IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204159 |

A Multi-Stage Machine Learning and Fuzzy Approach to Cyber-Hate Detection

S.Chandrasekar, Manikandan K, Mahisurya A, Manivas P

Assistant Professor, Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

Department of CSE, Muthayammal Engineering College (Autonomous), Rasipuram, Tamil Nadu, India

ABSTRACT: Sentiment analysis is a very popular application area of text mining and machine learning. The popular methods include Support Vector Machine, Naive Bayes, Decision Trees and Deep Neural Networks. However, these methods generally belong to discriminative learning, which aims to distinguish one class from others with a clear-cut outcome, under the presence of ground truth. In the context of text classification, instances are naturally fuzzy (can be multilabeled in some application areas) and thus are not considered clear-cut, especially given the fact that labels assigned to sentiment in text represent an agreed level of subjective opinion for multiple human annotators rather than indisputable ground truth. This has motivated researchers to develop fuzzy methods, which typically train classifiers through generative learning, i.e. a fuzzy classifier is used to measure the degree to which an instance belongs to each class. Traditional fuzzy methods typically involve generation of a single fuzzy classifier and employ a fixed rule of defuzzification outputting the class with the maximum membership degree. The use of a single fuzzy classifier with the above fixed rule of defuzzification is likely to get the classifier encountering the text ambiguity situation on sentiment data, i.e. an instance may obtain equal membership degrees to both the positive and negative classes. In this paper, we focus on cyberhate classification, since the spread of hate speech via social media can have disruptive impacts on social cohesion and lead to regional and community tensions. Automatic detection of cyberhate has thus become a priority research area. In particular, we propose a modified fuzzy approach with two stage training for dealing with text ambiguity and classifying four types of hate speech, namely: religion, race, disability and sexual orientation - and compare its performance with those popular methods as well as some existing fuzzy approaches, while the features are prepared through the Bag-of-Words and Word Embedding feature extraction methods alongside the correlation based feature subset selection method. The experimental results show that the proposed fuzzy method outperforms the other methods in most cases.

I.INTRODUCTION

Sentiment analysis is aimed at identifying the attitude or mood of people through natural language processing, text analysis and computational linguistics. In recent years, machine learning has become a very powerful tool for classifying sentiments. In particular, Support Vector Machines (SVM), Naive Bayes (NB), Decision Trees (DT) and its ensemble methods such as Gradient Boosted Trees (GBT) have been used extensively with good performance in broad application areas that involve sentiment analysis, such as cyberbullying detection [1], [2], abusive language detection [3], [4], movie reviews [5], [6] and cyberhate identification [7], [8].

In recent years, deep neural networks (DNN) have also been used for sentiment analysis and other types of text classification. In the context of machine learning, the above algorithms(SVM, NB, DT, GBT and DNN) are all considered to belong to discriminative learning, since they all aim to distinguish between one class and other classes. In fact, the above algorithms work based on the assumptions that different classes are mutually exclusive and each instance is clear-cut and provided with a ground truth label. However, in the context of text classification, the above assumptions do not always hold,especially when considering the following examples:

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204159 |

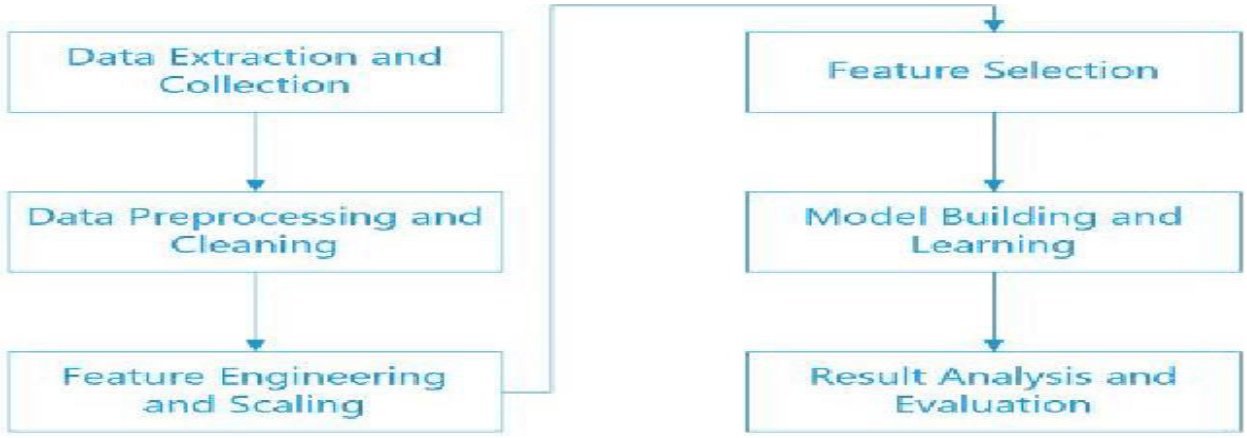

Fig 1: Detecting Hate Speech

In terms of the first assumption, for example, the same movie may belong to different categories, or the same book may belong to different subjects [9], [10], [6]. This example indicates that different classes may not necessarily be mutually exclusive, i.e. different classes could have overlaps, in terms of instances covered by these classes, and the instances can even be multilabelled in real applications. On the other hand, while different classes are truly mutually exclusive, instances could be very complex and are thus difficult to be classified uniquely to only one category. For example, text such as ”I LOVE my country but I HATE immigrants” involves both positive and negative speech [6]. This example indicates that an instance may not be clear-cut, i.e. an instance may partially belong to one class and partially belong to another class. Humans may agree this is hateful but for a discriminative algorithm this poses a challenge.

Furthermore, in sentiment analysis, the label assigned to each instance does not actually represent the ground truth but an agreed representation of the opinion of multiple human annotators, which means that different people may have different opinions about the polarity of a sentiment instance. Thus, sentiment analysis is essentially a task of opinion mining rather than discovery of externally verifiable patterns. The above examples indicate that textual instances are n aturally fuzzy and discriminative learning methods are likely to struggle to compute such fuzziness. This has motivated researchers to develop fuzzy methods for text classification, which are able to deal with fuzziness, imprecision and uncertainty of text. In this paper, we focus on detection of online hate speech (cyberhate) in short informal text posted to social media platforms. This has become a priority research topic due to the concern that the spread of online hate speech could lead to antisocial outcomes [7]. In particular, we deal with four types of online hate speech, namely: religion, race, disability and sexual orientation, by proposing a novel fuzzy approach grounded in generative learning, especially for dealing with text ambiguity, which could result from the following cases: a) the same word may be used in different contexts leading to different semantic meanings; b) that similar instances are assigned different labels by different annotators due to their different opinions. The proposed fuzzy approach is different from existing fuzzy systems in two aspects:

II.RELATED WORK

Shaukat (2020) The primary objective of this paper is to bridge the gap between machine learning (ML) techniques and threats to computer networks and mobile communication by providing a comprehensive survey of the crossovers between the two areas. The second one is the rapid growth of interest in machine learning and cybersecurity in both academic and industry. The machine learning models are not one-size-fits-all solutions for cybersecurity. Different cyber threats have unique characteristics, making it challenging for a single ML model to effectively handle all types of cyberattacks. In cybersecurity is their potential to achieve high accuracy in detecting cyber threats. ML models can analyze large datasets and learn to recognize patterns and anomalies, leading to effective threat detection.It is the unavailability of representative and benchmark datasets for each threat domain in cybersecurity. While datasets are crucial for training and testing ML models,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204159 |

the paper mentions that there is a lack of comprehensive and widely accepted datasets that cover the diversity of cyber threats. [1]

