THINK INDIA JOURNAL
ISSN: 0971-1260
Vol-22-NO-30 Special Issue- December-2019

Sentiment Analysis Importance in Digital Era

Karan Kaur Phull[1] ,DishaKaur Phull[2]
[1]*Senior Software Engineer, Teradata India Pvt. Ltd., Hyderabad, Telangana, India*
[2]*School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India*

ABSTRACT

The currently living period is a digital era where social media and technology has enabled almost everyone to openly share their views, thoughts or opinions. These views or thoughts in the form of posts, comments, general product review or blogs can play a powerful role in efficacy of the product, business or society trends. Here, analyzing the sentiments of such textual data has become very critical. Various factors such all types human emotions, lexical resource, topic and opinion play important roles in identifying the sentiment of a sentence. This paper discusses the challenges faced while performing efficient sentiment analysis. Here the growth in the study of the sentiment analysis is discussed along with the different approaches used. It also shows light on how Big Data and Machine Learning has played role in analyzing the sentiments of the such data.

KEYWORDS - Sentiment Analysis, Opinion Mining, Semantic, Machine Learning.

1. INTRODUCTION

Due to the digital revolution and the boom of the social media loads of data have been at the disposal for analyzing, extracting, loading in different forms – structured, unstructured or semi-structured. This large amount of textual and non-textual data is referred to as the Big Data. Analysis of such data can be done in different forms by performing the Data Mining or Text Mining techniques. In this paper, the textual category of the data has been chosen. The data collected from the blogs, social networks, news reports and e-commerce websites can be used to analyze and understand the opinions, preferences or the sentiments of the public. Such analysis of the data can play major role in the marketing, customer satisfaction, political movement, movie sales and many more.

Nowadays customers rely more on the reviews by existing customers and the producers try to improve the quality of the product based on such reviews. Online media is providing updates and feedbacks on the global issues via news, and textual posts etc, which can be analyzed to know the society trends. Various blogs and forums are enabling the public to be aware of the various issues or public experiences.

Sentiment Analysis is a research area which basically analyzes the sentiments of the people towards various entities such as events, services, products etc. It can be achieved by

ISSN: 0971-1260
Vol-22-NO-30 Special Issue- December-2019

knowing computational linguistics, data mining and machine learning while dealing with the psychology of the person. Sentiment Analysis requires to find the polarity of the sentence. The sentiments can be positive, negative or neutral [1]. Such analysis can be made at three different levels –

- **Document Level:** At this level, the classification is done at the whole document opinion for a service or product.

- **Sentence Level:** For reviews or comments, the analysis is done on a sentence level to classify its opinion whether its positive, negative or neutral.

- **Aspect Level:** At this level, the classification of sentiments is done based on the feature, which impacts the opinion mining at feature level.

The polarity of a word can be found using any lexicon, but the words need not affect the overall sentiment of the sentence. In this paper, the challenges in performing the Sentiment Analysis have been explained such as the human emotion, internet slang words detection, lexical resources, classification techniques. The growth in this research field over the last two decades and the impact of the machine learning and big data techniques have also been discussed here. This can help the new researchers under this domain to have an overview of the concept and use them for their Sentiment Analysis appropriately.

2. CHALLENGES IN SENTIMENT ANALYSIS

In this section the different challenges affecting efficient Sentiment Analysis have been discussed. Apart from the basic word polarity, there are few critical challenges which affect the sentiment of a sentence such as–

Attitude/ Emotion: Almost all emotions affect the sentiments. Analyzing the human emotions such as anger, sadness, happy, jealousy, sarcasm of the person plays a very important role in finding the overall sentiment[2]. For example- “We are living in such a beautiful world” can be treated as a positive sentence but if the person was angry or sarcastic in the context then this same sentence could change the sentiment of the text as a negative. The following Figure1 displays the structure of emotions also known as the psychological model.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue- December-2019

