A Binary Metaheuristic Algorithm for Wrapper Feature Selection

Shokooh Taghian, Mohammad H. Nadimi-Shahraki*
Faculty of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran
Big Data Research Center, Najafabad Branch, Islamic Azad University, Najafabad, Iran
sh.taghian@sco.iaun.ac.ir, nadimi@iaun.ac.ir

Abstract— ***The classification accuracy is strongly affected by the quality of the input features used to build a learned-model. Nowadays, datasets grow enormously both in size and number of features. One of the major difficulties confronted by huge datasets analysis is existing redundant, noisy, and irrelevant features, which may reduce the performance of the classifier. Feature selection is an important preprocessing task, which aims to select the most effective subset of features from the original dataset. Therefore, using feature selection method is essential for enhancing the classification accuracy and reducing the complexity of the built model. In this paper, a wrapper-based binary Sine Cosine Algorithm (SCA) for feature selection named WBSCA is proposed. The proposed algorithm was compared with three well-known binary algorithms over seven classification datasets from the UCI machine learning repository. The results show the competitive performance of the proposed algorithm in searching the optimal subset and selecting the salient feature.***

Keywords- Classification; Feature selection; Metaheuristic algorithm; Binary metaheuristic Algorithm; Sine Cosine Algorithm

I. INTRODUCTION

Over the past two decades, with the rapid advancement in science and technology, large datasets with the massive number of features are produced. Redundant and irrelative features significantly degrade the accuracy of learned models, reduce the learning speed, and increase the computational complexity of built models [1]. This problem is known as the curse of dimensionality, which is a major problem in the field of data mining, knowledge discovery, and machine learning [2]. Feature selection is a way to tackle this problem by selecting a subset of salient features, eliminating irrelevant and redundant features, and reducing the dimensionality of the dataset [3]. Consequently, this reduction helps to speed up the learning process, simplify the learnt model, and increase the performance [4].

Feature selection based on the subset evaluation process is categorized into filter-based and wrapper-based methods [5]. The filter-based methods evaluate features without considering any classification algorithms. Basically, in the first step, features rank by using measures such as distance, dependency, and consistency [6]. Then the model is created by learning on train data with the selected features. This method is independent of a learning model, so the feature selection is performed only once. Some advantages of using filter methods are easily scalable to high dimensional datasets, computationally simple, and different classifiers can be used for evaluation. Common disadvantages of filter methods are to ignore the relevance of features and lacks the influence of features on the performance of the classifier [7]. Contrary to the filter method, wrapper utilizes a learning algorithm to evaluate a candidate feature subset. The wrapper method is known to be more accurate, but it is computationally more expensive [8].

Feature selection is a binary optimization problem [9], where solutions are bounded to the binary values 0 or 1. Each solution in this problem is represented in a binary vector where the length of the vector is equal to the number of the features exist in the dataset. The value one indicates the feature is selected; otherwise, the feature is not selected.

In recent years, metaheuristic algorithms due to their ability in solving different problems in various fields have attracted the attention of many researchers [10-14]. Accordingly, there have been proposed many metaheuristic algorithms inspired from almost different phenomena in the nature, math, or physics such as particle swarm optimization (PSO) [15], bat algorithm (BA) [16], gravitational search algorithm (GSA) [17] and sine cosine algorithm (SCA) [18]. Also, with increasing the number of variables and complexity of the problems, the high dimensional problems are an emerging issue, and recently some metaheuristic algorithms such as conscious neighborhood-based crow search algorithm (CCSA) [19] have been proposed for solving large-scale optimization problems. The metaheuristic algorithms are known for their ability in finding near-optimum solutions for optimization problems within a reasonable time. However, the original version of these algorithms are suitable for continuous problems, and cannot be directly applied to binary problems [20].

Sine Cosine Algorithm was recently proposed to solve continuous optimization problems and used two trigonometric functions sine and cosine. To our best knowledge, the SCA has been utilized in different applications [21-23], and even hybridized with some other optimization algorithms [24, 25]. The SCA operates on the continues search space and cannot be performed on problems with discrete search space; thus, it needs to be developed in order to apply in the binary spaces. In this paper, a wrapper-based binary Sine Cosine Algorithm

(SCA) named WBSCA is proposed for feature selection. This algorithm has an SCA-based search behavior in continuous space, while uses an erf function to calculate the position of solutions in binary space. Therefore, in WBSCA, each solution has two position vectors with real and binary representation.

