International Journal of Natural Computing Research
Volume 10 • Issue 2 • April-June 2021

Development of a Predictive Model for Textual Data Using Support Vector Machine Based on Diverse Kernel Functions Upon Sentiment Score Analysis

Sheik Abdullah A., Thiagarajar College of Engineering, India
Akash K., Thiagarajar College of Engineering, India
Bhubesh K. R. A., Thiagarajar College of Engineering, India
Selvakumar S., GKM College of Engineering and Technology, India

ABSTRACT

This research work specifically focusses on the development of a predictive model for movie review data using support vector machine (SVM) classifier with its improvisations using different kernel functions upon sentiment score estimation. The predictive model development proceeds with user level data input with the data processing with the data stream for analysis. Then formal calculation of TF-IDF evaluation has been made upon data clustering using simple k-means algorithm. Once the labeled data has been sorted out, then the SVM with kernel functions corresponding to linear, sigmoid, rbf, and polynomial have been applied over the clustered data with specific parameter setting for each type of library functions. Performance of each of the kernels has been measured using precision, recall, and F-score values for each of the specified kernel, and from the analysis, it has been found that sentiment analysis using SVM linear kernel with sentiment score analysis has been found to provide an improved accuracy of about 91.18%.

KEYWORDS

Document Classification, Inverse Document Frequency, Movie Reviews, Predictive Modeling, Recommendation System, Support Vector Machine, Term Frequency, Text Analysis

1. INTRODUCTION

The impact of growth of data across various domains has been increasing with respect to volume and complexity. Leading organizations across various domains are using analytics to uncover insights from big data to achieve a significant outcome. With the advent of smart phones, cloud, and Internet of Things, data analytics has extended to Big Data. By providing the ability to predict trends before they happen, analytics can enable to stay one step ahead of the competition into the future. Data Analytics is needed to uncover hidden patterns and visualizing it to real-time has a good social and financial impact.

DOI: 10.4018/IJNCR.2021040101

Data Science is an emerging trend and Data Analysts or Data Engineers are the required market positions in the industrial world today. Data Analytics will support Teaching Learning effectively and promote Research activities in the areas like Predictive Analytics, Survival Analytics, Text analytics, Social Media Analytics, etc.

Among the analytical techniques used in practice the most significant type of analytics is the text analytics which is predominantly needed for analyzing various sorts of textual data. Text extraction provides the unique pathway for the opinions or the emotions extraction from the textual data which can be further deployed in making decision / decision analysis. In recent days, Sentiment Analysis is used in social and health care applications for collecting and reviewing customer responses.

Sentiment analysis basic task is to classify the given textual data to be as positive or negative. The advanced text analytic process is used to classify emotions of a person as happy, sad or angry. In simple terms, Sentiment Analysis predicts the psychological behavior of a person. In Social media, Sentiment analysis is used to find out reviews corresponding to products/movies/recommender systems. The aspect of sentiment analysis deals with the process of relative mining which in turn determines the specific information from the available data and it helps the vendor in identifying their product views from customer point of view, so the reviews of social media is restricted to count. With invent of deep learning algorithms text analysis is improved far better now. Artificial Intelligence techniques act as a tool for doing sentiment analysis in depth. In recent days Sentiment Analysis is used in Face book and Twitter data (Araque et al., 2019).

Social Media are the main source for Sentiment Analysis. Sentiment Analysis is used mainly to classify text, based on the dataset Sentiment Analysis classify text as binary (positive, negative or neutral) or multiclass. Preprocessing is the initial step in Sentiment Analysis, several techniques are used for preprocessing, some of them are remove numbers, remove punctuation, remove stop words, stemming, etc.

Sentiment analysis process is segregated into two types such as machine learning based and lexicon-based analysis. The mechanism behind the machine learning-based approach involves algorithms relative to supervised and un-supervised learning schemes. The scheme of supervised learning involves class labeled training data tuples for the given available dataset. In un-supervised scheme the training data tuples won't be provided with class labeled tuples, the tuples once trained then will be assigned for classification/prediction process. The mechanism of lexicon-based approach involves the process of determining the sentiments (positive/negative) from the given semantic context of the observed word/phrase from the given text (Mohey et al., 2018). This research work specifically focus on the design and development of a predictive model for the textual data which is specifically based on the movie review collections using SVM which is explicitly based on linear kernel function and the validation has been made using statistical incorporation. The model facilitates the analysis of text data with the recommendations from various users and ascertains the sentimental results in a real-time perspective (Boudad et al., 2018).

2. LITERATURE REVIEW

The authors (Ankit & NabizathSaleena, 2018) deployed the process of sentiment analysis for data corresponding to twitter (social media) with the analysis from the tweets collected from the users. The process is then segregated into positive/negative/neutral. The classification process has been carried out using naïve bayes, random forest algorithm and logistic regression analysis. The authors proposed new classifier approach called ensemble classifier which combines the base classifier into a single classifier, with the intention to improve the accuracy and the performance estimation of the sentiment analysis process. The implementation is carried out using python programming (data in the form of multi-dimensional array). For data pre-processing Scikit-learn and Natural Language Toolkit are used. Recall, Precision and F-Call are used as performance metrics. Ensemble classifier predicts higher accuracy when compared to traditional classification algorithms (Anandarajan et al., 2018).

PayCrypto Analtcoin Minting Application as Interest to Cryptocurrencies

G. K. Sandhia, R. Vidhya, K. R. Jansi, R. Jeya, M. Gayathri, S. Girirajan, S. Nagadevi, N. Ghuntupalli Manoj Kumarand J. Ramaprabha (2024). *Bio-Inspired Optimization Techniques in Blockchain Systems (pp. 207-220).*
www.irma-international.org/chapter/paycrypto-analtcoin-minting-application-as-interest-to-cryptocurrencies/338092