# A Comprehensive Study on Methods, Challenges and Applications of Sentiment Analysis: A Survey

# Ravina Nimmi Rodrigues

Department of Information Technology, Padre Conceição College of Engineering, Verna, Goa – India ravinarodrigues@gmail.com

**Abstract.** In today's modern era of the internet, a large number of people share their opinions on websites, social media, blogs and discussion forum. The opinions can be about services, events, people, organizations, products, and their attributes. Meaningful information can be derived from such data using sentiment analysis. Sentiment Analysis analyses people's thoughts, attitudes and feelings from their posting. The goal is to analyze the data and then categorize it as either positive, negative or neutral. This analysis is very beneficial for individuals, businesses, and even governments. This paper presents an outline of sentiment analysis along with a comprehensive survey of its methods, challenges and current applications.

Keywords: Sentiment analysis, Opinion mining, Movie review, Machine learning, Lexicon based, Hybrid.

Date of Submission: 26-07-2019

Date of Acceptance: 12-08-2019

#### I. INTRODUCTION

\_\_\_\_\_

\_\_\_\_\_

Since the evolution of internet, tremendous amount of data is being created every day. The data originates from various sources such as social media, microblogging, ecommerce, online transactions, and discussion forum. We can classify textual information into two main types: facts and opinions. Facts are objective expressions about entities such as services, organizations, people, issues, events, products. Opinions are subjective expressions that describes people's emotions towards entities. In today's world people share their opinion regarding anything such as films, news, products and even politics on the internet. The opinions are in the form of short to moderate text messages, which can be analyzed to get useful insights for consumers and business.

Sentiments can be described as emotions, judgements, opinions or ideas prompted or colored by emotions [1]. Sentiment Analysis is the area of study that examines people's sentiments, attitudes and emotions towards entities. Whether the given text has a positive, negative or neutral attitude, opinions are used to define polarity of the text. Sentiment analysis have many names such as opinion mining, opinion extraction, sentiment mining, emotion analysis and review mining. However, they are now all under the umbrella of sentiment analysis or Opinion mining [2].

Before the advent of the internet era, people would seek opinion from their family or friends before buying a product or watching a movie. But now people tend to seek out opinion from various customers of that product around the globe by just reading their reviews. Hence Sentiment analysis play a vital role in our day-today decisions which may range from buying a product such as a laptop to watching a movies or even making financial and business investments. All these decisions will have a tremendous influence on our day to day life. Nowadays people before buying any product/service will make a glance on review sites (e.g. CNET), e-Commerce sites (e.g. Amazon, eBay) and social media (e.g. twitter) to get a feedback about the specific product (or) service in market [3].The different levels of sentiment analysis are given below:

#### **1.1 Document level**

The aim of this level is to classify whether the entire document expresses a positive, negative, or neutral sentiment. A neutral sentiment has no opinion about the document. For example, given a review about an entity, the system expresses an overall opinion about that entity. This level is commonly known as document-level sentiment classification [2]. Such analysis is effective only if individual document expresses opinions on a one entity, i.e., a single product.

#### **1.2 Sentence level**

In sentence level the essential focus goes to each sentence and determine whether each sentence expresses a positive, negative, or neutral opinion. This level of analysis is closely related to subjectivity classification [4], which differentiates objective sentences that express factual information from subjective

sentences that express subjective views and opinions.

#### **1.3 Aspect level**

The document level as well as the sentence level analysis gives us the total sentiment and not specific to what exactly a person liked or disliked. Aspect level performs a detailed analysis by looking at the opinion itself. It is based on the fact that an opinion consists of a sentiment (positive, negative or neutral) and a target (of opinion). Aspect level was earlier called feature level [2]. Consider an example, "The Xiaomi's processor is slow, but its camera is good" evaluates the two aspects, processor speed and camera, of Xiaomi (entity). The sentiment on Xiaomi's processor is negative, but the sentiment on its camera is positive. Here the processor and camera are the opinion targets.



Fig. 1. Shows the complete procedure of sentiment analysis which begins from pre-processing of review documents till generation of sentiment polarity.

#### II. SENTIMENT ANALYSIS APPROACHES

Sentiment analysis approaches are basically divided into three categories. These three approaches are Machine learning, Lexicon based and Hybrid approaches [5].

#### 2.1 Machine Learning

Machine learning is a field of study in artificial intelligence. It has the ability to make a computer system, learn on its own and improve from experience without being specifically programmed. The learning process starts by looking for patterns in data and making decisions without human intervention. It is further categorized as supervised and unsupervised. Supervised learning is an approach in which the machine using data which is labelled, analyses new data and produces a correct outcome from the labelled data. Unsupervised learning is the training of system using data that is neither classified nor labelled and allowing the algorithm to act on it without any guidance.