The rest of the paper is organized as follows: in Section II, related works on feature selection are reviewed. The proposed algorithm is introduced in Section III. The experimental results are demonstrated in Section IV. Finally, Section V concludes the work and suggests for future works.

II. RELATED WORKS

Feature selection is an important preprocessing task, which aims to improve the classification accuracy by eliminating the redundant, irrelevant, and noisy features from datasets. Using a stochastic search approach like metaheuristic algorithms to explore a large portion of the search space and exploit near-optimal solutions is a way to tackle feature selection. Recently, population-based metaheuristic algorithms have been proposed and applied on feature selection and shown improved results. For instance, ant colony optimization (ACO) [26], which mimics the foraging behavior of ants, has been employed as a wrapper feature selection method [27]. Bat algorithm (BA) is another recent algorithm that has been successfully used in feature selection problem [28]. Gravitational search algorithm (GSA), which is based on Newton's law of motion and gravitation, has been employed as a wrapper feature selection [29]. Whale Optimization Algorithm (WOA) [30] is another recent algorithm that mimics the social behavior of humpback whales was used as a wrapper feature selection method [31]. Dragonfly algorithm (DA) [32] is another recent metaheuristic algorithm has been applied on feature selection problem [33].

III. WRAPPER-BASED BINARY SINE COSINE ALGORITHM (WBSCA)

In WBSCA, a population of solutions is randomly created within the continuous search space. Then, the SCA is used to search the continuous space for the best solution, in order to move other solutions toward it. The position updating of solutions is done by Eq. 1.

$$X_i^{t+1} = \begin{cases} X_i^t + \alpha \times \sin(\beta) \times |\sigma P_i^t - X_i^t| \ , & \varphi < 0.5 \\ X_i^t + \alpha \times \cos(\beta) \times |\sigma P_i^t - X_i^t| \ , & \varphi \geq 0.5 \end{cases} \tag{1}$$

where X_i^t is the value of i-th dimension of the current solution at iteration t, P_i^t is the position of the best solution in i-th dimension and t-th iteration. α, β, and σ are random numbers, in which α, σ are in [0, 1] and β is in [0, 2π]. The controlling parameter α is calculated by considering the maximum iterations T, the current iteration t, and a constant value a as shown in Eq. 2.

$$\alpha = a - t\frac{a}{T} \tag{2}$$

In the proposed algorithm, each solution with a real-valued position vector needs to have a binary-value position vector. The position vector is represented as an array of length N, where N is the total number of features in the dataset. Therefore, each solution has two position vector, which the binary position vector is a map of the real-valued vector. An example of these representations is shown in Fig. 1.

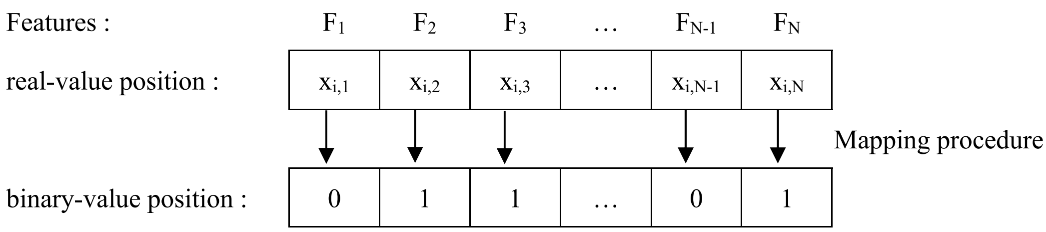

Figure 1. An example of mapping procedure

The real-valued positions are mapped into binary-values in two steps by using the probability and update equations. In the first step, the erf function, where erf is the Gauss error function, is utilized for calculating the probability of changing the solution's positions by Eq.3. Then, the updating position equation is employed to update the binary position vector of each solution, as shown in Eq. 4.

