



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Email Spam Detection Using Machine Learning

Prof. Kamble D. R., Yewale Pradip Dashrath, Chopade Sudam Sambhaji, Gujar Dattatray Navanath,

Assistant Professor, Department of Computer Engineering, S.B. Patil College of Engineering, Indapur, India

B. E. Department of Computer Engineering, S.B. Patil College of Engineering, Indapur, India

ABSTRACT- For business purposes, email is the preferred form of formal communication. Despite there being other forms of communication, email usage keeps growing. In the modern environment, where email volume is increasing daily, automated email management is crucial. More than 55 percent of the total emails are deemed to be spam. This demonstrates how these spams suck up user resources and time while producing little helpful results. It is crucial to comprehend the various spam email classification approaches and their workings because spammers employ sophisticated and inventive methods to carry out their illicit actions via spam emails. The machine learning algorithms used for spam classification are the main topic of this paper. Additionally, this paper offers a thorough analysis and assessment of previous research on various machine learning methodologies, email properties, and techniques. Additionally, it outlines potential paths for future study as well as difficulties encountered in the field of spam classification.

KEYWORDS: email spam detection; feature selection; classification; optimization.

I. INTRODUCTION

Real-world situations often have datasets with large dimensions and superfluous, worthless data, which makes it difficult to analyse this data. Feature selection (FS) is one of the pre-processing procedures in machine learning that can eliminate pointless and unnecessary characteristics from a dataset and identify the final subset of crucial features that will improve the performance of machine learning algorithms. 1–3 In fact, FS is a crucial and frequently used technique in data mining and machine learning for the reduction of dimensions through the removal of unnecessary and redundant data from the dataset in order to achieve the optimal feature subset that increases the efficiency and effectiveness of classification algorithms. Finding the ideal feature subset, however, poses a complicated optimization problem that cannot be resolved using standard techniques. In actuality, the objective of FS is to identify a subset of features from the total collection of features in order to enhance the efficiency of learning algorithms in terms of classification accuracy or learning time. Two frameworks, including the wrapper-based and lter-based techniques, have so far been suggested for effectively resolving the FS problem. 6, 7 the two essential elements of the former techniques are the search strategy and evaluation criterion. 8 The search strategy specifies the procedure for producing a solution for an ideal feature subset. A specific criterion is used to evaluate each created solution. The subset creation and evaluation procedure is continued until a stopping criterion is fulfilled in this technique, of course, in an effort to improve search methods in subsequent rounds. Contrary to the first set of approaches, the second set uses abundance, relationship, and connection between features to detect superfluous and unnecessary features. Problems can be divided into two groups: wrapper based approaches and filter-based solutions.

II. RELATED WORK

MajdiMafarja, Ibrahim Aljarah, HossamFaris, Abdelaziz I. Hammouri, Ala' M. Al-Zoubi, SeyedaliMirjalili- Real-world problems contain enormous amounts of data, which makes analysis difficult and frequently wrong owing to redundant or unnecessary data. In Mafarja and Mirjalili, Whale Optimization Algorithm (WOA)-based FS techniques have been proposed (2017). Different hybridization models between WOA and the Simulated Annealing method were put forth and used to the FS problem in this paper [1].

FarhadSoleimanian Gharehchopogh1,* , HojjatGholizadeh- Dedicated the increasing complexity of models and the requirement for speedy engineering decisions, research and engineering attention have been given to metaheuristics for decision-making purposes. This publication provides an updated review of WOA research for hybridization, improvements, and variants. Engineering, Clustering, Classification, Robot Path, Image Processing, Networks, Task



Scheduling, and other engineering applications are among the categories covered in the evaluations. The examined literature indicates that WOA is mostly employed in the engineering field to address optimization issues [2].

Maryam SamadiBonab, Ali Ghaffari, FarhadSoleimaniGharehchopogh, PayamAlemi- In this paper, a hybrid of the FFA and ALO has been used to perform feature selection on the IDS. In order to combine the advantages of the algorithms, the FFA-ALO has been developed to solve the problem of feature selection in IDS. A suitable diversity of population must be provided during the search phase to avoid being caught in local optima [3].

Sasan Gharehpasha, Mohammad Masdari, Ahmad Jafarian- In this paper, the advent of this technology, the cost of power computation, application hosting, content storage, resource wastage, and delivery is reduced significantly. Cloud computing provides the possibility of merely concentrating on business goals instead of expanding hardware resources for users. Challenging work in virtualization technology is the placement of virtual machines under optimal conditions on physical machines in cloud data centers [4].

Mehdi Hosseinzadeh, Hawkar Kamaran Hama Marwan Yassin Ghafour, Mohammad Masdari, Omed Hassan Ahmed, Hemn Khezri- Multi-criteria decision making (MCDM) methodologies have been widely used in the service selection process to assist a decision-maker in evaluating the weight of each QoS component and rating the services offered by various service providers. The state-of-the-art MCDM-based service selection strategies put forward in the literature are thoroughly examined in this work [5].

SankalpaArora, PriyankaAnand- In this paper, the best feature subset for classification purposes in a wrapper-mode is chosen using binary variations of the Butterfly Optimization Algorithm (BOA), which are proposed and applied in this paper. BOA is a freshly developed method that hasn't yet been used consistently to solve feature selection issues. By minimising a specific fitness function, BOA can efficiently search the feature space for an optimal or nearly optimal feature subset. Applying the two proposed binary BOA versions, the best feature combination that maximises classification accuracy while reducing the number of selected features is chosen [6].