#### 2.2 Lexicon Based Approach

In lexicon based approach, we assume that the overall orientation of the sentiment is the addition of the sentiment orientation of each word or phrase. Two methods are used under Lexicon based approaches first one is a dictionary-based method and the second one is the corpus-based method. For dictionary-based, dictionaries can be generated manually using initial words to expand the list of words. On the other side, the corpus-based method makes use of given initial list plus a general purpose sentiment lexicon used for a specific domain.

#### 2.3 Hybrid Approach

The hybrid approach uses a mixture of both machine learning and lexicon based techniques. Some researchers [6-7] have developed sentiment classification systems which are more accurate than machine learning or lexicon based approaches.



Fig. 2. Taxonomy of Sentiment Analysis Approaches

# III. SENTIMENT ANALYSIS CHALLENGES

#### 3.1 Language

Languages such as English, is very well used in opinion mining because of its resource availability such as corpora, dictionaries and lexicons. However with other languages, challenges surface for building such resources.

#### **3.2 Polarity Shift**

Polarity shift is a linguistic phenomenon which can contradict the sentiment polarity of a given word. For e.g. by adding a word don't in front of the word like ,to the sentence, "I like the new Samsung A50.", the sentiment polarity of the sentence completely changes from positive to negative.

#### 3.3 Sarcasm

Sarcasm is a form of statement that is deliberately used to mock somebody by using words with the opposite of their literal meaning. However, identification of sarcasm is challenging due to the gap between its exact and intended meaning.

#### 3.4Typo errors

Typo errors makes it difficult to comprehend the meaning of words. Such errors include spelling mistakes, abbreviation, punctuation and grammatical errors.

#### 3.5 Fake Opinion

It is also known as Fake review. It states to spurious reviews which misguide the readers or customers by providing them dishonest opinions related to any entity.

#### **IV. APPLICATIONS OF SENTIMENT ANALYSIS**

#### 4.1 Social Media Monitoring of public figures

In addition to traditional news media, social media platforms have become an important site for conversations on public figures throughout the world. The conversations can range from science, culture, business to celebrities and even politics. Moreover business executive or public figures choose social media to promote biography or to publish important activity that they want the world to know. Such reviews can then be analyzed to obtain people's sentiment towards that activity.

#### 4.2 Brand Monitoring

Internet contains a wealth of information. People commonly share their opinion on various brands on social media sites as well as blogs and forums. Brands can now look largely across the web to understand public sentiments. They should observe the quantity as well as the quality of those mentions. Sentiment analysis is useful in brand monitoring because it helps to understand how a brand reputation progresses over time.

#### 4.3 Customer Feedback/Voice of Customer (VoC)

Sentiment analysis can analyze millions of online reviews, free-text survey responses and social media comments. Understanding exactly how customers feels, and why they feel that way can grow the market, deliver innovative products and build customer loyalty.

#### 4.4 Workforce Analytics & Voice of the Employee

For any company to do well, it is very essential to have a good employees as well as good working environment. There is always a constant pressure on the management to retain good and qualified employees. Performing sentiment analysis on employee feedback or survey can not only track changes in employee sentiment over time but also give actionable ideas on how to improve the organization.

#### 4.5 Market Research and Analysis

Sentiment analysis endows all kinds of competitive market research and analysis. Whether it is keeping an edge on the competition, anticipating future trends, or exploring a new market, sentiment analysis can make all the difference.

## V. LITERATURE REVIEW

A brief description of the papers used in the survey is given below. All of the research work has been executed in the problem domain of sentiment analysis of movie reviews.

The authors [8], proposed a document based opinion mining system that classify the documents sentiment. The system also handles negation. WordNet is used as a dictionary to determine the opinion words and their synonyms and antonyms. The polarity of the given documents is calculated based on the majority of opinion words.

This paper [9], presents the document level sentiment classification addressing the polarity shift problem using a three-stage cascade model. It uses a hybrid sentimental method which is a combination of hybrid polarity shift detection, polarity shift elimination in negations, and polarity shift based ensemble model.

The paper [10] proposes a rule based domain independent sentiment analysis. The suggested technique classifies subjective and objective sentences from reviews and blogs by making use of lexicon based sentiment approach on sentence level.

The paper [11] includes sentiment analysis of movie reviews using feature-based opinion mining and supervised machine learning. The main focus is to determine the polarity of reviews using nouns, verbs, and adjectives as opinion words. All opinion words are selected from the sentence to find the polarity from it.

In this paper [12], a new hybrid classification method is suggested based on coupling classification methods using arcing classifier. Its performance is analyzed in terms of accuracy. A Classifier ensemble was designed using Naive Bayes (NB) and Genetic Algorithm (GA).