$$V(x_i^d(t+1)) = \left| erf(\frac{\sqrt{\pi}}{2}x) \right| = \left| \frac{\sqrt{2}}{\pi} \int_0^{(\frac{\sqrt{\pi}}{2})x} e^{-t^2} dt \right| \tag{3}$$

$$x_i^d(t+1) = \begin{cases} complement(x_i^d(t)), & if\ rand \le V(x_i^d(t+1)) \\ x_i^d(t), & otherwise \end{cases} \tag{4}$$

Then, found solutions are mapped into binary solutions, in order to find the best feature subset combination. The best combination is the one which has the maximum classification accuracy and the minimum number of selected features. The fitness of each solution calculates by Eq. 5 and assigns to each real-valued solution.

$$Fitness = \alpha \gamma_R(D) + \beta \frac{|S|}{|N|} \tag{5}$$

where γ_R (D) is the classification error rate of feature subset R, |S| is the number of selected features, |N| is the total number of features in the dataset, α and β are two parameters related to the importance of accuracy and number of selected features, $\alpha \in [0,1]$ and $\beta = 1 - \alpha$ [33].

IV. EXPERIMENTAL EVALUATION

A. Datasets and parameter settings

In this section, the performance of WBSCA is evaluated and compared to other binary algorithm exists in the literature. For the evaluation process, the experiments are performed on seven datasets. Table I shows the used datasets, which are from the UCI repository [34]. Each dataset is split into %80 for training and %20 for testing. All the experiments were repeated for 30 runs to obtain meaningful results. In this work, the k-nearest neighbor classifier (KNN) is used to indicate the classification error rate of the selected feature subset with k=5. All the experiments are performed on PC with Intel Core(TM) i7-3770 3.4GHz CPU and 8.00 GB RAM using MATLAB 2014 software.

TABLE I. STATISTICAL INFORMATION OF THE DATASETS

Dataset	No. of features	No. of instances	No. of classes
Iris	5	150	3
Glass	10	214	7
Exactly	14	1000	2
Exactly2	14	1000	2
Zoo	17	101	7
Waveform	22	5000	3
Sonar	61	208	2

The proposed WBSCA is compared with BBA, BGSA, and BDA. The initial and specific parameters of each algorithm are reported in Table II and can be found in the given references. In order to have a fair comparison, all the algorithms use the same initial settings. Each algorithm is randomly initialized with the population size, and the number of iterations are set to 20 and 300. The α parameter in the fitness function has a value of 0.99. Evaluation criteria for all the algorithms are considered as average classification fitness accuracy and number of selected features.

TABLE II. PARAMETER SETTINGS FOR ALGORITHMS

Algorithm	Parameter	Value
BBA	Q_{min}	0
	Q_{max}	2
	A	0.5
	r	0.5
BGSA	G_0	100
WBSCA	a	2

B. Numerical results

In this section, the results of the proposed algorithm, WBSCA, is compared with other state-of-the art binary metaheuristic algorithms which are widely used to solve the feature selection problem. Table III outlines the result of BBA, BGSA, BDA, and WBSCA based on the average and standard deviation of the accuracy. Note that the best results are highlighted in bold. As per results reported in Table Ⅲ, the WBSCA algorithm provides the competitive or even better results on used datasets. It achieves the maximum accuracy on Iris, Glass, Exactly2, Zoo, and Waveform in comparison to other algorithms, while on Exactly, and Sonar competes with BGSA, and BDA.

TABLE III. THE PROPOSED ALGORITHM VS OTHER BINARY METAHEURISTIC ALGORITHMS BASED ON AVERAGE ACCURACY

Dataset		BBA	BGSA	BDA	WBSCA
Iris	AVE	**1.0000**	**1.0000**	0.7867	**1.0000**
	STD	0.0000	0.0000	0.2657	0.0000
Glass	AVE	0.5829	0.5798	0.4000	**0.5891**
	STD	0.0432	0.0465	0.1407	0.0523
Exactly	AVE	0.7703	**0.9457**	0.8915	0.9380
	STD	0.1092	0.0846	0.2214	0.0785
Exactly2	AVE	0.7280	0.7392	0.6107	**0.7503**
	STD	0.0110	0.0091	0.0926	0.0125
Zoo	AVE	0.8633	0.9117	0.7850	**0.9183**
	STD	0.0320	0.0215	0.2604	0.0278
Waveform	AVE	0.8034	0.8191	0.7971	**0.8260**
	STD	0.0132	0.0083	0.1203	0.0070
Sonar	AVE	0.8619	0.8794	**0.9690**	0.9127
	STD	0.0392	0.0207	0.0199	0.0157

In terms of the number of selected features, in Table IV, the WBSCA obtains the smallest value as same as BGSA and BBA in Iris dataset. For the rest of the datasets, although the number of selected features of other algorithms is the smallest, they have a great degraded on the classification accuracy.

