Survey on Meta-Heuristic Algorithms for Feature Selection

R. Suresh[1], C. Punitha Devi[2], S. Akasthiya[3], M. Arthi[4], A.Keerthana[5]

[1]*Associate Professor, Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry, India*

[2]*Professor, Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry, India*

[3,4,5]*Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry, India*

[1]sureshramanujam78@smvec.ac.in, [2]punithadevi@smvec.ac.in, [3]akasthiya2899@gmail.com, [4]arthimurugan99@gmail.com,[5]akkeerthana225@gmail.com

***Abstract*— Selection of features is method of choosing a set of relevant options to be used in model construction. It's different from spatial property reduction. Each strategy gets to scale back the number of attributes within the dataset. The features that are most useful in a dataset are a key step in understanding the machines. Unnecessary features reduce the training speed, decrease the model's interpretability and, most importantly, decrease the test output produced by generalization. A high level or heuristic approach is a metaheuristic one developed to identify, create or choose a heuristic that provides a sufficiently appropriate Optimization approach problem. We have discussed about particle swarm optimization, harmony search, genetic algorithm, firefly algorithm, whale algorithm, grey wolf algorithm in this paper.**

Keywords — Feature Selection, Machine learning, Meta heuristic algorithm, Optimization problems

I. INTRODUCTION

A feature is a property or characteristic of an object, where the property can be measured. Any machine learning algorithm takes the features along with their values as input and produce the optimal classifier. Feature selection is one of the longest existing methods that deals with these problems. Its goal is to select a minimal sub-set of features according to some appropriate parameters so that the original task can be performed in good if not better quality. Selection of features is a method of selecting a minimum subset of n features from the original set of N features based on certain evaluation criterion [1]. It is the best subset which contains the least amount relevant features by discarding the unimportant features from original set of N features.

In general, Process of feature selection consists of three basic steps: Assessment of subsets, discontinuation of criteria, result validation and three main categories of feature selection methods: wrapper, filter and embedded.

Filter method:

The subset selection is performed without involving any mining algorithms, only be relaying on general characteristics of the data. In this approach, the subset selection is independent of any particular classification algorithm and the selected subsets are supposed to be useful for several different classification algorithms. This method is quite popular mainly due to their computational efficiency even for large datasets but the less computational effort and quality of selected features are the main disadvantages of this method.

Wrapper method:

The selection of features takes into consideration the classification algorithm to be applied to the selected functions. It selects a sub pack of apps that is “optimized” For a given ranking algorithm to improve the mining performance. This method selects the best subset with higher predictive accuracy with the expensive computation effort, but the selected feature cannot be used for several different classification algorithms.

Embedded method:

The embedded method has the advantage in both wrapper and filter methods by using their various evaluation method in different search stages.

II. RELATED WORKS

In most of the classification problems, huge number of features are introduced into the dataset to describe the target concepts in an efficient way. Optimization of Particulate Swarm (PSO) is one among the meta-heuristic algorithms. The paper aims to propose a list of the applications using PSO based on filter method with the view of selecting minimum features to achieve high classification accuracy than making use of all features. PSO which was introduced in 1995 is technique for an evolutionary computation [2]. PSO Mimics individual attitudes such as flocking of birds and schooling of fish. In the Optimization of Particle Swarm algorithm, candidate solutions are concealed as particles within the search area and are referred as population or swarm. The initial stage of PSO is the random initialization of the population of particles. In order to search for the most appropriate solution, the entire swarm moves in the search space and the Place of every particular particle is updated based on the experience of its own and its neighboring particles [3]. When PSO was proposed it was mainly used for the purpose of solving problems in real-number search spaces. Yet, most of the optimization problems like feature selection comes in a discrete search space. Optimization of Binary Particle Swarm was proposed for this purpose, to solve discrete problems. This proposed methodology is capable of reducing the number of features efficiently thus achieving better accuracy in most of the cases.