Md. MonirulKabira, Md. Shahjahan b, Kazuyuki Murase- In this paper, we offer ACOFS, a novel hybrid ant colony optimization (ACO) approach for feature selection (FS). The selection of a subset of prominent features that is less in size is a crucial component of this approach. Utilizing both the benefits of wrapper and filter techniques, ACOFS employs a hybrid search method. New sets of rules for pheromone update and heuristic information measurement were created in order to support such a hybrid search. While building graph (subset) paths using a limited strategy, the ants are instead led in the right direction at every stage of the process [7].

HossamFaris, Majdi M. Mafarja, Ali AsgharHeidari, Ibrahim Aljarah, Ala' M. Al-Zoubi, SeyedaliMirjalili, Hamido Fujita- In order to address FS issues, an improved binary SSA-based optimizer using transfer functions and a crossover strategy was put forth in this study. 22 reputable benchmark datasets were used to test the suggested methodologies. The classification accuracy, features, and fitness metrics were examined in order to find the best TF for binary versions. Detailed statistical tests were also supplied. The BSSA with S3-shaped TF and crossover outperforms other hybrid variants, according to comparisons between the suggested versions. The effectiveness of the BSSA S3 CP approach was evaluated against three cutting-edge techniques and a number of algorithms from earlier publications. Comparative analyses of the BSSA S3 CP against bGWO, BGSA, and BBA revealed that the suggested technique was more effective in terms of accuracy and fitness values for various FS problems [8].

Nailah Al-Madi, Hossam Faris, Seyedali Mirjalili- In order to handle discrete problems, we present a new binary version in this work that is based on the MVO algorithm. Each solution in the suggested BMVO is represented by binary values, and continuous values are converted to binary values using a v-shaped function. Various benchmark functions will be used to gauge the performance of the BMVO algorithm. Along with the two real-world issues that are the knapsack problem and feature selection, which are covered in the following sections, It is significant to note that whereas earlier works have addressed the problem using multi-objective optimization, our challenge in this study is defined as a single objective optimization problem [9].

Yongquan Zhou, Fahui Miao, QifangLuo- The paper It is more challenging to calibrate a fractional fuzzy controller since it has more control parameters. This research suggests using the symbiotic organisms search technique to optimise the parameter controller tuning process because the fuzzy controller's parameter controller tuning values obviously affect how well it controls its environment. The symbiotic organisms search-based parameter controller



tuning approach was put to the test against parameter controller tuning based on five other representative swarm intelligence algorithms in a significant number of simulation tests. The experimental findings demonstrate that the symbiotic organisms search algorithm more effectively optimises the fractional fuzzy controller's parameters [10].

III. EXISTING SYSTEM/OPEN ISSUES

Automatic email filtering may be the most effective method of detecting spam but nowadays spammers can easily bypass all these spam filtering applications easily. Negative comments the risk of using an email filter is that messages could be mistakenly classified as spam due to an error in the algorithm. Even with the strongest spam filters available, mails might still end up with incorrect labels, claims Steven Scott, a Bayesian specialist.

IV. CONCLUSION

In order to reduce dimensions by removing redundant characteristics, FS is one of the crucial procedures in pre-processing information due to the abundance of information available. Accuracy of the classifier while choosing the crucial features and eliminating the redundant ones. As a result, FS directly affects the precision and quickness of machine learning classifiers. Recent meta-heuristic algorithms have caught the interest of numerous researchers due to their randomness and simplicity, as well as the fact that they are successful and promising FS techniques. One of the most effective meta-heuristic algorithms is called SOS, and it draws its inspiration from how organisms in nature compete with one another.

REFERENCES

- [1] Majdi Mafarja, Ibrahim Aljarah, Hossam Faris, Abdelaziz I. Hammouri, Ala' M. Al-Zoubi, Seyedali Mirjalili, Binary Grasshopper Optimisation Algorithm Approaches for Feature Selection Problems, Expert Systems With Applications (2018)
- [2] Farhad Soleimanian Gharehchopogh^{1,*}, Hojjat Gholizadeh, A Comprehensive Survey: Whale Optimization Algorithm and its Applications
- [3] Maryam Samadi Bonab¹ Ali Ghaffari, Farhad Soleimanian Gharehchopogh, Payam Alemi, A wrapper-based feature selection for improving performance of intrusion detection systems
- [4] Sasan Gharehpasha¹ · Mohammad Masdari¹ · Ahmad Jafarian, Virtual machine placement in cloud data centers using a hybrid multi-verse optimization algorithm.
- [5] Mehdi Hosseinzadeh, Hawkar Kamaran Hama, · Marwan Yassin Ghafour, Mohammad Masdari, Omed Hassan Ahmed, Hemn Khezri, Service Selection Using Multi-criteria Decision Making: A Comprehensive Overview
- [6] Sankalpa Arora^a, Priyanka Anand, Binary butterfly optimization approaches for feature selection, September 1, 2018
- [7] Md. Monirul Kabir^a, Md. Shahjahan^b, Kazuyuki Murase, A new hybrid ant colony optimization algorithm for feature selection, 2011
- [8] Hossam Faris, Majdi M. Mafarja, Ali Asghar Heidari, Ibrahim Aljarah, Ala' M. Al-Zoubi, Seyedali Mirjalili, Hamido Fujita, An Efficient Binary Salp Swarm Algorithm with Crossover Scheme for Feature Selection Problems, Knowledge-Based Systems (2018)
- [9] Nailah Al-Madi, Hossam Faris, Seyedali Mirjalili, Binary multi-verse optimization algorithm for global optimization and discrete problems (2019)
- [10] Yongquan Zhou, Fahui Miao^a, Qifang Luo, Symbiotic organisms search algorithm for optimal evolutionary controller tuning of fractional fuzzy controllers



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 **6381 907 438** **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details