TABLE IV. THE PROPOSED ALGORITHM VS OTHER BINARY METAHEURISTIC ALGORITHMS BASED ON AVERAGE NUMBER OF SELECTED FEATURES

Dataset		BBA	BGSA	BDA	WBSCA
Iris	AVE	2.13	**1.00**	**1.00**	**1.00**
	STD	0.94	0.00	0.00	0.00
Glass	AVE	3.57	1.53	**1.00**	1.50
	STD	1.79	1.25	0.00	0.86
Exactly	AVE	**4.87**	7.27	6.00	7.40
	STD	1.74	1.14	0.00	0.97
Exactly2	AVE	**5.17**	6.70	5.73	6.33
	STD	2.29	1.73	2.15	1.88
Zoo	AVE	**6.23**	7.13	6.83	6.60
	STD	2.16	1.11	0.38	1.33
Waveform	AVE	**7.20**	13.40	15.10	14.40
	STD	2.25	2.08	1.94	2.01
Sonar	AVE	22.53	27.83	**19.30**	25.43
	STD	5.24	4.07	3.31	5.69

V. CONCLUSION

In this paper, a wrapper-based binary metaheuristic algorithm is proposed to deal with feature selection problems. The WBSCA inherits the SCA characteristics to search the continuous space, and it uses a mapping procedure in order to calculate the binary-valued position vector for each solution. The performance of the proposed algorithm was compared with the well-known binary algorithms in a wrapper-based feature selection method on seven datasets from the UCI repository. The experimental results were reported from the two different aspects of classification accuracy and the length of the subset features. It reveals that the WBSCA, in terms of classification accuracy, has merit among other binary algorithms. For future studies, it is recommended to use WBSCA in different practical applications. Also, it would be interesting to investigate the impact of other classifiers like SVM to verify and extend this approach.