Harmony search (HS) is a newly developed Meta-heuristic algorithm of optimization imitating the phenomenon of musical improvisation, in which, each musician plays a note with a view of discovering a best harmony all together. Harmony search proved to work efficiently in numerous engineering optimization problems and machine learning tasks. The proposed algorithm tries to find a solution vector that enhances the given objective function [4]. In such an iterative search process, each musician (decision variable) would choose a note (variable value) to find out the most suitable harmony (a potential global optimum) along with the other musicians. The main difficulty in the set of characteristics the value of feature subsets effectively, at the same time maintaining their original semantics [5]. The iterative refinement method employed by HS feature selection intents to address the above mentioned issue which is made possible by its flexible mapping of musical concepts onto their associated elements in FS. In particular, a musician is not tied to a specific feature, thereby becoming an independent, single-feature selector. This forms a sharp contrast to many alternative methods that rely on binary-valued feature subset representation.

Ant colony Optimization (ACO) is one among several meta-heuristic algorithm that emulates the conduct of the ants. In the proposed system, the authors have used ACO algorithm to eliminate the features that are unimportant, irrelevant and redundant and selects the most fitting features from the data set that has wide number of functionalities [6]. The calculation of the Optimization of Ant Colony (ACO) is based on the social activity of ants when they search for a food source they leave behind a smelly material known as pheromone used to that mark the way between the food and source. At that time when an ant goes looking for food it smells the material that was left and follows the particular path that has more odorous material. This ant in the same way will leave the smelly substance along its way thus building the way's quality, which will later be used to attract the other ants in similar way to select the same path. If any ant needs to pick among different ways, it inclines toward the path, where the quantity of pheromone is high, which proves that most of the ants has gone through that particular way. The ants select the way that is shorter to take the nourishment of their settlements, the short routes get the highest level of pheromone when compared to longer routes [7]. If a way is not navigated by any subterranean insect, at that point the pheromone dissipates over time, subsequently diminishing the pheromone level of that way. In different words, diminish in pheromone powers that force the ants to investigate new ways for sustenance.

Artificial Colony of Bees (ABC) algorithm mimics the foraging conduct of honey bee colony and was proposed in 2005. In the view of the problem of optimization, the solutions that are portable and its respective fitness values are represented based on the sources of food and its amount of nectar, respectively. The previously explored food sources are exploited of the bees working and the information regarding the quality and position of food sources are shared with onlooker bees through waggle dance. The onlooker will take the decision on the food chain to be selected and that is to be exploited by using the information obtained by employed bees. Depending on the internal rule or any external clues, the scout bees searches for a different source of food [8]. The overall goal this is the article propose a new modified binary edition of a colony of artificial bees (ABC) to address feature selection problems. To attain the goal, the discrete binary ABC (DisABC) algorithm based on the similarity of Jaccard coefficient among individuals is further improved by introducing the neighborhood selection mechanism of the differential evolution (DE) strategy. The drawbacks are the form the structure of feature selection that is very complicated and the fragility of DisABC in the problems that are high-dimensional. This motivates to improvise the ability of searching in DisABC algorithm for the problem of feature selection.

Genetic algorithms (GA) are stochastic search based algorithms the natural selection mechanism that underlies biological growth. We can be applied to several problems related to search, optimization, and machine learning. The GA, when implemented as a selection method to features, starts by randomly developing the first population, which consists of a collection of subsets of features, each subset being one. The characteristics are represented in the form of n-binary vectors in each chromosome, in which the bit having value' 1' represents the respective role is selected and' 0' represents that the function isn't selected [9]. Using a fitness feature, the value of every chromosome is assessed. Fit chromosomes are chosen for creating new ones chromosomes. Then, new chromosomes are created from old chromosomes by the process of crossover and mutation. The key challenges in applying GA to any issue are creating initial populations, building new populations, and choosing an appropriate role for assessment. To address this problem, a new approach was proposed to change the GA by generating the initial population and enhancing genetic operators based on other variables, including chromosome partitioning and deciding the weight of each partition based on the meaning of chromosome entries (features) and taking in multiple populations.