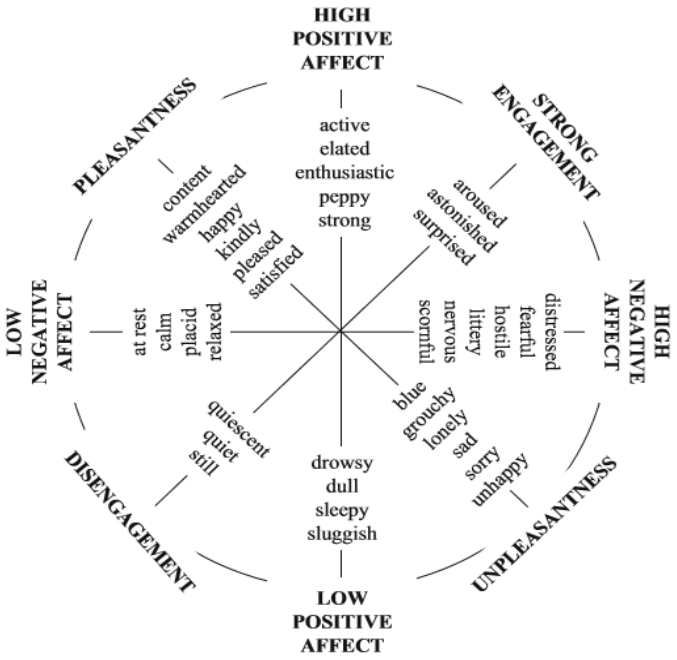

Fig. 1 Watson and Tellegen's Structure of Emotions in two dimensions.

- **Opinion Holder or Time of Opinion:** The sentiment of the text also depends on the opinion holder and the time of the opinion[2]. For example – "Alice hated the flavor of tomatoes." Here the hatred and negativity towards tomatoes is not expressed by author but is that of Alice. Similarly, "Alice hated the flavor of tomatoes during her middle school." Here hatred of Alice is limited only during the period of middle school, that doesn't mean she hates it in current period. If we consider the time and the holder then this statement is not a negative sentiment.
-
- **Aspect/Feature:** The object, feature or service for which the analysis is being conducted also affects the sentiment of the statement[3]. For example – "This APP is very useful and nice just that it's customer service is a little tardy." here the review is not negative about the overall services of the APP but only the customer service feature of the APP is negative.
-
- **Detecting and Handling Internet Slang or Acronyms:** Sentiment analysis majorly deals with the text analysis where we try to identify the sentiment-oriented words and contextual words from the text. But with the advent of the social media and chat functionality when people try to express their views, it's not necessary to follow the proper English Grammar. Hence the usage of abbreviations is prominent which are also know as the internet slangs. Few of the examples are - "life" as "lyf", "brb" for "be right back", "tc" for "take care", "lol" for "laugh out loud" etc. These need to be identified and handled as preprocess before the sentiments could be determined for the statements[4].
-

All these factors can pose as a challenge that needs to be overcome to give provide a proper sentiment for any sort of data.

ISSN: 0971-1260
Vol-22-NO-30 Special Issue- December-2019

3. APPROACHES

After addressing the challenges mentioned in the previous section, here the approaches available to classify the sentiments into different levels of polarity have been discussed. Big data plays role when the data for the sentiment detection is very large. In such cases the use of Hadoop has been noted by the researchers. Machine learning plays a significant role in the sentiment classification. This classification can be via supervised such as classification techniques or unsupervised techniques such as clustering which is a semantic approach. A survey even summarized the use of a combination of unsupervised and semi-supervised learning known as hybrid approach for sentiment classification [5]. Apart from these a high usage of the lexical resources such as SentiWordNet, Sentiment Tree Bank etc have been observed.

