S-shaped Binary Whale Optimization Algorithm for Feature Selection

Abdelazim G. Hussien, Aboul Ella Hassanien, Essam H. Houssein, Siddhartha Bhattacharyya and Mohamed Amin

Abstract Whale optimization algorithm is one of the recent nature-inspired optimization technique based on the behavior of bubble-net hunting strategy. In this paper, a novel binary version of whale optimization algorithm (bWOA) is proposed to select the optimal feature subset for dimensionality reduction and classifications problem. The new approach is based on a sigmoid transfer function (S-shape). By dealing with the feature selection problem, a free position of the whale must be transformed to their corresponding binary solutions. This transformation is performed by applying an S-shaped transfer function in every dimension that defines the probability of transforming the position vectors' elements from 0 to 1 and vice versa and hence force the search agents to move in a binary space. K-NN classifier is applied to ensure that the selected features are the relevant ones. A set of criteria are used to evaluate and compare the proposed bWOA-S with the native one over eleven different datasets. The results proved that the new algorithm has a significant performance in finding the optimal feature.

Keywords Whale optimization algorithm · Binary whale optimization algorithm S-shaped · Feature selection · Classification · Dimensionality reduction

A. G. Hussien (✉)
Faculty of Science, Fayoum University, Faiyum, Egypt
e-mail: aga08@fayoum.edu.eg

A. E. Hassanien
Faculty of Computers and Information, Cairo University, Giza, Egypt
e-mail: aboitcairo@gmail.com

E. H. Houssein
Faculty of Computers and Information, Minia University, Minya, Egypt
e-mail: essam.halim@mu.edu.eg

S. Bhattacharyya
Department of Computer Application, RCC Institute of Information Technology, Kolkata 700 015, India
e-mail: dr.siddhartha.bhattacharyya@gmail.com

M. Amin
Faculty of Science, Menoufia University, Shibin El Kom, Egypt

S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*, Advances in Intelligent Systems and Computing 727,
https://doi.org/10.1007/978-981-10-8863-6_9

1 Introduction

The feature selection (FS) is considered as a process of pre-processing in the machine learning. Selecting the subset of optimal features is one of the most difficult tasks for a huge and complex dataset [1–4]. The objectives of FS are to reduce the dimensionality of data and improve performance [5, 6]. FS has been proven to effectively eliminate irrelevant and redundant features. In recent years, data have become increasingly large in both attributes and instances. FS has been successfully applied in many applications such as text categorization [7], genome projects [8], customer relationship management [9], and image retrieval [10]. Therefore, at present, FS for high-dimensional data becomes very necessary for the learning tasks of the machine.

Meta-heuristic nature-inspired algorithms are now the most widely used algorithms to solve optimization problems. There are a plenty of natural meta-heuristic algorithms that are hired with FS problem such as binary bat algorithm (BBA) in [11], binary cuckoo search algorithm (BCSA) in [12], binary flower pollination algorithm (BFPA) in [13], binary flower pollination algorithm (BFPA) [14], firefly optimization algorithm for FS [15], and sine–cosine algorithm [16]. Moreover, Emary et al. have proposed the binary ant lion and the binary gray wolf optimization [17, 18], respectively. Whale optimization algorithm is a recent nature-inspiring algorithm that mimics the humpback whale in searching for prays. In this paper, we try to propose a novel approach based on sigmoid function [19].

This paper is organized as follow: a brief overview of WOA is provided in Sect. 2. In Sect. 3 we introduce our algorithm. Results of comparing new algorithm with native one is shown in Sect. 5. In Sect. 6 we summarize the proposed work.

2 Whale Optimization Algorithm

Mirjalili et al. [20] proposed the whale optimization algorithm (WOA) inspired from the whales' behavior. Their foraging behavior is called bubble-net feeding method. However, in WOA, the current best candidate solution is set either the target prey or close to the optimum. The other will try to update their position toward the best. Mathematically, the WOA simulates the swarming behavior as follows:

$$\mathbf{D} = |\mathbf{C} \cdot \mathbf{X}^*(t) - \mathbf{X}(t)| \tag{1}$$

$$\mathbf{X(t+1)} = \mathbf{X}^*(t+1) - \mathbf{A} \cdot \mathbf{D} \tag{2}$$

where t is the current iteration, X is the position vector, X^* is the position vector coincide to the best solution found, and A and C are coefficient vectors. A and C are defined as following:

$$\mathbf{A} = 2 \cdot \mathbf{a} \cdot \mathbf{r} - \mathbf{a} \tag{3}$$

$$\mathbf{C} = 2 \cdot \mathbf{r} \tag{4}$$

where r is located randomly in the range [0, 1] and a is decreased linearly from 2 to 0 over the iterations. This algorithm as many optimization algorithms has two phases: exploration and exploitation. **The exploitation phase**: Exploitation phase divided into two processes; (1) shrinking encircling mechanism: This can be obtained by decreasing a value's according to Eq. 4. Note that a is a random value between $[-a, a]$. (2) Spiral updating position: This approach calculates the distance between the whale and the prey. A spiral equation is used to mimic the helix-shaped movement as follow:

$$\mathbf{X}(t+1) = \mathbf{D}^{l} e^{bl} \cdot \cos(2\pi l) + \mathbf{X}^{*}(t) \tag{5}$$

where l is a random number in the range $[-1, 1]$ and b is a constant. A probability of 50% to choose between either shrinking encircling mechanism or the spiral model is assumed. Consequently, the mathematical model is as follow:

$$\mathbf{X}(t+1) = \begin{cases} \mathbf{X}^{*}(t) - \mathbf{A} \cdot \mathbf{D} & if\ p < 0.5 \\ \mathbf{D}^{\mathbf{l}} \cdot e^{bl} \cdot \cos(2\pi l) + \mathbf{X}^{*}(t) & if\ p \geq 0.5 \end{cases} \tag{6}$$

where p is a random number in a uniform distribution. **The exploration phase**: In another hand side, in the exploration phase, A has used random values within $1 \prec A \prec -1$ to force the agent to move away from this location and mathematically formulated as follow:

$$\mathbf{D} = |\mathbf{C} \cdot \mathbf{X}_{rand} - \mathbf{X}| \tag{7}$$

$$\mathbf{X}(t+1) = X_{rand} - \mathbf{A} \cdot \mathbf{D} \tag{8}$$

3 S-shaped Binary Whale Optimization Algorithm

In the basic WOA, the whales move in the search space to modify their positions to any point in the space, and this is called the continuous space. As regards to the nature of the FS issues, the solutions are limited to the binary space {0, 1} values. To solve FS problem, the continuous (free position) must be transforming to their corresponding binary solutions {0, 1} and hence motivates to propose a new version of the bWOA-S. The conversion is performed by applying an S-shaped transfer function. Algorithm 1 shows the steps of the proposed S-shaped binary whale optimization. The probability of transforming the position vectors' elements from 0 to 1 and vice versa has been adopted by the transfer functions and hence force the search agents to move in a binary space (Fig. 1).

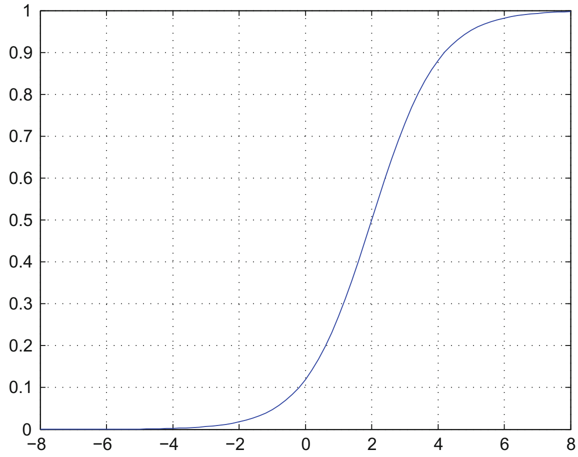

Fig. 1 S-shaped transfer function

Algorithm 1 Binary Whale Optimization S-shaped

Input: n number of whales in the population, and $MaxIter$ number of iteration.
Output: Optimal whale position.

Initialize a and n.
Find $X* = best$ search agent.
while Stopping criteria not meet to **do**
  for Whale$_i$ belong to whales **do**
    Calculate and Update a; A, C, p and l.
    if $p < 0.5$ **then**
      if $(|A| < 1)$ **then**
        Update position by Equation 2.
      else$(|A| \geq 1)$
        Choose search agent randomly (X_{rand})
        Update position by Equation 8.
      end if
    else$(p \geq 0.5)$
      Update position by Equation 5.
    end if
    Update $\mathbf{X}(t+1)$ from Equation 9
  end for
  if There exiting a search agent belong to the search space **then**
    Calculate the fitness for each agent.
    Update $X*$ if there is a better solution.
    $t = t + 1$
  end if
end while

The common S-shaped function is updating as shown in Eq. 9. Algorithm 1 shows the pseudo code of the proposed bWOA-S version.

$$y^k = \frac{1}{1 + e^{-v_i^k(t)}} \tag{9}$$

$$X_i^d = \begin{cases} 1 \; if \, rand < S(x_i^k(t+1)) \\ 0 \; otherwise \end{cases} \tag{10}$$

4 S-shaped Binary Whale Optimization Algorithm for Feature Selection

In order to solve the problem of FS, a new binary version of whale optimization algorithm bWOA-S was proposed, and hence, for feature reduction, N the number of different feature combinations would be 2^N which is a huge space of features to be searched exhaustively. So the proposed bWOA-S is applied to adaptive the search space and provided the best feature combination. The best feature combination is specifically obtained as well as the maximum classification accuracy with the minimum number of selected features. Equation 11 illustrates the fitness function used in bWOA-S to evaluate the individual whale positions.

$$Fitness = \alpha \gamma_R(D) + \beta \frac{|C - R|}{|C|} \tag{11}$$

where R is selected feature subset length, C is the total number of features, $\gamma_R(D)$ is the classification accuracy of condition attribute set R relative to decision D, and α and β are two parameters symmetric to the subset length and the classification accuracy, illustrates as $\alpha \in [0; 1]$ and $\beta = 1 - \alpha$. Therefore, this leads to the fitness function that maximizes the classification accuracy. Therefore, Eq. 11 is converted exactly into a minimization problem based on the classification error rate and selected features rather than the classification accuracy and unselected feature size. Thus, the obtained minimization problem defined as in Eq. 12.

$$Fitness = \alpha E_R(D) + \beta \frac{|R|}{|C|} \tag{12}$$

where $E_R(D)$ is the classification error rate, C is the total number of features, and R is the selected feature subset length. α and β are two parameters corresponding to the classification accuracy and feature reduction, illustrates as $\beta = 0.01$ and $\beta = 1 - \alpha$.

Eventually, a simple common K-NN classifier is applied. Therefore, K-NN is used to ensure that the selected features are the relevant ones. In the other hand, bWOA-S is adapted to maximize the feature evaluation criteria as defined in Eq. 12. Also, the individual feature is represented as a single dimension in the search space, and hence, the single feature represents the whale's position.

5 Experimental Results and Discussion

To provide a fair comparison, the experimental results are performed on different of 11 UCI benchmark datasets. Table 1 summarizes the 11 datasets from UCI machine learning that are used for the experiments and comparison, while Table 2 shows the parameter settings values. The datasets were selected to have different numbers of instances and attribute to represent various kinds of issues. In each dataset, instances are divided randomly into three different equal subsets, namely training, testing, and validation subsets in cross-validation manner. K-NN is applied in the experiment results using trial and error basis, and the best choice of K is 5, overall the different datasets. Meanwhile, every whale position produces one attribute subset through the training process. The training set is used to evaluate the K-NN classifier performance on the validation subset throughout the optimization process, and the bWOA-S is used to guide the FS process simultaneously.