Jaksic (2023) The primary objective of this article is to provide a comprehensive survey of bio-inspired optimization methods, covering the most recent information and developments in the field. It is the potential for bio-inspired optimization methods to tackle complex and multi-objective optimization problems. These methods often excel in exploring large search spaces and finding solutions that may not be apparent through traditional optimization approaches. Due to the abundance of metaheuristic algorithms and the evolving nature of the field, selecting the most suitable method can be challenging and may require both expertise and trial-and-error. In the context of optimization, the accuracy of a method can be related to its ability to find nearoptimal solutions efficiently. [2]
Bacanin (2022) The primary objective of this manuscript is to introduce a novel version of the Sine Cosine Algorithm (SCA) metaheuristic, referred to as the diversity-oriented SCA (DOSCA), and to implement it within a machine learning framework. It is the introduction of the DOSCA algorithm, which is designed to address the limitations of the original SCA variant. It does not delve into the specific limitations or challenges associated with the DOSCA algorithm. The DOSCA algorithm, when applied to LR training and XGBoost hyper parameter optimization, is reported to achieve superior accuracy levels. It mentions the intention to conduct future experiments on more real-world datasets. [3]

Jantan (2017) The primary objective is to introduce and utilize a recent meta-heuristic algorithm called EBAT (Evolved Bat Algorithm) for training feedforward neural networks (FFNN) in the context of spam email detection. It is the utilization of the EBAT algorithm, a recent meta-heuristic approach derived from the original bat algorithm. It is that it does not delve into the specific limitations or challenges associated with the EBAT algorithm or the FFNN-EBAT approach. While it emphasizes the superiority of FFNN-EBAT over other algorithms. It is that the FFNN-EBAT approach yields better quality results in terms of accuracy when compared to other training algorithms, including ant-colony optimization, bat algorithm, differential evolution algorithm. Investigating the robustness of the EBAT algorithm. [4]

Al- Rawashdeh (2019) The primary objective of this research paper is to develop and evaluate hybrid algorithms that combine the Water Cycle Algorithm (WCA) with the Simulated Annealing (SA) metaheuristic for the purpose of feature selection in the context of spam email detection. It is the innovative use of hybrid metaheuristic algorithms, combining WCA and SA, to optimize feature selection for spam email classification. It is that it does not explicitly discuss the limitations or challenges associated with the proposed hybrid algorithms. These have various hybridizations of WCA and SA, including lowlevel, interleaved, and high-level hybridizations. [5]

III.METHODS

This model endeavors to boost the performance efficiency of the network Intrusion Detection System (IDS) through a hybrid approach that integrates several meta-heuristic algorithms, namely PSO, MVO, GWO, MFO, WOA, FFA, and BAT. The proposed hybrid model architecture is depicted in Fig. 1. The primary goal is to improve performance by minimizing the number of relevant features during the classification of the dataset for the detection of generic attacks. Each subsection below provides a detailed explanation of the stages within the proposed model.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204159 |

