Chapter 14
A Systematic View of Sentiment Analysis on Different Techniques, Challenges, and Future Directions in COVID-19

A. Sathya
Bishop Heber College, Bharathidasan University, India

M. S. Mythili
Bishop Heber College, Bharathidasan University, India

ABSTRACT

In the last few years, social networking sites have exploded at a rapid pace. Sentiment analysis (SA) is a way of mining data and reading text using natural language processing(NLP) In today's era, corona virus is a hot issue and pandemic all over the world. Millions of citizens use blogs or microblogging tools to communicate their viewpoints. Twitter is one of the handiest platforms and most popular social media for users to share their thoughts and ideas. Through blogging sites, the public can share their feelings and opinions such as "panic, anger, misery," and it can be divided the text into three categories: positive, negative, and neutral. Sentiment analysis is employed in a variety of industries, including healthcare, finance, sports, politics, hospitality, and tourism. According to a comparison of ten primary studies, Naive Bayes (NB) and support vector machine (SVM) are widely used algorithms for sentiment analysis in COVID-19.

1. INTRODUCTION

Nowadays, people like to write reviews about things they use or about anything that's trending on the internet (Zhou, Z. et al., 2019). Online media consists of a huge number of posts, so knowing one's perspective can be quite a challenge. Sentiment analysis is a method that uses text analytics to analyse and classify emotions using Twitter information through an online review process (Nirmala et al., 2023).

DOI: 10.4018/979-8-3693-1301-5.ch014

The purpose is to understand how the public views fear, depression, and disgust, which are directly related to the COVID-19 pandemic in India (Cirillo et al., 2023; Wikipedia, 2020). A text can convey multiple polarities of sentiment at the same time. Sentiment analysis is a natural language processing technique that divides the text into positive, negative, and neutral categories (Pandit, 2023). Opinions can be based on subjective, objective, or both types of data (Priscila et al., 2023). The subjective form includes a positive or negative view, while the objective form contains facts (Abdul Mohaimin Rahat & Kahir, 2019). Using opinion mining, the subjectivity and objectivity of knowledge are created (Saxena & Chaudhary, 2023). The outcome may be either positive or negative, or a combination of both (Vinoth Kumar et al., 2022). After the world war, "tsunami and COVID-19" were the biggest problems faced by the people. In today's world, people rely a lot on social networking sites like Instagram, Facebook, and Twitter (Suganthi & Sathiaseelan, 2023). It is predicted that posts shared on social media will provide people with accurate information (Jeba et al., 2023). But in most cases, the information led to incorrect decisions, like when COVID-19 information circulated over the mass media (Imran et al., 2020). In addition, most online information is unstructured. So proper pre-processing stages should be used before applying sentiment analysis (Pradha et al., 2019). To overcome the complication of many research challenges, a set of feature vectors can be taken by applying effective feature extraction techniques like "TF-IDF (Term Frequency-Inverse Document Frequency), gramme extraction, and Bag of Words (BOW)" (Deepak Kumar Jain et al., 2022). Different sorts of machine learning algorithms like "Nave Bayes" and "Support Vector Machine" should be applied to categorise the polarity of the problem statement (Sathya & Mythili, 2022). This survey explores the levels, approaches, and steps of sentiment analysis that can be used to categorise the sentiments of people about COVID-19 (Vashishtha & Dhawan, 2023).

2. RELATED WORKS

Gulati et al. (2022) presented sentiment analysis classifications and methodologies. The author emphasised sentiment analysis's useful applications. Lexicon techniques and machine learning were described in detail. Mrityunjay Singh et al. (2021) used pre-processing steps to separate the essential data from the noise and retrieve the information that is important. The CORONA virus categorization of public opinion has been determined using the BERT model.

It has been obtained from the GitHub source; validation and emotion categorization are also used. Pradha et al. (2019) discussed effective text pre-processing, including "stemming, lemmatization, and spelling correction." To handle Twitter data, "Support Vector Machine," "Deep Learning," and "Naive Bayes Classifier" are used. The accuracy of SVM's model performance was higher. Naseem et al. (2021) described the Twitter API to retrieve the data from the Twitter dataset. To eliminate certain unrelated COVID datasets, pre-processing will be performed. TF-IDF and word embedding vectorization algorithms are used. Machine learning algorithms like Naive Bayes, Decision Trees, and Support Vector Machines are used for classification in order to categorise the polarity.

Deep learning classifiers such as "LSTM" and "CNN" can be used to demonstrate greater accuracy. Jeganathan et al. (2023) described the classification methods, illustrated the accuracy of SVM and Naive Bayes measures, and labelled how to extract and recognise sentiment from the text. Sathya & Mythili (2022) compared the study of machine learning algorithms like Support Vector Machine, Nave Bayes, and finally exposed SVM, which produces results with higher accuracy.