3.1 Classification based Analysis

Classification methods such as Support Vector Machine, Maximum Entropy Classifier and Rule Based classifiers are some of the techniques which can be used to identify the sentiment. SVM is used for analysis where the data needs classification or regression. Researchers have identified that feature set training done on SVM resulted in improved classification of sentiment whether its binary or multi polarity classification by taking proper information-oriented feature set. Similarly, Maximum entropy classification has been used to detect the parallel sentences between language pairs [6]. Decision rule-based classifiers allow the overlaps in the decision space based on set of rules [7]. The classification can be done for the emotions and the sentiments.

3.2 Clustering based Analysis

Researchers have proposed models using various clustering algorithms such as hierarchical clustering, density-based clustering for sentiment analysis. A dendrogram is created by clustering groups of data on variety of scales, clusters at one level join as clusters on the next level, which is called as hierarchical clustering. In density-based clustering, grouping the points which are closely packed together leading to outliers in the low-density regions [5].

3.3 Lexical Resource based Analysis

Lexical resources such as the WordNets, SentiWordNet and Sentiment Tree have become very popular as they deal with the semantic of the sentence rather than the contextual orientation of the opinion. It checks how much the word is inclined towards positive or negative sentiment. This does not require any prior training of the data which saves time. First the opinionated words need to be extracted and then lookup for the scores in the wordNets. The success of this majorly depends on the underlying Part Of Speech (POS) tagger which helps in extracting the opinionated words. The semantic approach directly provides the sentiment of the words.

3.4 Statistical Approach based Analysis

ISSN: 0971-1260
Vol-22-NO-30 Special Issue- December-2019

The patterns of co-occurrence and seed of opinion words can be found using statistical techniques. Point-Wise Mutual Inclusion is seen to be used in sentiment analysis to find the polarity by detecting the frequency of the words which occur with other words in positive or negative context [5]. Latent Semantic Analysis is also a statistical approach which produces a set of meaningful patterns related to the document and terms by analyzing the relationship between the documents and the terms. It has been used to examine the impact of the features on review texts by finding the semantic characteristics.

4. CONCLUSION

Sentiment Analysis is an important means for products, services and people to know the opinion of people all over the world. Social media, online news, e-commerce, blogs have abundant data waiting to be analyzed and highlight the sentiments of the people. In this era technology and innovation has the capability to even motivate and impact people's lives in a better way. In this paper, no new approach has been defined but rather summarizes the world of Sentiment Analysis in this modern digital era. This paper aims to help the new researchers and developers get an overview of the challenges and the approaches to take a plunge into the world of the sentiment analysis and dig deeper to innovate. It also shows how machine learning algorithms helps in the process of analyzing the sentiments.

REFERENCES

[1] Patil, Priyanka, and PratibhaYalagi. "Sentiment Analysis Levels and Techniques: A Survey." space 1 (2016): 6.
[2] Yue, Lin, Weitong Chen, Xue Li, WanliZuo, and Minghao Yin. "A survey of sentiment analysis in social media." Knowledge and Information Systems (2018): 1-47.
[3] Lazarus, Nancy. "5 Key Challenges in Sentiment Analysis." Adweek. Adweek, July 21, 2015. https://www.adweek.com/digital/5-key-challenges-in-sentiment-analysis/.
[4] Kundi, FazalMasud, Shakeel Ahmad, Aurangzeb Khan, and Muhammad ZubairAsghar. "Detection and scoring of internet slangs for sentiment analysis using SentiWordNet." Life Science Journal 11, no. 9 (2014): 66-72.
[5] Ravi, Kumar, and Vadlamani Ravi. "A survey on opinion mining and sentiment analysis: tasks, approaches and applications." Knowledge-Based Systems 89 (2015): 14-46.
[6] Medhat, Walaa, Ahmed Hassan, and HodaKorashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams engineering journal 5, no. 4 (2014): 1093-1113.
[7] Vasuki, M., J. Arthi, and K. Kayalvizhi. "Decision making using sentiment analysis from twitter." International Journal of Innovative Research in Computer and Communication Engineering 2, no. 12 (2014).