approach

In this approach the various connections among highlights and parts is broke down for feeling arrangement task. Such connections might be connections between various members, connections between item includes. For instance, to know the feeling of clients about an item brand, one might process it as an element of the opinions on various highlights or parts of it.

2.2.3 Keyword-based approach

Keyword based sentiment analysis consist of three main steps namely keyword extraction, sentiment prediction, sentiment classification and optional summarization module. Sentiment analysis tools categorize pieces of writing as positive, neutral, or negative. Positive sentiment may be expressed using words such as "good", "great", "wonderful", and "fantastic". Negative sentiment may be expressed using words such as "bad", "terrible", "awful", and "disgusting".

2.2.4 Discourse-based approach

Discourse-based sentiment analysis refers to the task of analyzing the sentiment expressed in a larger text unit, such as a document or a conversation, taking into account not just individual words or sentences, but also their relationships and context within the larger discourse structure. This type of sentiment analysis is more challenging than simple sentence-level sentiment analysis, as it requires understanding of the discourse and rhetorical structure of the text and ability to track sentiment shifts and changes throughout the text. Discourse structures and semantics approach involves discourse connection between text parts for arrangement. In many surveys, the general opinion is typically communicated toward the finish of the text [3]. In this talk driven approach the opinion of the entire survey is extracted by deciding feeling between various text parts and the topic relations that exist between them. In such a methodology, the last section of the survey may be given more weight in the assurance of the feeling of the entire review.

3. Applications of Sentiment Analysis

3.1 E-Commerce

Ecommerce operations are where sentiment analysis is used the most frequently. Websites give users the option to submit comments regarding their shopping and product-quality experiences. By appointing rates or scores, they offer a description of the product and its various aspects. Customers can access opinions and recommendation data on both the entire product as well as certain product features with ease. Users are shown a graphic description of the whole product and its characteristics. Popular shopping websites like amazon.com include editorial and customer reviews along with ratings. A well-known website that offers reviews on hotels and tourist places is <http://tripadvisor.in>. 75 million thoughts and reviews are present in them globally. Such websites benefit from sentiment analysis since it turns unhappy customers into advocates by examining this vast array of viewpoints.

3.2. Brand monitoring and reputation management

The management of your brand's reputation in the marketplace is a concern. Customer or other parties' opinions have the power to improve or harm your reputation. Brand Reputation Management (BRM) is more customer-focused than product- or company-focused. Online one-to-many communications are now happening often. This gives businesses the chance to control and enhance their brand reputation. Now Advertising, public relations, and corporate messaging are merely a few of the factors that affect brand perception. Today, brand discussions make up the whole. Sentiment analysis aids in evaluating how the community is perceiving a company's brand, product, or service online.

3.3. Government

By examining public sentiment, sentiment analysis aids government in determining their strengths and flaws. "If this is the state, how do you expect the truth to emerge, for instance? The MP who is looking into the 2G scandal is seriously corrupt. This illustration amplifies public dissatisfaction with the administration.

We can see the possibilities for sentiment analysis in many other contexts, such as monitoring public opinion on a new 108 system, determining the successes and failures of a

government job recruitment effort, evaluating the success of electronic tax return submission, and many other situations.

3.4.Voice of Market

The Voice of the Market is a service that offers the information required to connect with potential clients, hold a competitive advantage, and expand your company. Precision marketing is what Voice of the Market does. Early detection of such information aids in the targeting and directing of important marketing initiatives. Sentiment analysis enables businesses to obtain real-time client feedback. They can use this real-time data to create new marketing strategies, enhance product features, and forecast the likelihood that a product will fail. Zhang et al. [11] proposed weakness finder system which can help manufacturers find their product weakness from Chinese reviews by using aspects based sentiment analysis.

3.5.Voice of Customer

Customer experience's voice of the customer (VOC) element focuses on the requirements, desires, expectations, and preferences of the customer. The level of customer satisfaction is typically a crucial differentiator from rival companies. Obtaining client feedback also aids in identifying non-functional criteria for items, such as performance and price.