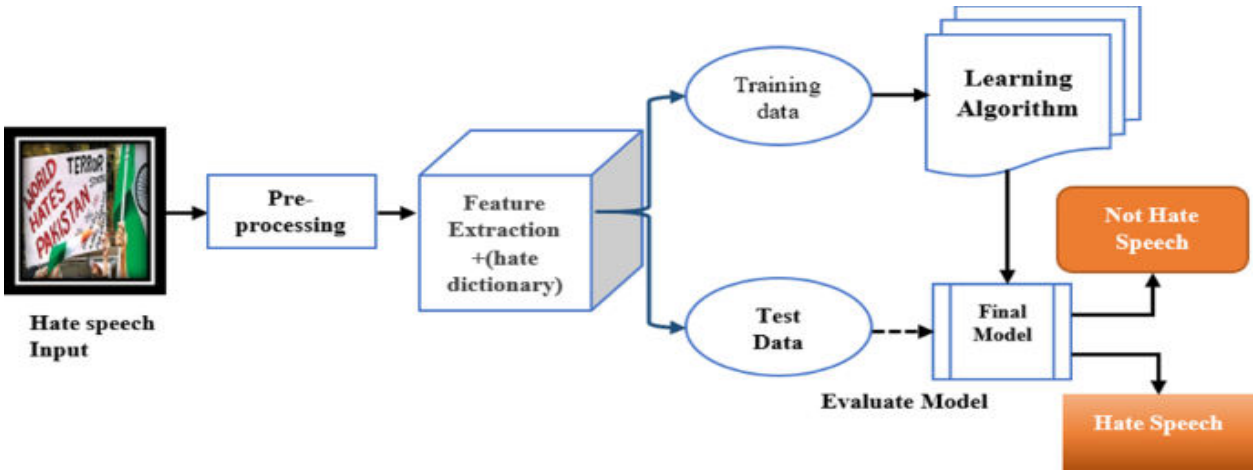

Fig 2: Hate speech, toxicity detection

The UNSW-NB15 dataset serves as a valuable resource in the realm of cybersecurity research and the assessment of Intrusion Detection Systems (IDS). Originating from the University of New South Wales (UNSW) in Australia, this network traffic dataset is specifically tailored for the evaluation of IDS capabilities. The inclusion of "NB15" in its nomenclature denotes its association with the NSL-KDD dataset, a widely recognized dataset within the intrusion detection domain. This connection underscores the dataset's significance and builds on the foundations laid by NSL-KDD, contributing to the advancement of intrusion detection methodologies

There are a variety of data reduction techniques that can be used, such as sampling and oversampling. Sampling is a technique that can be used to reduce the size of a data set by randomly selecting a subset of the data. Resampling is a technique that can be used to reduce the size of a data set by creating new data points that are similar to existing data points. The pre-processing step is a critical step in NIDS development. By carefully pre-processing your data, you can improve the performance of your machine learning models and detect malicious network traffic more effectively.

IV.RESULT ANALYSIS

Particle Swarm Optimization (PSO) is a population optimization algorithm inspired by the social behavior of birds and fish. It was introduced by James Kennedy and Russell Eberhart in 1995. PSO is part of the broader category of swarm intelligence, where multiple groups of people (particles) collaborate to search for optimal solutions in a multidimensional search space. PSO is based on the ability to interpret each solution in the swarm as a particle. If we consider each particle as a position in the search space, we have:

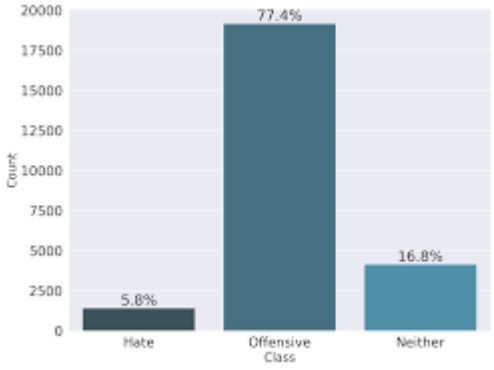

Fig 3: Result analysis

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204159 |

Moth Flame Optimization (MFO) is a nature-inspired optimization algorithm introduced in 2019 by Gai-Ge Wang. Inspired by the mating behavior of butterflies and their attraction to fire, MFO aims to effectively solve optimization problems by modeling the natural behavior of moths. moth. The basic concept of MFO came from studies of the light-seeking cycle of butterflies in nature, called transverse orientation. Moths move in a spiral and tend to maintain an angle similar to that of the light emitted by humans.

According to my data, the proposed hybrid model improves network IDS by reducing features. Time required to build a detection model. Besides this my results show dominance J48 in SVM and RF within the required time. functional decline and The classification results show that the MFO-WOA and FFA-GWO models reduce the number of features to 15. The MVO-BAT model provides features with similar accuracy, sensitivity, and F-measure for all features. We reduce the number of features to 24, using the same accuracy, sensitivity, and F-measure as all features for all classifiers. According to my data, the proposed hybrid model improves network IDS by reducing features. Time required to build a detection model. Besides this my results show dominance J48 in SVM and RF within the required time. functional decline and The classification results show that the MFO-WOA and FFA-GWO models reduce the number of features to 15. The MVO-BAT model provides features with similar accuracy, sensitivity, and F-measure for all features. We reduce the number of features to 24, using the same accuracy, sensitivity, and F-measure as all features for all classifiers.