Most swarm intelligent computation or optimization methods are based on higher organisms with more complex behaviors. Several studies have focused on the behaviors of microorganisms which are easier to be described qualitatively. Bacteria, as the simplest unicellular organisms, have simple patterns of behaviors which can be easily described [10]. Besides, as one of the oldest biological creatures on earth, bacteria's strong vitality and abilities to flexibly adapt to the complicated environment fully demonstrate their optimization instinct in the process of survival activities. For the two advantages above, research communities have developed the bacteria-inspired (BI) based methods from a new perspective. Such BI-based approaches are inspired by low-grade microorganism bacteria's social behaviors, and further consider the bacteria's foraging cycle as the method of problem solving optimization. To be more specific, the bacteria-inspired algorithm primarily stimulates three typical bacterial foraging behaviors, including chemotaxis, reproduction, and elimination/dispersal. Multi-Bacterial Foraging Optimization (MOBFO) was suggested to further boost the bacteria-algorithm efficiency and add a new variant to the device selection problem [11]. This algorithm is used to seek the most appropriate feature subset, every bacterium generates a potential solution satisfying problem constraints. More specifically, each bacterium is endowed with three attributes, i.e. the features being selected, value of corresponding classification error rate and the size of feature subset.

Moth-Optimization (MFO)-based function selection is one of the meta-optimization techniques. Moths have adapted to

travel in the night using moon light and are dependent on a navigation system called transverse orientation. A moth flies through this process holding a fixed angle relative to the moon (i.e., the light source). The technique ensures a straight line flight, since the moon is far away from the moth. When moths see an artificial light created by man, they seek to keep the sun at a similar angle to travel in straight line [12]. However, holding the light source at a similar angle produces a useless or hazardous spiral flight path for moths, because this light is very close to the moon. It may be observed that for artificial lights the moth eventually converges toward the light. An algorithm inspired from this type of motion is MFO. In MFO, moths continuously change their positions to whatever point in the space depending on the spiral equation. MFO's performance is substantially better than GA and PSO which are popular in wrapper-based selection of features.

Ant lion optimizer (ALO) is a recently proposed optimization algorithm. The ALO algorithm imitates the hunting process of nature's ant-lions. Mostly, the killing of ant lions (doodlebugs) in larvae and their period of adulthood is for reproduction. An ant lion larva digs a cone-shaped whole (trap) in the sand by pushing along a circular path with its giant jaw and tossing out sand. The larvae then hide under the bottom of the spike, waiting for insects (preferably ant) get caught in the pit. When the ant lions have learned that this prey is in pit, they attempt to capture it by tossing it back into the pit burying a mouse. When the prey is trapped in the mouth , it is dragged under the ground and consumed. They throw the leftovers outside the trap after consuming their prey, and amend it for the next hunting process. Another curious behavior observed during ant lion's lifestyle is that trap size is affected by two things: hunger level and moon shape [13]. The ACO algorithm was proposed based on this behavior of ant lion.

Firefly Algorithm (FA) serves as a target of vital significance formulated through Yang which gets motivation from social insects called fireflies [14]. Fireflies belong to insects, the primary features of which involves admirable flashing lights. The fireflies flashing patterns which is produced by a bioluminescence procedure, enjoy a special place for each of 2000 current living fireflies' species. Two main purposes of these flashing are to attract the potential prey and to mate partners. To simplify the FA, there exist the following 3 idealized rules (1) the total amount of fireflies have one sex. As a result, sexual attraction can happen to them. (2) The level of attractiveness has positive relationship to the brightness, meaning that the one with less bright tends to be attracted by the one having more light. Under the circumstance that there is no one with the largest amount of light, random attractiveness will happen. (3) The light intensity of a firefly gets influenced by the outlook of the unbiased function. In the quest process, an improved FA employs opposition-based learning in population initialization and opposition strategy that specifies the convergence rate to achieve the global optima.

Grasshopper Optimization Algorithm is a novel, bio-inspired algorithm that imitates the grasshoppers' swarming behavior in nature. The original GOA version was developed to give a solution to the problems of continuous optimization. However, many optimization Problems (e.g., FS) have discrete search space and decision variables. We suggest new binary variants of GOA in this paper and are built for FS problem solving [15]. Two transition features in the previous two methods, which pertains to two different groups of families are used to convert the continuous solutions in GOA to binary ones. Both approaches are known to be BGOA- S and BGOA-V. With the third approach, a novel method is used the binary version of the solution is developed in accordance to the position of the most appropriate solution found and the phase vector, rather than depending on the phase vector as in the first two approach. It is called the BGOA method. The output of the approaches proposed is evaluated and compared in the literature with similar methods, and with some thoughtfully chosen traditional FS approaches. GWO proposed in has been inspired from the social intelligence of grey wolves that prefer living in a group of 5-12 individuals. For this algorithm, four stages are known for order to represent the GWO leadership hierarchy: alpha, beta, delta, and omega. Alpha identified as male and female, and team leaders, decisions (e.g. hunting, Alpha's main duty is sleeping position and wake-). The input suggestions are Beta known to help alpha make choices, and the key responsibility of beta. Virgin performs as scouts, sentinels, caretakers, elders and hunters. Delta controls omega wolves those are dogs from Alpha and Beta. The wolves called as omega wolves must follow every other wolves. One of the core participants at GWO components to tune exploration and exploitation. The Grey Wolf Algorithm when tested with certain datasets provided good exploration and exploitation.

In 2016, Whale Optimization Algorithm (WOA) was introduced as a nature-inspired meta-heuristic optimization algorithm, which imitates the hunting behavior of humpback whales to solve complex continuous nonlinear functions [16]. The idea of this algorithm is inspired by the behavior of humpback whales which favor hunting school of krill or small fishes near to the surface. They plunge down around 12 m and then start spiral around the prey to trigger bubble and swim up to the surface, known as the bubble-net feeding process. The spiral bubble-net feeding man oeuvre is an essential idea adopted by WOA. The WOA algorithm consists of three steps of encircling prey, spiral bubble-net feeding man oeuvre, and search for prey. 1. Encircling prey The WOA algorithm assumes that the current best candidate solution is the target prey in the search space that it is not a priori known, or is near to the optimum. This replicates to recognize the location of prey and encircle them by humpback whales. The new agents will update their positions across the best search agent after the best search agent is defined. 2. Spiral bubble-net attacking method the humpback whales attack the prey with the bubble-net strategy which can be regarded in WOA as "exploitation phase". The simultaneous behavior of swimming humpback whales around the prey within a shrinking circle and along a spiral-shaped path simultaneously can be modelled by selecting between either the shrinking encircling mechanism or the spiral model with a probability of 50% to update the position of whales during the optimization.

In 2019, a new algorithm called Squirrel Search Algorithm (SSA) was proposed, which is a book optimization inspired by nature algorithm, with four suppositions condensed in the flying squirrel search process: flying squirrels and n trees in a deciduous forest, and one squirrel in a single tree; only one hickory tree and three oak trees in the forest [17]. On the acorn nuts the flying squirrels will shift towards the hickory nut tree. Some randomly selected squirrels are believed to fulfil their daily energy needs and are then considered heading towards the hickory nut tree. And the remaining squirrels should migrate

toward the acorn-nut trees to satisfy their daily need for energy. Yet predators have an effect on the flying squirrels foraging behavior.

III. Conclusion

Feature selection is one among the most essential phases in machine learning and image processing. As an effective technique for dimensionality reduction, feature selection has a broad application in different research areas. Selection of roles is technique to discard repetitive and irrelevant features without compromising the classification accuracy. If there is a massive amount of data set is taken, it very important to select the features appropriately since selecting any unwanted features or missing an essential feature would affect the overall performance and accuracy. So it's really necessary select a proper algorithm for feature selection. Over the past years many new meta-heuristic algorithms for feature selection have been proposed. The purpose of this paper explores on existing methods for feature selection to provide the most efficient algorithm that provides best selection of features.

References

[1] Pradnya Kumbhar, Manisha Mali(2016). A Survey on Feature Selection Techniques and Classification Algorithms for Efficient Text Classification. International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064

[2] Liam Cervante, Bing Xue, Mengjie Zhang, Lin Shang(2012). Binary Particle Swarm Optimisation for Feature Selection: A Filter Based Approach. WCCI 2012 IEEE World Congress on Computational Intelligence June, 10-15, 2012 - Brisbane, Australia

[3] Amir Rajabi Behjat1, Aida Mustapha, Hossein Nezamabadi-Pour, Md. Nasir Sulaiman, and Norwati Mustapha (2014). A New Binary Particle Swarm Optimization for Feature Subset Selection with Support Vector Machine. Recent Advances on Soft Computing and Data Mining SCDM 2014, Advances in Intelligent Systems and Computing 287, DOI: 10.1007/978-3-319-07692-8_5.

[4] Ling Zheng, Ren Diao, Qiang Shen (2014). Self-adjusting harmony search-based feature selection. Soft Computing, DOI 10.1007/s00500-014-1307-8.

[5] Messaouda Nekkaa, Dalila Boughaci (2015). Hybrid Harmony Search Combined with Stochastic Local Search for Feature Selection. Neural Process Lett., DOI 10.1007/s11063-015-9450-5

[6] Shima Kashef, Hossein Nezamabadi, An advanced ACO algorithm for feature subset selection, Neurocomputing,http://dx.doi.org/10.1016/j.neucom.2014.06.067

[7] Rajesh Dwivedi, Rahul Kumar, Ebenezer Jangam, Vishnu Kumar (2019). An Ant Colony Optimization Based Feature Selection for Data Classification. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5S4, February 2019

[8] Emrah Hancer, Bing Xueb, Dervis Karaboga, Mengjie Zhang (2015). A modified binary ABC algorithm based on advanced similarity scheme for feature selection. Applied Soft Computing, http://dx.doi.org/10.1016/j.asoc.2015.07.023,1568-4946/

[9] Yanan Mao, Zuoquan Zhang, Dingyuan Fan (2016). Hybrid feature selection based on improved genetic algorithm for stock prediction.

[10] Yu-Peng Chen , Ying Li , Gang Wang , Yue-Feng Zheng , Qian Xu , Jia-Hao Fan , Xue-Ting Cui , A Novel Bacterial Foraging Optimization Algorithm For Feature Selection, Expert Systems With Applications (2017), doi: 10.1016/j.eswa.2017.04.019

[11] Ben Niu, Wenjie Yi, Lijing Tan, Shuang Geng, Hong Wang, A multi-objective feature selection method based on bacterial foraging optimization. Natural Computing https://doi.org/10.1007/s11047-019-09754-6

[12] Hossam M. Zawbaa, E. Emary, B. PARV, Marwa Sharawi, Feature Selection Approach based on Moth-Flame Optimization Algorithm (2016)

[13] Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary ant lion approaches for feature selection. Neurocomputing, 213, 54–65. doi:10.1016/j.neucom.2016.03.101

[14] Xu, H., Yu, S., Chen, J., & Zuo, X. (2018). An Improved Firefly Algorithm for Feature Selection in Classification. Wireless Personal Communications. doi:10.1007/s11277-018-5309-1

[15] Mafarja, M., Aljarah, I., Faris, H., Hammouri, A. I., Al-Zoubi, A. M., & Mirjalili, S. (2018). Binary Grasshopper Optimisation Algorithm Approaches for Feature Selection Problems. Expert Systems with Applications. doi:10.1016/j.eswa.2018.09.015

[16] Mohamed A. Tawhid, Abdelmonem M. Ibrahim , Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm (2019). International Journal of Machine Learning and Cybernetics https://doi.org/10.1007/s13042-019-00996-5

[17] Hongping Hu , Linmei Zhang , Yanping Bai , Peng Wang, And Xiuhui Tan, A Hybrid Algorithm Based on Squirrel Search Algorithm and Invasive Weed Optimization for Optimization. IEEE Access, Digital Object Identifier 10.1109/ACCESS.2019.2932198