Neurocomputing 000 (2017) 1–11

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Hybrid Whale Optimization Algorithm with simulated annealing for feature selection

Majdi M. Mafarja [a,*], Seyedali Mirjalili [b]

[a] Department of Computer Science, Birzeit University, Birzeit, Palestine
[b] School of Information and Communication Technology, Griffith University, Brisbane, 4111 QLD, Australia

ARTICLE INFO

Article history:
Received 22 September 2016
Revised 9 April 2017
Accepted 22 April 2017
Available online xxx

Keywords:
Feature selection
Hybrid optimization
Whale Optimization Algorithm
Simulated annealing
Classification
WOA
Optimization

ABSTRACT

Hybrid metaheuristics are of the most interesting recent trends in optimization and memetic algorithms. In this paper, two hybridization models are used to design different feature selection techniques based on Whale Optimization Algorithm (WOA). In the first model, Simulated Annealing (SA) algorithm is embedded in WOA algorithm, while it is used to improve the best solution found after each iteration of WOA algorithm in the second model. The goal of using SA here is to enhance the exploitation by searching the most promising regions located by WOA algorithm. The performance of the proposed approaches is evaluated on 18 standard benchmark datasets from UCI repository and compared with three well-known wrapper feature selection methods in the literature. The experimental results confirm the efficiency of the proposed approaches in improving the classification accuracy compared to other wrapper-based algorithms, which insures the ability of WOA algorithm in searching the feature space and selecting the most informative attributes for classification tasks.

1. Introduction

Data mining is one of the fastest growing research topics in the information industry in recent years due to the wide availability of the huge amounts of data and the crucial need to transform such data to useful information [1]. Data mining is an essential part in knowledge discovery (KDD) process that consists of a set of iterative sequence of tasks data preprocessing (data cleaning, data integration, data reduction, data transformation), data mining, pattern evaluation and finally knowledge presentation [1]. The preprocessing step, that precedes the data mining step in the Knowledge Discovery (KDD) process, has a large impact on the performance of the data mining techniques (*e.g.,* classification) [2] on either the quality of the extracted patterns or the running time required to analyze the complete dataset. Feature selection is one of the major preprocessing steps since it aims to eliminate the redundant, irrelevant variables within a dataset. Feature selection methods are categorized as wrappers and filters [3]. While the filter model tries to evaluate the features subsets depending on data itself using designated methods (*e.g.,* information gain (IG) and principal component analysis (PCA) [4]), wrappers utilize a learning algorithm (*e.g.,* classification algorithm) to evaluate the selected feature subset during the search process [5]. Filters usually perform faster than wrappers since they measure the information gain, distance between features, features dependency, which are computationally cheaper than measuring a classifier accuracy [3]. However, wrappers have been widely investigated for classification accuracy since they are proved to be beneficial in finding feature subsets that suite a predetermined classifier [6]. Generally, three factors should be determined when using a wrapper feature selection model: classifier (*KNN, SVM, DT, etc.*), feature subset evaluation criteria (accuracy, false-positive elimination rate, area under the ROC), and a searching technique to find the best combination of features [7].

It is worth mentioning here that there are also feature selection techniques based on matrix computations in the literature. Some of the recent and promising ones the column-subset selection problem [8–10] are known to do feature selection with provable theoretical bounds. These methods have been used to perform feature selection on *k*-means [11], SVM [12,13], Ridge Regression [14], which provide provable performance guarantees. These methods are known to outperform existing methods like mutual information, recursive feature elimination, *etc.*

Searching a (near) optimal subset from the original set is a challenging problem. In the last two decades, metaheuristics have been very reliable when solving diverse optimization problems such as engineering design, machine learning, data mining scheduling and production problems, and so on [15]. Feature selection is one of the domains where metaheuristics have been investigated.

* Corresponding author.
E-mail addresses: mmafarja@birzeit.edu, mmafarjeh@gmail.com (M.M. Mafarja).

http://dx.doi.org/10.1016/j.neucom.2017.04.053
0925-2312/

Please cite this article as: M.M. Mafarja, S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.04.053

Metaheuristics show superior performance when compared to the exact search mechanisms since the complete search tends to generate all possible solutions for the problem. For instance, if a dataset contains N features, then 2^N solutions should be generated and evaluated which requires high computational cost [16]. Random search is another search strategy for selecting features which searches the next set randomly [17]. Metaheuristics are considered better than the random search too, since they may perform as a complete search in the worst case [3,15]. Although the strategy metaheuristic does not guarantee finding the best solution in every run, it may determine an acceptable solution in a reasonable time [15]. Various metaheuristics including Tabu Search (TS) [18], Simulated Annealing (SA) [19], Record-to-Record Travel Algorithm (RRT) [20,21], Genetic Algorithm (GA) [22], Particle swarm optimization (PSO) [23], Ant Colony Optimization (ACO) [24], Differential Evolution (DE) [25], and Artificial Bee Colony (ABC) [26] have been used in the literature to search feature subset space for selecting (sub)optimal feature set [25].

