International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 5 ǁ 2022 ǁ PP. 52-56

An Overview of Sentiment Analysis: Approaches, Applications, and Challenges

Minakshee Narayankar
Ishita Tambat
Nisha Gharpure
Suja Jayachandran
Department Of Computer Engineering, Vidyalankar Institute of Technology, Vidyalankar College Rd., Wadala East, Mumbai - 400037, Maharashtra, India.
Corresponding Author: Minakshee Narayankar

Abstract
The rapid proliferation of Internet-based applications, such as social media platforms and blogs, has resulted in comments and evaluations about everyday activities. The practice of obtaining and analyzing people's ideas, thoughts, and perceptions about various themes, products, subjects, and services is known as sentiment analysis. Corporations, governments, and individuals can benefit from people's opinions in terms of gathering information and making decisions based on that knowledge. However, the process of sentiment analysis and evaluation is fraught with difficulties. These difficulties make it difficult to interpret emotions effectively and choose the appropriate sentiment polarity. Sentiment analysis uses natural language processing and text mining to identify and extract subjective information from text. This article covers various state-of-the-art mechanisms to perform sentiment analysis. It delves into the use of sentiment analysis in a variety of fields. Finally, in order to determine future paths, the problems of sentiment analysis are discussed.
Keywords: *Sentiment analysis, Opinion mining, Text analysis, Machine learning, Social media.*

Date of Submission: 18-04-2022 Date of acceptance: 03-05-2022

I. INTRODUCTION

Sentiment analysis has gained widespread acceptability in recent years, not only among researchers, but also among enterprises, governments, and organizations. The Internet's expanding popularity has elevated it to the status of the primary source of universal knowledge. Many people use a variety of online resources to convey their thoughts and opinions. We must use user-generated data to evaluate it automatically in order to keep track of public opinion and improve decision-making. As a result, sentiment analysis has grown in popularity in recent years among research communities. Opinion analysis or opinion mining are other terms for sentiment analysis. The sentiment analysis task has recently grown in popularity. In the early stages of sentiment analysis, numerous research papers presented an outline of Opinion mining.

The proliferation of social networking sites has spawned a host of professions dedicated to studying these networks and their contents in order to extract useful data. Sentiment analysis is the process of determining the emotions conveyed by a piece of text based on its content. Sentiment analysis is a branch of NLP, and given the long and distinguished history of public opinion in decision-making, there must be a plethora of early publications on the subject. However, it continues to develop sentiment analysis into the next millennium. Sentiment analysis uses Natural Language Processing to extract attitudes, opinions, and emotions from text, audio, and database sources. Sentiment analysis is the process of categorizing textual opinions into categories such as "positive," "negative," or "neutral." Subjectivity analysis, opinion mining, and assessment extraction are all terms used to describe this process.

Subjectivity Detection, Sentiment Prediction, Aspect Based Sentiment Summarization, Text Summarization for Opinions, Contrastive Viewpoint Summarization, Product Feature Extraction, and Detecting Opinion Spam are the main subjects of research in Sentiment Analysis. The task of determining whether or not the text is opinionated is known as subjectivity detection. Predicting the polarity of text, whether positive or negative, is what sentiment prediction is all about. Sentiment summaries are provided in the form of star ratings or feature scores via Aspect Based Sentiment Summarization. Text Summarization creates a few sentences that summarize a product's reviews. The emphasis of Contrastive Viewpoint Summarization is on opposing viewpoints. Product Feature Extraction is a task that takes a review and extracts the product's features. Identifying fake or fraudulent opinions from reviews is a concern when it comes to detecting opinion spam.

Sentiment prediction can be done at three levels: document, sentence, and phrase. The overall sentiment of the document is characterized as positive, negative, or objective at the document level. Individual sentiment-bearing sentences are classified using sentence-level prediction. The polarity of phrases in a sentence is categorized at the phrase level. The goal of this work is to introduce the notion of sentiment analysis in the context of natural language processing, as well as to investigate its applications and limitations.

II. APPROACHES AND TECHNIQUES

2.1. Approaches

The technique can be discourse-driven, relationship-driven, language model-driven, or keyword-driven, depending on the task at hand and the viewpoint of the individual performing sentiment analysis.

2.1.1 Knowledge-based approach

The creation of word lexicons that signify the positive or negative class is the key effort in this technique. Prior to the sentiment analysis, the sentiment values of the terms in the lexicon are established. Lexicons can be made in a variety of ways. It can be made by starting with a few seed words and then adding more using linguistic heuristics, or by starting with a few seed words and adding more words depending on their frequency in a text. SENTIWORDNET 3.0 is a freely accessible lexical resource that was created specifically to enable sentiment classification and opinion mining applications.

2.1.2 Relationship-based approach

The different links between characteristics and components are investigated in this technique for the sentiment classification job. Relationships between different participants or product attributes are examples of such relationships. For example, if you want to know how people feel about a product brand, you may calculate it as a function of how they feel about particular aspects or components of the product.

2.1.3 Language models

The n-gram language models are created using this method. It is possible to use the presence or frequency of n-grams. The frequency of n-grams in text classification yields better outcomes. In most cases, the frequency is transformed to TF- IDF to account for the term's importance in a document. Term presence, on the other hand, can occasionally produce greater results than term frequency. Unigram's presence is better suited for sentiment analysis, according to their investigation of movie reviews. When it came to categorizing product reviews, bi-grams and tri-grams were sometimes more effective than unigrams.

2.1.4 Discourse structures and semantics

This method classifies text components based on their discourse relationships. The overall impression is frequently conveyed at the end of numerous reviews. The sentiment of the entire review is determined in this discourse-driven technique by determining the sentiment between different discourse components and the discourse interactions that exist between them. In this case, the final paragraph of the review may be given more weight in determining the overall sentiment of the review.

2.2. Techniques

2.2.1 Supervised Techniques

Building a classifier can be used to implement supervised approaches. Examples that can be manually labeled are used to train this classifier. Support Vector Machines (SVM), Naive Bayes classifier, and Maximum Entropy are the most commonly used supervised algorithms. Supervised Techniques have been found to outperform unsupervised techniques in terms of performance [1].

SVMs, according to Cui et al. [2], are better for sentiment classification since they may perform better when a review contains both positive and negative words. When the amount of training data is limited, however, a Naive Bayes classifier may be preferable because SVMs require a big amount of data to develop a high-quality classifier. Selecting an adequate set of features is one of the most critical jobs in sentiment categorization. Below are the most widely used features in sentiment categorization.

- Term presence and their frequency:

Unigrams or n-grams, as well as their frequency or presence, are among these characteristics. In sentiment categorization, these features have been frequently and successfully used. In movie review sentiment analysis, Pang et al. [1] say that unigrams produce better results than bi-grams, however, Dave et al. [3] argue that bi-grams and tri-grams produce superior product-review polarity categorization.

- Part of speech information:

The part-of-speech method is used to decipher sense, which is then used to guide feature selection [4]. Part-of-speech tagging is effective for detecting adjectives and adverbs in sentences that identify opinion words and nouns that define product characteristics.

- Negations:

Negation is a key aspect to consider because it has the ability to reverse a sentiment [4]. "The movie is great" and "The movie is not fantastic," for example. The negation word "not" in the second statement makes it negative.

- Opinion words and phrases:

Words and phrases expressing favorable or negative opinions include "like," "lovely," "hate," "I'd propose," and "I'd suggest..." Statistical and lexicon-based approaches are the most common methods for determining the semantic orientation (positive or negative) or polarity of opinion words.

The main limitation of supervised learning is that it is dependent on the amount and quality of the training data and may fail when training data are insufficient.

2.2.2 Unsupervised Techniques

The properties of a given text are compared against word lexicons whose sentiment values are determined prior to their use in an unsupervised technique. Start with positive and negative word lexicons, then analyze the document for the sentiment you're looking for. The document is positive if it has more positive word lexicons; else, it is negative. Turney [5] is the most well-known example of applying unsupervised algorithms for opinion mining and sentiment identification. He uses "poor" and "excellent" seed words as they appear more on the web for calculating the semantic orientation of phrases, where orientation is measured by pointwise mutual information. The average semantic orientation of all such sentences is used to calculate the sentiment of a document.

Ting-Chun Peng and Chia-Chun Shih [6] employed part-of-speech (POS) patterns to extract the sentiment phrases of each review. They used an unknown sentiment phrase as a query term and used a search engine to find the top-N relevant terms. The feelings of unknown sentiment phrases are then calculated using lexicons based on the sentiments of neighboring known relevant phrases. Based on the k-means clustering algorithm, Gang Li and Fei Liu [7] devised an approach for clustering documents into positive and negative groups.