The authors [13] evaluate the sentiment expression to classify the polarity of the movie review on a scale of 0 to 4. They execute feature extraction and ranking and use them to train their multilabel classifier to classify the movie review into its correct label. The methodology based on structured N-grams is used.

Tuble 1. Comparison of Sentiment Approaches				
Sr.No	Author name	Sentiment	Sentiment	Accuracy
		Level	Approach	
1.	Richa Sharma ,Shweta Nigam and Rekha	Document	Lexicon Based	63%.
	Jain	Level		
2.	RuiXiaa,d, Feng Xub, JianfeiYua, Yong	Document	Hybrid	87.1%
	Qia,b , Erik Cambriac	Level		
3.	Aurangzeb Khan, BaharumBaharudin and	Sentence Level	Lexicon Based	86.6%
	Khairullah Khan			
4.	Gurshobit Singh Brar 1,* , Asst. Prof.	Feature Level	Machine	81.22%
	Ankit Sharma2		Learning	
5.	M.Govindarajan	Feature Level	Hybrid	93.80%
6.	Tirath Prasad Sahu and Sanjeev Ahuja	Feature Level	Lexicon Based	88.95%

**Table 1.** Comparison of Sentiment Approaches

### VI. CONCLUSION

Sentiment analysis has become a widespread research subject lately. This paper presented a summary on sentiment analysis as well as its approaches, challenges and current applications. A survey on six research papers was implemented in the domain of sentiment analysis of movie reviews. All of the papers used different approaches and achieved different accuracy level. It was observed that by performing sentiment analysis on aspect level of the reviews using a hybrid model gave the best accuracy. So I conclude that a hybrid model on aspect level of the opinion gives the optimal sentiment analysis.

#### REFERENCES

- [1]. Boiy, E. Hens, P., Deschacht, K. & Moens, M.F (2007), "Automatic Sentiment Analysis in Online Text," In Proceedings of the Conference on Electronic Publishing, pp. 349-360, 2007.
- [2]. Bing Liu (2012), Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers.
- [3]. Ms.A.M.Abirami,Ms.V.Gayathri (2016), A survey on sentiment analysis methods and Approach, IEEE Eighth International Conference on Advanced Computing (ICoAC),978-1-5090-5888-4
- [4]. Wiebe, J.; Bruce, R.; and O'Hara, T. (1999), Development and use of a gold standard data set for subjectivity classifications. In Proc. 37th Annual Meeting of the Assoc. for Computational Linguistics (ACL-99), 246–253. University of Maryland: ACL.
- [5]. Priyavrat, A. J Singh (2017), Sentiment Analysis: A Comparative Study of Supervised Machine Learning Algorithms Using Rapid miner, International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887 Volume 5 Issue XI, November 2017.
- [6]. A. Mudinas, D. Zhang, M. Levene (2012), Combining Lexicon and Learning Based Approaches for

Concept-Level Sentiment Analysis, Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, New York, NY, USA, Article 5, pp. 1-8, 2012.

- [7]. L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B.Liu(2011), Combined Lexicon-Based and Learning-Based Methods for Twitter Sentiment Analysis, Technical report, HP Laboratories.
- [8]. RichaSharma ,Shweta Nigam and Rekha Jain (2014), Opinion mining of movie reviews at document level, International Journal on Information Theory (IJIT), Vol.3, No.3.
- [9]. RuiXiaa,d, Feng Xub, JianfeiYua, Yong Qia,b, Erik Cambriac (2016), Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis, Volume 52, Issue 1, January 2016, Pages 36-45
- [10]. Aurangzeb Khan, Baharum Baharudin and Khairullah Khan (2011). Sentiment Classification Using Sentence-level Lexical Based Semantic Orientation of Online Reviews. Trends in Applied Sciences Research, 6: 1141-1157.
- [11]. Gurshobit Singh Brar, Ankit Sharma (2018), Sentiment Analysis of Movie Review Using Supervised Machine Learning Techniques, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, pp. 12788-12791.
- [12]. M.Govindarajan (2013), Sentiment Analysis of Movie Reviews using Hybrid Method of Naive Bayes and Genetic Algorithm, International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-3 Number-4 Issue-13.
- [13]. Tirath Prasad Sahu and Sanjeev Ahuja (2016), Sentiment Analysis of Movie Reviews: A study on Feature Selection & Classification Algorithms, 2016 International Conference on Microelectronics, Computing and Communications (MicroCom)978-1-4673-6621-2.

IOSR Journal of Engineering (IOSRJEN) is UGC approved Journal with Sl. No. 3240, Journal no. 48995.

\_\_\_\_\_

Ravina Nimmi Rodrigues. "A Comprehensive Study on Methods, Challenges and Applications of Sentiment Analysis: A Survey." IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 08, 2019, pp. 12-16.