Exploration of the search space (diversification) and exploitation of the best solutions found (intensification) are two contradictory criteria that must be taken into account when designing or using a metaheuristic [15]. According to these criteria, metaheuristics algorithms can be classified into two families; population-based (*e.g.*, swarm intelligence, evolutionary algorithms) algorithms that are exploration oriented and single-solution based (*e.g.*, local search and simulated annealing) algorithms that are exploitation oriented. A good balance between these two goals will enhance the performance of the searching algorithm. One choice to achieve this balance is to use a hybrid model where at least two techniques are combined to enhance the performance of each technique. The resulting algorithm is called *memetic algorithm*. In this study, our aim is to use the recently proposed Whale Optimization Algorithm (WOA) and SA to build a new hybrid (memetic) wrapper method to improve the performance of general classification tasks [27].

According to Talbi [15], one of the metaheuristic hybridization models is to combine a metaheuristic with a complementary metaheuristic. In this manner, two levels of hybridizations can be distinguished: low-level and high-level. On one hand, in the low-level hybridization, a given function in a metaheuristic (*e.g.,* crossover or mutation in GA) is replaced with another metaheuristic (*e.g.,* local search). In this model, the local optimization has the responsibility of the local search algorithm, while the global optimization is carried by the population-based algorithm [27]. On the other hand, in high-level hybrid algorithms, self-contained metaheuristics are executed in a sequence. In each level two hybridization mechanisms are possible; relay and teamwork hybridization. While in relay hybridization a set of metaheuristics is acting in a pipeline fashion where each metaheuristic uses the output of the previous algorithm, in the teamwork hybridization, many cooperating agents evolve in parallel; each agent carries out a search in a solution space [27]. In this paper, we aim to use the low-level teamwork hybrid (LTH) and the high-level relay hybrid (HRH) models for feature selection problems, where a population-based algorithm (WOA) will be hybridized with another single-solution based algorithm (SA). In other words, SA will be enhancing the exploitation in WOA algorithm. In literature, different forms of hybridization of heuristic algorithms were developed for feature selection problem, but to the best of the authors' knowledge this is the first time that a hybrid model using WOA and SA algorithms is applied to feature selection problems.

SA [28] is a single solution based metaheuristic algorithm, introduced by Kirkpatrick et al., and can be considered as a hill-climbing based method that iteratively try to improve a candidate solution with respect to the objective function. The improving move will be accepted, while the worse move is accepted with a certain probability to help the algorithm escape from the local optima. The probability of accepting a worse solution is determined by the Boltzmann probability, $P = e^{-\theta/T}$, where θ is the difference of the evaluation of the objective function between the best solution (Sol_{best}) and the trial solution (Sol_{trial}), and T is a parameter (called the temperature) that periodically decreases during the search process according to some cooling schedule.

WOA [29], proposed by Mirjalili and Lewis, is a recent optimization algorithm that mimics the intelligent foraging behavior of the humpback whales. WOA has good properties such as fewer number of parameters to control since it includes only two main internal parameters to be adjusted, easy implementation, and high flexibility. WOA algorithm smoothly transit between exploration and exploitation depending on only one parameter. In the exploration phase the position of the search agents (solutions) are updated according to a randomly selected search agent instead of the best search agent find so far. Due to the simplicity of WOA algorithm in implementation and the less dependency on parameters in addition to using a logarithmic spiral function which makes the algorithm cover the border area in the search space, many researchers in many fields become motivated to use this algorithm in solving variant optimization problems like the economic dispatch problem on IEEE 30-Bus [30], sizing optimization problems of truss and frame structures [31] and unit commitment problem solution [32]. The powerful properties of WOA and SA algorithms can be combined to produce a hybrid model to obtain better results than using each one separately. This hybridization is to enhance the exploitation property of the WOA algorithm. To enhance the exploration in the same algorithm, tournament selection mechanism is used instead the random selection.

This paper proposes a hybrid WOA and SA algorithms in a wrapper feature selection method. The proposed approach aims to enhance the exploitation of the WOA algorithm. To enhance the exploitation, SA algorithm is employed in two hybridization models; namely low-level teamwork hybrid (LTH) and high-level relay hybrid (HRH). In LTH model, simulated annealing is used as a component in the WOA algorithm. It is used to search the neighborhood of the best search agent so far to insure that it's the local optima. In the second model, SA is employed in a pipeline mode after the WOA terminates to enhance the best found solution. To preserve the diversity of the algorithm, tournament selection is employed to select search agents from the population because it gives the chance to all individuals to be selected.