III. APPLICATIONS

Sentiment analysis can be used in a variety of sectors and for a variety of objectives. This section goes over a few of the more frequent ones. The examples in this section aren't exhaustive; they're only a taste of what's possible.

3.1 E-Commerce

The most common application of sentiment analysis is in e-commerce. Users are able to share their buying and product quality experiences on websites. They provide ratings or scores to the product to provide a summary of the product and its various characteristics. Customers can quickly access customer reviews and recommendations for the entire product as well as individual product aspects. Users are given a graphical description of the whole product and its features.

Popular merchant websites, such as amazon.com, include reviews from editors as well as customer reviews with ratings. http://tripadvisor.in is a well-known website that offers hotel and tourist location reviews. They have 75 million reviews and comments from all over the world. By evaluating this massive amount of data, sentiment analysis assists such websites in converting dissatisfied clients into promoters.

3.2 The Market's Voice (VOM)

The goal of Voice of the Market research is to learn how customers feel about competitors' products and services. The Voice of the Market provides accurate and timely information that aids in competitive advantage and new product development. Early detection of such information aids in the targeting and direction of critical marketing activities. Sentiment Analysis enables businesses to obtain client feedback in real-time. This real-time data aids in the development of new marketing strategies, the enhancement of product features, and the prediction of product failure.

Zhang et al. [8] suggested a vulnerability detector technique that uses aspects-based sentiment analysis to help manufacturers identify product flaws in Chinese evaluations. Radiant6, Sysomos, Viralheat, Lexalytics, and other commercial and free sentiment analysis services are available. There are also some free programs accessible, such as www.tweettfeel.com and www.socialmention.com.

3.3 The Customer's Point of View (VOC)

Concern for what individual customers have to say about products or services is known as the "voice of the customer." It entails examining client testimonials and feedback. Customer Experience Management relies heavily on VOC. VOC aids in the discovery of new product development potential. Customer feedback also aids in the identification of product functional and non-functional needs such as performance and cost.

3.4 Brand Reputation Management

Brand Reputation Management refers to the process of controlling your brand's reputation in the marketplace. Customer or third-party opinions can either hurt or help your reputation. Brand Reputation Management (BRM) is more concerned with the product and company than with the customer. One-to-many discussions are becoming commonplace on the internet. As a result, businesses can better manage and improve their brand reputation. Advertising, public relations, and corporate messaging no longer determine brand perception. Brands have evolved into a sum of the discourse surrounding them. Sentiment analysis aids in determining how the public perceives a company's brand, product, or service on the internet.

IV. CHALLENGES

Sentiment analysis is a text categorization task that categorizes text as positive, negative, or otherwise objective. There are numerous classes in text categorization since there are various themes, but there are only three in sentiment analysis. However, when compared to traditional text classification, there are a number of issues that make sentiment analysis problematic. Some of the factors are as follows.

4.1 Coreference resolution

The challenge of determining what a pronoun or a noun phrase refers to is known as coreference resolution. "We went to supper and watched the movie; it was terrible," for example. What exactly does "It" mean? For sentiment analysis based on topics or aspects, coreference resolution may be effective. Opinion mining accuracy could be improved by using coreference resolution.

4.2 Temporal Relations

For sentiment analysis, the time of the reviews may be relevant. In 2008, the reviewer may have thought Windows Vista was wonderful, but in 2009, he may have had an unfavorable opinion due to the new Windows 7. As a result, examining these types of attitudes that fluctuate over time may increase the sentiment analysis system's effectiveness. This allows us to see if a product improves over time or if people's opinions about a product alter.

4.3 Sarcastic sentences

Sarcastic and sarcastic sentences may appear in the text. "What a great car," for example, "it stopped working on the second day." Positive phrases can take on a negative connotation in this situation. Sarcastic or sarcastic statements can be difficult to spot, leading to incorrect opinion mining.

4.4 Requirement of World Knowledge

To correctly categorize the text, knowledge of world facts, events, and individuals is frequently essential. Consider the following example [9]: "Casablanca with a rice and fish lunch: a pleasant Sunday."
The above sentence is classified as positive by the system because of the word "good," however it is an objective sentence because Casablanca is the title of a popular film.