REFERENCES

[1] M. M. Kabir, M. Shahjahan, and K. Murase, "A new local search based hybrid genetic algorithm for feature selection," Neurocomputing, vol. 74, 2011, pp. 2914-2928.
[2] H. Liu and H. Motoda, Computational methods of feature selection: CRC Press, 2007.
[3] H. Liu and H. Motoda, Feature selection for knowledge discovery and data mining vol. 454: Springer Science & Business Media, 2012.
[4] Guyon and A. Elisseeff, "An introduction to variable and feature selection," Journal of machine learning research, vol. 3, 2003, pp. 1157-1182.
[5] R. Kohavi and G. H. John, "Wrappers for feature subset selection," Artificial intelligence, vol. 97, 1997, pp. 273-324.
[6] C. Dhaenens and L. Jourdan, Metaheuristics for big data: John Wiley & Sons, 2016.
[7] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," bioinformatics, vol. 23, 2007, pp. 2507-2517.
[8] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," IEEE Transactions on Evolutionary Computation, vol. 20, 2015, pp. 606-626.
[9] E.-G. Talbi, Metaheuristics: from design to implementation vol. 74: John Wiley & Sons, 2009.
[10] H. Zamani and M. H. Nadimi-Shahraki, "Swarm Intelligence Approach for Breast Cancer Diagnosis," International Journal of Computer Applications, vol. 151, 2016, pp. 40-44.
[11] M. Banaie-Dezfouli, M. H. Nadimi-Shahraki, and H. Zamani, "A Novel Tour Planning Model using Big Data," in 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018, pp. 1-6.
[12] H. G. Arjenaki, M. H. Nadimi-Shahraki, and N. Nourafza, "A low cost model for diagnosing coronary artery disease based on effective features," International Journal of Electronics Communication and Computer Engineering, vol. 6, 2015, pp. 93-97.
[13] S. Taghian and M. H. Nadimi-Shahraki, "Binary Sine Cosine Algorithms for Feature Selection from Medical Data," arXiv preprint arXiv:1911.07805, 2019.
[14] E. S. Fard, K. Monfaredi, and M. H. Nadimi-Shahraki, "An Area-Optimized Chip of Ant Colony Algorithm Design in Hardware Platform Using the Address-Based Method," International Journal of Electrical and Computer Engineering, vol. 4, 2014, pp. 989-998.
[15] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995, pp. 39-43.
[16] X.-S. Yang and A. Hossein Gandomi, "Bat algorithm: a novel approach for global engineering optimization," Engineering Computations, vol. 29, 2012, pp. 464-483.
[17] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "GSA: a gravitational search algorithm," Information sciences, vol. 179, 2009, pp. 2232-2248.
[18] S. Mirjalili, "SCA: a sine cosine algorithm for solving optimization problems," Knowledge-Based Systems, vol. 96, 2016, pp. 120-133.
[19] H. Zamani, M. H. Nadimi-Shahraki, and A. H. Gandomi, "CCSA: Conscious Neighborhood-based Crow Search Algorithm for Solving Global Optimization Problems," Applied Soft Computing, in press, 2019, https://doi.org/10.1016/j.asoc.2019.105583.
[20] S. Taghian, M. H. Nadimi-Shahraki, and H. Zamani, "Comparative Analysis of Transfer Function-based Binary Metaheuristic Algorithms for Feature Selection," in 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), 2018, pp. 1-6.
[21] S. Das, A. Bhattacharya, and A. K. Chakraborty, "Solution of short-term hydrothermal scheduling using sine cosine algorithm," Soft Computing, vol. 22, 2018, pp. 6409-6427.
[22] K. Rajesh and S. Dash, "Load frequency control of autonomous power system using adaptive fuzzy based PID controller optimized on improved sine cosine algorithm," Journal of Ambient Intelligence and Humanized Computing, vol. 10, 2019, pp. 2361-2373.
[23] R. Sindhu, R. Ngadiran, Y. M. Yacob, N. A. H. Zahri, and M. Hariharan, "Sine–cosine algorithm for feature selection with elitism strategy and new updating mechanism," Neural Computing and Applications, vol. 28, 2017, pp. 2947-2958.
[24] K. Chen, F. Zhou, L. Yin, S. Wang, Y. Wang, and F. Wan, "A hybrid particle swarm optimizer with sine cosine acceleration coefficients," Information Sciences, vol. 422, 2018, pp. 218-241.
[25] H. Nenavath and R. K. Jatoth, "Hybridizing sine cosine algorithm with differential evolution for global optimization and object tracking," Applied Soft Computing, vol. 62, 2018, pp. 1019-1043.
[26] M. Dorigo and G. Di Caro, "Ant colony optimization: a new meta-heuristic," in Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406), 1999, pp. 1470-1477.
[27] U. Singh and S. N. Singh, "A new optimal feature selection scheme for classification of power quality disturbances based on ant colony framework," Applied Soft Computing, vol. 74, 2019, pp. 216-225.
[28] D. Gupta, J. Arora, U. Agrawal, A. Khanna, and V. H. C. de Albuquerque, "Optimized Binary Bat Algorithm for classification of White Blood Cells," Measurement, 2019.
[29] E. Rashedi, H. Nezamabadi-Pour, and S. Saryazdi, "BGSA: binary gravitational search algorithm," Natural Computing, vol. 9, 2010, pp. 727-745.
[30] S. Mirjalili and A. Lewis, "The whale optimization algorithm," Advances in engineering software, vol. 95, 2016, pp. 51-67.
[31] H. Zamani and M. H. Nadimi-Shahraki, "Feature selection based on whale optimization algorithm for diseases diagnosis," International Journal of Computer Science and Information Security, vol. 14, 2016, pp. 1243-1247.
[32] S. Mirjalili, "Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems," Neural Computing and Applications, vol. 27, 2016, pp. 1053-1073.
[33] M. M. Mafarja, D. Eleyan, I. Jaber, A. Hammouri, and S. Mirjalili, "Binary dragonfly algorithm for feature selection," in 2017 International Conference on New Trends in Computing Sciences (ICTCS), 2017, pp. 12-17.
[34] C. L. Blake and C. J. Merz, "UCI repository of machine learning databases," 1998.