

Literature Review on Sentiment Analysis in Social Media

Ved Panpaliya - 20200802147

Aadarsh Krishna - 20200802127

Shruti shete – 20210812060

Somesh Roy - 20200802097

ABSTRACT

In today's world, sentiment analysis stands as the prominent research topic in demand under the Natural Language Processing (NLP). The fundamental objective of this research topic is to spot out the emotions and opinions of the customers or users via a text basis. Even though numerous research works have been carried out in this field through diverse models, sentiment analysis is still considered a challenging problem with so many conflicts to be solved. Some of the existing challenges are due to the slang words, new accents, grammatical and spelling mistakes etc. This paper plans to make a literature review using different machine learning algorithms with various data. The current literature review aims to survey nearly 20 contributions, which covers different types of applications being used for sentimental analysis. At first, the analysis focuses on illustrating the contributions of each work and observes the type of machine learning algorithms used. Moreover, the analysis also concentrates on the identification of the type of data used. Further, the utilized environment and the performance measures covered in each work is evaluated, and concluded with proper research gaps and challenges, which helps to identify the non-saturated application for which the sentimental analysis is needed most in upcoming research.

INTRODUCTION

There is a vigorous improvement in the micro blogging websites as well as social networks. One of the major web destinations to the users is micro blogging websites, which are helpful for expressing the user's attitudes, opinions, and thoughts regarding various contexts. The most used social networking services and the micro blogging platform is twitter, which provides more data. At present, for the sentiment analysis of the user's opinions on the product, event, or context, researchers make use of social data. Moreover, the other name for sentiment analysis is opinion mining, which is the significant NLP task. This sentiment analysis defines orientation of sentiments related to text as either neutral, positive, or negative. Moreover, sentiment analysis represents the text analytics, computational linguistics, and NLP implementations for recognizing and categorizing the opinions of the user. In general, the main intention of the sentiment analysis is to define the author's point of view concerning the similar context or the entire document's contextual polarity. The view can be either a user's judgment or assessment, affective state or the deliberated communication of emotion.

In general, the classification of text expressions in source materials into facts and opinions is done by the sentiment analysis. Facts are the objective expressions regarding the events and their attributes as well as entities. The opinions are the subjective expressions of

sentiments, emotions, feelings, events and attributes, and attitudes. This must be specified that not all the objective sentences include no opinions and not all subjective sentences include opinions. Thus, for sentiment analysis, it is significant for recognizing and extracting the facts and opinions from source materials. However, this seems to be quite complex for attaining precisely. In recent times, important approaches are related to machine learning, rule-based, and the combination of both techniques. Machine learning models consist of conventional approaches like deep learning and conditional random field approaches, whereas the rule-based models consist of lexicon-based approach. Object detection, network optimization, image recognition, system security, sensor networks and transportation are based on deep learning methods, which are mostly utilized in different fields. Several researchers have combined deep learning as well as machine learning algorithms into text sentiment analysis by sentiment lexicon formulation and best results are obtained.

Literature Review

Related Works,

In 2019, Saad and Yang have aimed for giving a complete tweet sentiment analysis on the basis of ordinal regression with machine learning algorithms. The suggested model included pre-processing tweets as first step and with the feature extraction model, an effective feature was generated. The methods such as SVR, RF, Multinomial logistic regression (SoftMax), and DTs were employed for classifying the sentiment analysis. Moreover, twitter dataset was used for experimenting the suggested model. The test results have shown that the suggested model has attained the best accuracy, and also DTs were performed well when compared over other methods. In 2018, Fang et al. have suggested multi-strategy sentiment analysis models using the semantic fuzziness for resolving the issues. The outcomes have demonstrated that the proposed model has attained high efficiency.

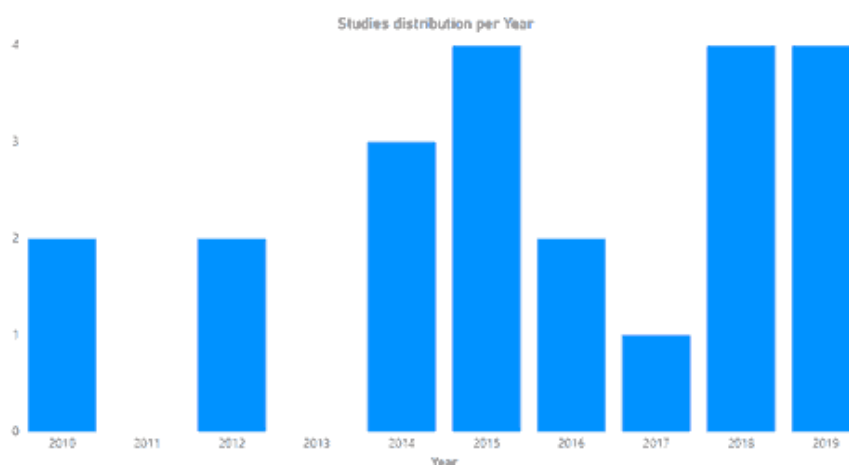
In 2019, Afzaal et al. have recommended a novel approach of aspect-based sentiment classification, which recognized the features in a precise manner and attained the best classification accuracy. Moreover, the scheme was developed as a mobile application, which assisted the tourists in identifying the best hotel in the town, and the proposed model was analyzed using the real-world data sets. The results have shown that the presented model was effective in both recognition as well as classification.

In 2019, Feizollah et al. have concentrated on tweets related to two halal products such as halal cosmetics and halal tourism. By utilizing Twitter search function, Twitter information was extracted, and a new model was employed for data filtering. Later, with the help of deep learning models, a test was performed for computing and evaluating the tweets. Moreover, for enhancing the accuracy and building prediction methods, RNN, CNN, and LSTM were employed. From the outcomes, it was seemed that the combination of LSTM and CNN attained the best accuracy.

In 2018, Mukhtar et al. have performed the sentiment analysis to the Urdu blogs attained from several domain with Supervised Machine learning and Lexicon-based models. In Lexicon-based models, a well-performing Urdu sentiment analyzer and an Urdu Sentiment Lexicons were employed, whereas, in Supervised Machine learning algorithm, DT, KNN, and SVM were employed. The data were combined from the two soruces for performing the best sentiment analysis. Based on the tests conducted, the outcomes were shown that the Lexicon-based model was superior to the supervised machine learning algorithm.

In 2020, Kumar et al. have presented a hybrid deep learning approach named ConVNetSVMBoVW that dealt with the real-time data for predicting the fine-grained sentiment. In order to measure the hybrid polarity, an aggregation model was developed. Moreover, SVM was used for training the BoVW to forecast the sentiment of visual content. Finally, it was concluded that the suggested ConvNet-SVMBoVW was outperformed by the conventional models.

In 2018, Abdi et al. have proffered a machine learning technique for summarizing the opinions of the users mentioned in reviews. The suggested method merged multiple kinds of features into a unique feature set for modelling accurate classification model. Therefore, a performance investigation was done for four best feature selection models for attaining the best performance and seven classifiers for choosing the relevant feature set and recognized an effective machine learning algorithm. The suggested method was implemented in various datasets. The outcomes have demonstrated that the combination of IG as the feature selection approach and SVM-based classification approach enhanced the performance.



Studies Distribution per Year

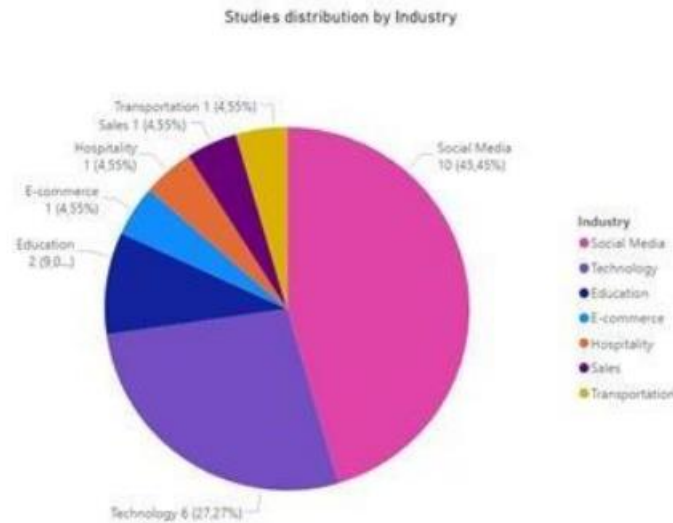


FIGURE 2
STUDIES DISTRIBUTION BY INDUSTRY

RESEARCH GAPS AND CHALLENGES

The machine learning algorithm uses linguistic features by an objective of system's performance optimization with the example data. The big data models like Pentaho and Mahout consists of plugins and library related to the machine learning algorithm that is evaluated for performing the classification of sentiment. In evaluation of big data, the user must define the method type, which should be given to the data and that method has been performed using big data analytics tools in order to solve the certain problem like predictive analytics. In most of the cases, two document sets are needed for performing the machine learning-based categorization. These sets are considered as the training as well as testing sets. In order to learn the features of the document, the training set has been given to the machine learning algorithm, and for evaluating the performance of the classifier, testing set is employed.

In sentiment analysis, many machine learning algorithms have been employed for classification. The famous machine learning algorithms, which have attained more successful in classifying the text are NB, ME, and SVM. The remaining machine learning algorithms in NLP are ID3, centroid classifier, N-gram model, K-Nearest Neighbour, C5, and winnow classifier. Many of the conventional models are associated with the public-related sentiments from social network, and text applications. However, there is less amount of work, in which the ontology and semantics are seemed to be more important research works in sentiment analysis. Now-a-days, the conventional research models are experimented using the public review datasets. However, this kind of review has not been evaluated keenly by concerning the sentiments. By categorizing the sentiments as positive or negative, it will not provide the original and the concealed information beyond the actual concepts of sentiments. In addition, there are some specific sentences that are quite

complicating and accurate classification cannot be performed. There have been many constraints such as assessment of sentiment during the review, and document exploration using many subjects. Moreover, traditional models concentrated on major problems rather than minor ones, in which the accuracy is not seemed to be optimized. It is observed that there are few research methodologies, which have been recognized as standard models. There is only little number of researches other than standard models, whose results are seemed to be less efficiency over a suitable approach. The evaluation of less dimension text might utilize less number of resources. However, the collection of sentiments from the collaborative environment will utilize more amounts of resources. It is unfortunate that in the conventional researches, the authors didn't found several confirmations related to the computational expenses of efficient techniques for performing huge data sentiment analysis.

Conclusion : The present paper has developed the review of earlier contributions with various machine learning models using discrete information. The present review has explored 6 research works that covered various implementations employed for sentiment analysis. Initially, the assessment has concentrated on clarifying the contribution of every task and observed the type of machine learning algorithm utilized. The evaluation also focused in recognizing the type of data employed. Later, the environment utilized and the performance metrics covered in each contribution was analysed. Finally, the research gaps and challenges were mentioned that were useful for recognizing the non-saturated implementation for which the sentiment analysis was required in further researches.