4.5 Domain Considerations

The domain of the items to which sentiment classification is applied can affect its accuracy. The reason for this is that many words have different meanings depending on the domain. "Read the book," for example [4]. This sentence has a positive tone in the book domain but a negative tone in the movie domain.

4.6 Grouping synonyms

Many times, a text will have multiple words with the same meaning. For appropriate classification, such words should be identified and grouped together. The challenge of identifying these terms is tough since people frequently use different phrases to describe the same trait. In phone reviews, for example, the terms "voice" and "sound" both refer to the same feature.

4.7 Thwarted Expectations

Some sentences in a text begin with a different context and end with a different context. "The cast was not good, and the actors played horribly," for example, "but I liked it."

The last statement in the previous review makes the entire review good. Because there are more negative words in the review, the aforementioned assertions would be classified as negative if term frequency were taken into account.

4.8 Negation

Small variations between two bits of text don't impact the meaning very much in traditional text classification. Sentiment analysis, on the other hand, distinguishes between "the movie was wonderful" and "the movie was not great." Negative handling in sentiment analysis is tough since it reverses the polarity. Negation can also be expressed through sarcasm and implied phrases that do not include any negative words

4.9 Review Spam Detection

Many people publish fake reviews, often known as review spam, on product review sites to promote their own items by giving undeservedly positive recommendations or to slander their competitors' products by giving fraudulent bad reviews. The task of detecting opinion spam has a significant impact on industrial communities. The users' experience will be harmed if the opinion-given services contain a big quantity of spam. Furthermore, if the offered opinion deceives the user, he will never use the system again. experience. Furthermore, if the offered opinion deceives the user, he will never use the system again.

V. CONCLUSION

The use of sentiment analysis to harvest massive amounts of data has become a major academic topic. This document outlines some of the most often utilized sentiment analysis applications and obstacles. Now, commercial groups and academia are collaborating to develop the finest sentiment analysis technology. Although various algorithms have been utilized in sentiment analysis and have produced positive results, no algorithm can tackle all of the problems. Support Vector Machines (SVM) have higher accuracy than other algorithms, according to the majority of academics, however, it also has drawbacks. The classification of sentiment is found to be domain-dependent. Different types of classification algorithms should be integrated to overcome their unique flaws, profit from each other's strengths, and improve classification accuracy.

Such apps are in high demand in the market since every company wants to know how customers feel about their products and services, as well as those of their competitors. New applications for sentiment analysis can be created. Although sentiment analysis techniques and algorithms have progressed, there are still many obstacles to overcome in this discipline. More research can be done in the future to address these issues.

VI. REFERENCES

[1]. Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10 (EMNLP '02). Association for Computational Linguistics, USA, 79–86. 10.3115/1118693.1118704

[2]. Hang Cui, Vibhu Mittal, and Mayur Datar. 2006. Comparative experiments on sentiment classification for online product reviews. In Proceedings of the 21st national conference on Artificial intelligence - Volume 2 (AAAI'06). AAAI Press, 1265–1270

[3]. Dave, Kushal & Lawrence, Steve & Pennock, David. (2003). Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. 775152. 10.1145/775152.775226

[4]. Bo Pang and Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. Found. Trends Inf. Retr. 2, 1–2 (January 2008), 1–135. 10.1561/1500000011

[5]. Turney, Peter. (2002). Thumbs Up or Thumbs Down? {S}emantic Orientation Applied to Unsupervised Classification of Reviews. Computing Research Repository - CORR. 417-424. 10.3115/1073083.1073153

[6]. T. Peng and C. Shih, "An Unsupervised Snippet-Based Sentiment Classification Method for Chinese Unknown Phrases without Using Reference Word Pairs," 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2010, pp. 243-248, doi: 10.1109/WI-IAT.2010.229

[7]. Gang Li and Fei Liu, "A clustering-based approach on sentiment analysis," 2010 IEEE International Conference on Intelligent Systems and Knowledge Engineering, 2010, pp. 331-337, doi: 10.1109/ISKE.2010.5680859

[8]. Wenhao Zhang, Hua Xu, and Wei Wan. 2012. Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. Expert Syst. Appl. 39, 11 (September, 2012), 10283–10291. 10.1016/j.eswa.2012.02.166

[9]. Joshi, Aditya & R, Balamurali & Bhattacharyya, Pushpak & Mohanty, Rajat. (2011). C-Feel-It: A Sentiment Analyzer for Micro-blogs.. 127-132. 10.13140/2.1.3834.7200