Each algorithm has been performed 20 runs with random positioning of the search agents. Repeated runs of the compared algorithms were used to test the convergence

Table 1 List of data sets used in the experiments results

No.	Name	Features	Samples
1	Breastcancer	9	699
2	Tic-tac-toe	9	958
3	Zoo	16	101
4	WineEW	13	178
5	SpectEW	22	267
6	SonarEW	60	208
7	IonosphereEW	34	351
8	HeartEW	13	270
9	CongressEW	16	435
10	KrvskpEW	36	3196
11	WaveformEW	4	5000

Table 2 Parameter setting

Parameter	Value
Iterations number	70
Search agents number	8
Dimension	No. of features in the data
Search domain	$[0, 1]$
No. repetitions of runs	20
α parameter in the fitness function	0.99
β parameter in the fitness function	0.01

Table 3 Statistical mean fitness measure calculated for the compared algorithms on the different datasets using small, mixed and large initialization

No.	Small		Mixed		Large	
	WOA	bWOA-S	WOA	bWOA-S	WOA	bWOA-S
1	0.05305	0.04466	0.09977	0.04725	0.20643	0.15028
2	0.31558	0.22432	0.20914	0.20715	0.21441	0.20004
3	0.25203	0.11632	0.16824	0.139783	0.16629	0.16952
4	0.92738	0.897567	0.91507	0.915064	0.93309	0.91703
5	0.33557	0.28321	0.31313	0.307875	0.32361	0.30543
6	0.32991	0.19923	0.30304	0.278656	0.29493	0.28113
7	0.13436	0.11809	0.17146	0.155077	0.16974	0.16243
8	0.27104	0.22607	0.31744	0.297508	0.34913	0.34216
9	0.37778	0.35613	0.37771	0.371929	0.39039	0.39359
10	0.40440	0.06748	0.07364	0.06953	0.07259	0.07137
11	0.45201	0.19543	0.19458	0.19173	0.19331	0.19126

Table 4 Average classification accuracy for the compared algorithms on the different datasets using small, mixed and large initialization

No.	Small		Mixed		Large	
	WOA	bWOA-S	WOA	bWOA-S	WOA	bWOA-S
1	0.78597	0.69400	0.74091	0.62365	0.60900	0.61214
2	0.66008	0.77569	0.79885	0.80002	0.79352	0.80803
3	0.72842	0.85853	0.83688	0.82693	0.83083	0.82647
4	0.03378	0.06876	0.07879	0.07376	0.05314	0.07389
5	0.62664	0.68082	0.68478	0.66492	0.66067	0.68208
6	0.63837	0.73029	0.69856	0.71048	0.69529	0.71586
7	0.82579	0.82443	0.82449	0.83414	0.83011	0.83607
8	0.61415	0.64104	0.65244	0.64659	0.64370	0.65259
9	0.58876	0.59702	0.60087	0.58403	0.58991	0.59334
10	0.56964	0.93287	0.93437	0.93613	0.93066	0.93097
11	0.53258	0.80330	0.81143	0.81185	0.81026	0.81140

capability. In each run, eight well-known measures are recorded to investigate the performance of comparison algorithms.

Tables 3 and 4 outlines the statistical mean fitness values obtained from the two algorithms WOA and bWOA-S ,and average classification accuracy respectively in three initialization methods we can see that the proposed algorithm overcomes the native algorithm. For assessing the repeatability of results, Table 5 shows standard deviation to the obtained fitness function. There is a question here as to why we need more algorithms despite the many algorithms proposed so far. The answer to this question is in the no free lunch (NFL) theorem [21] that logically has proven that there is no optimization technique for solving all optimization problems. This

Table 5 Standard deviation fitness function for the compared algorithms on the different datasets averaged over the three initialization methods

No.	Standard deviation		Selection averaged		Average time	
	WOA	bWOA-S	WOA	bWOA-S	WOA	bWOA-S
1	0.03023	0.01226	4.72307	6.66245	0.57625	0.63875
2	0.05362	0.04834	7.13699	11.0659	0.79861	0.94444
3	0.03724	0.03441	2.89027	4.17675	0.69688	0.74765
4	0.20649	0.20237	10.05066	11.30218	0.65288	0.68750
5	0.06923	0.06874	3.36067	3.67051	0.66591	0.72159
6	0.06848	0.06318	3.64014	3.76398	0.64000	0.69604
7	0.03861	0.03683	4.21395	3.97656	0.59228	0.62610
8	0.06492	0.06904	3.18767	3.13704	0.49135	0.56250
9	0.08452	0.08923	3.85402	4.39146	0.52344	0.61797
10	0.03209	0.01394	74.21914	86.55317	0.71632	0.88924
11	0.06597	0.04137	596.25106	216.99145	0.73250	0.92344

theorem has motivated the rapidly increasing number of algorithms proposed over the last decade and is one of the motivations of this paper as well.

6 Conclusion and Future Work

In this paper, a new binary version of the basic whale optimization algorithm called bWOA-S to solve the FS problem was proposed. To convert the native version of WOA to a binary version, S-shaped transfer functions are employed. In order to investigate the performance of the proposed two binary algorithms, the experiments are applied on 11 benchmark datasets from UCI datasets and five evaluation criteria are performed. The experimental results revealed that the proposed algorithms have achieved superior results versus the native algorithm. Furthermore, the results proved that bWOA-S has been achieved the smallest number of features with better classification accuracy. For future work, the proposed algorithm introduced here will be used with more common classifiers such as SVM and ANN to verify the performance.

References

1. Chizi, B, Rokach L, Maimon O (2009) A survey of feature selection techniques. In: Encyclopedia of data warehousing and mining, 2nd edn, pp 1888–1895. IGI Global
2. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. J Mach Learn Res 1157–1182
3. Hamad, A, Houssein EH, Hassanien AE, Fahmy AA (2016) Feature extraction of epilepsy EEG using discrete wavelet transform. In: 2016 12th international computer engineering conference (ICENCO), pp 190–195. IEEE
4. Houssein, EH, Kilany M, Hassanien AE, Snasel V (2016) A two-stage feature extraction approach for ECG signals. In: International Afro-European conference for industrial advancement, pp 299–310. Springer, Cham
5. Girish C, Ferat S (2014) A survey on feature selection methods. Comput Electr Eng 40(1):16–28
6. Manoranjan D, Huan L (1997) Feature selection for classification. Intel Data Anal 1(1–4):131–156
7. Yang Y, Pedersen JO (1997) A comparative study on feature selection in text categorization. ICML, vol 97
8. Blum Avrim L, Pat L (1997) Selection of relevant features and examples in machine learning. Artif Intel 97(1):245–271
9. Kiansing N, Huan L (2000) Customer retention via data mining. Artif Intel Rev 14(6):569–590
10. Yong R, Huang Thomas S, Shih-Fu C (1999) Image retrieval: current techniques, promising directions, and open issues. J Visual Commun Image Represent 10(1):39–62
11. Nakamura Rodrigo YM et al (2012) BBA: a binary bat algorithm for feature selection. In: 2012 25th SIBGRAPI conference on graphics, patterns and images (SIBGRAPI). IEEE
12. Rodrigues D, Pereira L, Almeida T, Papa J, Souza A, Ramos C, Yang X-S (2013) BCS: a binary cuckoo search algorithm for feature selection. In: Proceedings of the IEEE international symposium on circuits and systems (ISCAS), pp 465–468. IEEE, Beijing
13. Douglas R et al (2015) Binary flower pollination algorithm and its application to feature selection. Recent Adv Swarm Intel Evol Comput, pp 85–100. Springer International Publishing
14. AbdEl-Fattah SS, Nabil E, Badr A (2016) A binary clonal flower pollination algorithm for feature selection. Pattern Recog Lett 77:21–27
15. Eid E et al (2015) Firefly optimization algorithm for feature selection. In: Proceedings of the 7th Balkan conference on informatics conference. ACM
16. Hafez, AI et al (2016) Sine cosine optimization algorithm for feature selection. In: International symposium on innovations in intelligent systems and applications (INISTA). IEEE
17. Eid E, Zawbaa HM, Hassanien AE (2016) Binary ant lion approaches for feature selection. Neurocomputing 213:54–65
18. Eid E, Zawbaa HM, Hassanien AE (2016) Binary grey wolf optimization approaches for feature selection. Neurocomputing 172:371–381
19. Seyedali M, Andrew L (2013) S-shaped versus V-shaped transfer functions for binary particle swarm optimization. Swarm Evol Comput 9:1–14
20. Seyedali M, Andrew L (2016) The whale optimization algorithm. Adv Eng Softw 95:51–67
21. Wolpert David H, Macready William G (1997) No free lunch theorems for optimization. IEEE Trans Evol Comput 1(1):67–82