3.6.Product analysis

As soon as a new product is released, find out what the public is saying about it. You may also examine feedback from years ago that you may have never seen. For a certain product feature (interface, UX, functionality), you can conduct a keyword search and apply aspect-based sentiment analysis to find only the data you require. Learn how a product is viewed by your target market, what needs to be changed about it, and what would satisfy your most valuable clients with sentiment analysis.

3.7.Market and competitor research

Use sentiment analysis for market and competitor research. Find out who's receiving positive mentions among your competitors, and how your marketing efforts compare. Analyze the positive language your competitors are using to speak to their customers and weave some of this language into your own brand messaging and tone of voice guide.

CONCLUSION

A significant research issue is the use of sentiment analysis to mine the enormous volume of data. The most popular applications and difficulties in sentiment analysis are outlined in this study. Researchers and corporate organizations are currently competing to develop the most effective sentiment analysis system. Even though some algorithms employed in sentiment analysis produce outstanding results, no algorithm is able to handle every problem. Support Vector Machines (SVM), according to the majority of researchers, offers higher accuracy than other algorithms but also has drawbacks. It is discovered that sentiment classification depends on the domain. Combining several categorization algorithms will help them overcome their flaws, capitalize on each other's strengths, and improve overall performance. Every firm wants to know how customers feel about their goods and services and those of their rivals, thus there is a great need in the market for such applications. Sentiment analysis can be created for new applications. Although there has been significant advancement in sentiment analysis methods and algorithms, there are still many challenges that need to be resolved. Future research can be conducted to address these challenges.

REFERENCES:

[1]. Pandya, V.M. and Bhatwala, P., redefining sentiment analysis based on related concepts and tasks. International Journal of Research in Technology and Management (IJRTM) Volume 2 Issue 5, JULY 2016
 [2]. B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and

Trends in Information Retrieval 2(1-2),2018, pp. 1–135.
 [3]. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," Proceedings of the ACL- 02 conference on Empirical methods in natural language processing, vol.10, 2020, pp. 79-86.
 [4]. H. Cui, V. Mittal, and M. Datar, "Comparative Experiments on Sentiment Classification for Online Product Reviews." In Proceedings of AAAI-06, 2020, pp.1265-1270.
 [5]. P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," Proceedings of the Association for Computational Linguistics (ACL),2020, pp.417–424.
 [6]. T. Peng, C. Shih, "An Unsupervised Snippet- Based Sentiment Classification Method for Chinese Unknown Phrases without Using Reference Word Pairs." Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technology,2019, pp.243-248.
 [7]. G. Li, and F.Liu, "A Clustering-Based Approach on Sentiment Analysis," IEEE International Conference on Intelligent System and Knowledge Engineering, Hangzhou, China, vol. 2019/ 1, pp.331-337.
 [8]. G. Gebremeskel, "Sentiment Analysis of Twitter Posts about News." Master's Thesis, University of Malta, May 2019.<http://proudtobeindian.net/indian-mediaexposed>
 [9]. S. Baccianella, A. Esuli, and F.Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," Proceedings of the Seventh conference on International Language Resources and Evaluation,2020,pp.2200-2204.
 [10]. K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," Proceedings of WWW,2020, pp.519–528.
 [11]. W. Zhang, H. Xu, W. Wan, "Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis," Expert Systems with Applications, Elsevier, vol. 39, 2019, pp. 10283-10291.
 [12]. M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," Computational Linguistics, vol. 37, 2019, pp. 267-307.
 [13]. M. Hu and B. Liu, "Mining and summarizing customer reviews," Proceedings of the tenth ACM international conference on Knowledge discovery and data mining, Seattle, 2019, pp.168-177.
 [14]. S. Baccianella, A. Esuli, and F.Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," Proceedings of the Seventh conference on International Language Resources and Evaluation,2020,pp.2200-2204.