The rest of this paper is organized as follows: Section 2 presents the related works. The basics of the WOA algorithm and SA are presented in Section 3. Section 4 presents the details of the proposed approach. In Section 5, the experimental results are presented and result are analysed. Finally, in Section 6, conclusions and future work are given.

2. Related works

In recent years, hybrid metaheuristics have been used by many researchers in the field of optimization [15]. Hybrid algorithms showed superior performance in solving many practical or academic problems [27]. In 2004, the first hybrid metaheuristic algorithm for feature selection was proposed [33]. In this approach, local search techniques were embedded into GA algorithm to control the search process.

In [34], Martin and Otto introduced a hybrid algorithm between Markov chain and simulated annealing, in which the Markov chain is only dedicated to explore local optima. The algorithm was designed to solve the travel salesman problem. Recently, TS and SA algorithms have been proposed to solve reactive power problem [35] and Symmetrical Traveling Salesman Problem [36]. In addition, SA has been hybridized with genetic algorithm in [37–40].

Please cite this article as: M.M. Mafarja, S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.04.053

These studies evidence that the hybrid models shown a better performance when compared with other local or global search algorithms.

In feature selection domain, many hybrid metaheuristic algorithms have been proposed with much success. Mafarja and Abdullah proposed a filter feature selection hybrid algorithm that used SA to enhance the local search capability of genetic algorithm in [41]. The algorithm was tested on eight UCI datasets and showed a good performance in terms of the number of selected attributes when compared with the state-of-the-art approaches. In [42], a hybrid GA and SA has been proposed and tested on the hand-printed Farsi characters. Another wrapper feature selection hybrid algorithm was proposed in [43], by incorporating the metropolis acceptance criterion of simulated annealing into crossover operator of GA. Again, GA was hybridized with SA to produce a hybrid wrapper feature selection algorithm to classify the power disturbance in the Power Quality (PQ) problem, and to optimize the SVM parameters for the same problem [44]. Olabiyisi et al. in 2012 [45] proposed a novel hybrid GA–SA metaheuristic algorithm for feature extraction in timetabling problem, where the GA selection process replaced by the one in SA to avoid stacking at the local optima. GA was hybridized with TS in a wrapper feature selection which used the FUZZY ARTMAP NN classifier as an evaluator [46]. Two filter feature selection memetic algorithms were proposed by Mafarja et al. [47]. In these approaches, the fuzzy logic was employed to control the main parameters in two local search algorithms (record to record and great deluge) who combined with GA later.

Moradi and Gholampour [48] proposed a local search algorithm to guide the search process in PSO algorithm to select the minimal reducts based on their correlation information. Moreover, in [49], GA and PSO algorithms were proposed in a hybridized method with SVM classifier called GPSO and were applied to microarray data classification. A multi objective PSO with a hybrid mutation operator is proposed in the same field in [50]. In [51], a new combination between GA and PSO was proposed to optimize the feature set for Digital Mammogram datasets. Two hybrid wrapper feature selection algorithm based on the combination between ACO and GA [52,53]. In the same manner, ACO was combined with Cuckoo Search (CS) algorithm in [54]. Also, a new combination between Harmony Search Algorithm (HSA) and a Stochastic Local Search (SLS) was proposed for a wrapper feature selection method in [55]. Artificial Bee Colony (ABC) has been recently combined with differential evolution algorithm (DE) to propose a new wrapper feature selection method [25]. For further reading about metaheuristics and feature selection, interested readers are referred to the survey papers in [56,57].

Despite the merits of the above-mentioned memetic algorithms for feature selection, one might ask if we need any other new memetic. There is a theorem in the field of optimization called No-Free-Lunch (NFL) that logically proves: there is no algorithm for solving all optimization problems. In the scope of this work, this can be stated as: none of the heuristic wrapper feature selection is able to solve *all* feature selection problems. In other words, there is always room for the improvements of the current techniques to better solve the current of new feature selection problems. This motivated our attempts to propose yet another memetic algorithm for feature selection in the next section.

3. Methods

3.1. Whale optimization algorithm

WOA is a new metaheuristic algorithm proposed by Mirjalili and Lewis [29] and mimics the foraging of humpback whales. The humpback whales hunt school of krill or small fishes close to the surface by swimming around them within a shrinking circle and creating distinctive bubbles along a circle or '9'-shaped path (see Fig. 1). Encircling prey and spiral bubble-net attacking method were represented in the first phase of the algorithm; exploitation phase, the second phase where search randomly for a prey (exploration phase). The following subsections discuss the mathematical model of each phase in details. