A Hybridized GA-Based Feature Selection for Text Sentiment Analysis

Gyananjaya Tripathy
National Institute of Technology, Raipur, India

Aakanksha Sharaff
National Institute of Technology, Raipur, India

INTRODUCTION

The advancement of today's internet technology has changed the lifestyle of society. Due to this advancement, the current generation has upgraded their lifestyle up to a certain extent. Different social forums are commonly used to share helpful information and new ideas for advertisement and service improvement. The social platform is often watched with various perspectives. These include compiling business marketing strategies for product and promotional services, observing harmful actions to detect and reduce cyber-attacks, and sentiment analysis to analyze human responses and feedback (Saberi & Saad, 2017). Sentiment analysis is often referred to as archaeology, uprooting and classifying sentiments from text using Natural Language Processing (NLP), mathematics, or Machine Learning (ML) methods. ML methods use various approaches and a database that can be trained to distinguish and find sentiments (Fiok et al., 2021). Authors have widely studied the field of sentiment analysis over the past few years. In this state of affairs, different methods have been tested after development. The most usual process is ML which requires a robust database to train and learn the relationship between various aspects and sentiments.

Sentiment analysis is a form of written assessment or language spoken to determine whether speaking is negative, positive, or neutral and to what extent. Current analysis Market tools can handle a lot of price customer criticism honestly and accurately. Collectively, sentiment analysis finds customers' ideas on various topics, including procurement, the provision of services, or the presentation of promotions (Alsaeedi & Khan, 2019). Sentiment analysis is often used in the case of a review. Reviews can be taken from various resources for various reasons, such as product reviews, political reviews, and community reviews. When feedback from customers using any product, further questions will be included: Is the product usable? Is this product satisfactory? Is this product worth the money? Some helpful information always comes out of updates in positive or negative feedback (Birjali et al., 2021). Sentiments need to be learned using these practical answers. The semantic position estimates submission and ideas in the text data. The rules-based analysis searches for different words in a text and categorizes them based on positivity and negativity.

The proposed paper is based on Amazon's review dataset's hybrid sentiment analysis process. The dataset contains several responses and equally separates the positive and negative labels. Authors have developed an integrated novel algorithm based on the Genetic Algorithm (GA) to minimize the feature (Iqbal et al., 2019). Iqbal et al. (2019) have explained the feature selection method using GA by evaluating the fitness value with sentiment score whereas in the proposed model the fitness of each solution is evaluated using the accuracy score of each feature subsets. Support Vector Machine (SVM) (Preeti et al.,

DOI: 10.4018/978-1-7998-9220-5.ch112

Copyright © 2023, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

2020) is used to check the validity of the words concerning the label to find an effective solution. This evolutionary process of selecting the right element improves accuracy with increasing scalability. This customized method offers a 45% reduced feature set with better accuracy. In addition to demonstrating the feasibility of this proposed method, the authors conducted a detailed study with other mitigation strategies such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD). Using these two algorithms as a comparison, the authors obtained the proposed model results, which provides up to 14.5% increased accuracy over PCA and 16.2% increased accuracy over SVD through the Naïve Bayes learning process and this reduction feature strategies. As a comparison of the number of features of all three feature reduction strategies, the proposed method gives 13% better results compared to PCA and a 10% better result compared to SVD. With a small amount of variable set, the proposed system exceeds the other two algorithms.
The main contributions to the proposed work are as follows:

- Text reviews are undergone pre-processing before being applied to the proposed framework. This pre-processing includes 'stop words removal', 'lemmatization,' and 'tokenization.'
- The authors have designed, developed, and evaluated a hybridized sentiment analysis model by integrating dictionary-based methods and ML methods to perform better than each individual.
- The authors have proposed a hybrid feature minimization method using a GA-based technique with a customized fitness calculation. Fitness calculation uses SVM to explore feasible solutions that improve system performance.
- The authors have analyzed the proposed method that demonstrates improved accuracy compared to the existing feature reduction algorithms.
- The authors have examined the novel algorithm to decrease the number of features to an excellent yield with an increment in accuracy esteem.

In evaluating the results of the suggested model, execution is performed with the Amazon cell review dataset (discussed in section 5) using Python language. The evaluation process is dependent on several measuring parameters like precision, recall, F1_Score, and accuracy. By applying all these performance measuring parameters, the proposed hybrid model outperforms the other two techniques. The further portion is divided in different sections where section 2 represents the work done in the related field. Section 3 represent the architecture behind the proposed model. Section 4 explains the proposed algorithm. Section 5 gives the insight of result analysis. Section 6 shows the future research direction and the final Section concludes the proposed work.

BACKGROUND

This section focused on outstanding correlated research performed in sentimental and textual analysis mines. The proposed comparison method is based on two factors discussed earlier. The important thing is to check how the user's feedback and social practices can help analyze the current circumstances.

Medhat et al. (2014) have proposed a novel technique in sentiment analysis. Fields related to transfer reading, sentimental acquisition, and construction resources were also their discussion point. The authors have attempted to provide a complete picture of the sentiment analysis approach. Khan et al. (2011) have proposed a method that detached different sentiments like objective or subjective directly from the review and blog feedback. SentiWordNet helped to evaluate points and determine the polarity. The proposed

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/a-hybridized-ga-based-feature-selection-for-text-sentiment-analysis/317591

Related Content

Network Support Data Analysis for Fault Identification Using Machine Learning
Shakila Basheer, Usha Devi Gandhi, Priyan M.K.and Parthasarathy P. (2022). *Research Anthology on Machine Learning Techniques, Methods, and Applications (pp. 586-595).*
www.irma-international.org/chapter/network-support-data-analysis-for-fault-identification-using-machine-learning/307473

Intelligent System for Credit Risk Management in Financial Institutions
Philip Sarfo-Manu, Gifty Siawand Peter Appiahene (2019). *International Journal of Artificial Intelligence and Machine Learning (pp. 57-67).*
www.irma-international.org/article/intelligent-system-for-credit-risk-management-in-financial-institutions/238128

Programming Language Support for Implementing Machine Learning Algorithms
Anitha Elavarasi S.and Jayanthi J. (2020). *Handbook of Research on Applications and Implementations of Machine Learning Techniques (pp. 402-421).*
www.irma-international.org/chapter/programming-language-support-for-implementing-machine-learning-algorithms/234135

Power Consumption Prediction of IoT Application Protocols Based on Linear Regression
Sidna Jeddou, Amine Baina, Najid Abdallahand Hassan El Alami (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 1-16).*
www.irma-international.org/article/power-consumption-prediction-of-iot-application-protocols-based-on-linear-regression/287585

Palmprint And Dorsal Hand Vein Multi-Modal Biometric Fusion Using Deep Learning
Norah Abdullah Al-johaniand Lamiaa A. Elrefaei (2020). *International Journal of Artificial Intelligence and Machine Learning (pp. 18-42).*
www.irma-international.org/article/palmprint-and-dorsal-hand-vein-multi-modal-biometric-fusion-using-deep-learning/257270