V.CONCLUSION

A comprehensive analysis of the provided table reveals that the NN model stands out as the most accurate, achieving an impressive 99.81% accuracy. The SVM model also demonstrates remarkable performance with an accuracy of 96.90%. The phishing detection model, C4.5 Decision Tree, IWTSObased HAN, and Grey Wolf Optimization (GWO) models also exhibit noteworthy accuracy levels, ranging from 96.67% to 99.37%. While the Multinomial Naive Bayesian (MNB) model falls behind with an accuracy of 85.21%, the ATD-SGAN model lags significantly, achieving an accuracy of only 42.29%. In terms of dataset usage, Spam-Base emerged as the most prevalent choice, employed in two of the studies. The remaining datasets included Reuter, SOMLAP, Davidson, PhishTank, BLTE, ATD-SGAN, UK-2011 Webspam, and NSL-KDD. Overall, the NN and SVM models stand out as the most effective approaches, while the ATD-SGAN model requires further refinement. The choice of dataset also plays a crucial role in model performance, with Spam-Base proving to be a favorable choice. These insights provide valuable guidance for selecting and evaluating spam detection models. the presented analysis of spam detection models highlights the NN and SVM models as the most effective approaches, offering superior accuracy and reliability. For situations demanding high-accuracy spam detection, NN and SVM should be prioritized. Additionally, C4.5 Decision Tree, IWTSObased HAN, and Grey Wolf Optimization models provide a balance between accuracy and efficiency.

REFERENCES

[1] Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., & Xu, M. (2020). A survey on machine learning techniques for cyber security in the last decade. IEEE access, 8, 222310-222354.

[2] Jaksic, Z., Devi, S., Jaksis, O., & Guha, K. (2023). A Comprehensive Review of Bio-Inspired Optimization Algorithms Including Applications in Microelectronics and Nanophotonics. Biomimetic, 8(3), 278.

[3] Bacanin, N., Zivkovic, M., Stoean, C., Antonijevic, M., Janicijevic, S., Sarac, M., & Strumberger, I. (2022). Application of natural language processing and machine learning boosted with swarm intelligence for spam email filtering. Mathematics, 10(22), 4173.

[4] Jantan, A. M. A. N., Ghanem, W. A. H. M., & Ghaleb, S. A. (2017). Using modified bat algorithm to train neural networks for spam detection. J. Theor. Appl. Inf. Technol., 95(24), 1-12.

[5] Al- Rawashdeh, G., Mamat, R., & Abd Rahim, N. H. B. (2019). Hybrid water cycle optimization algorithm with simulated annealing for spam e-mail detection. IEEE Access, 7, 143721-143734.

[6] Almousa, B. N., & Uliyan, D. M. (2023). Anti-Spoofing in Medical Employee's Email using Machine Learning Uclassify Algorithm. International Journal of Advanced Computer Science and Applications, 14(7).

[7] Ketsbaia, L., Issac, B., Chen, X., & Jacob, S. M. (2023). A Multi-Stage Machine Learning and Fuzzy Approach to Cyber-Hate Detection. IEEE Access.

[8] Kumar, N., & Sonowal, S. (2020, July). Email spam detection using machine learning algorithms. In 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 108-113). IEEE.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204159 |

[9] Darvishpoor, S., Darvishpour, A., Escarcega, M., & Hassanalian, M. (2023). Nature-Inspired Algorithms from Oceans to Space: A Comprehensive Review of Heuristic and Meta-Heuristic Optimization Algorithms and Their Potential Applications in Drones. Drones, 7(7), 427.

[10] Beegum, T. R., Idris, M. Y. I., Ayub, M. N. B., & Shehadeh, H. A. (2023). Optimized routing of UAVs using Bio-Inspired Algorithm in FANET: A Systematic Review. IEEE Access.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details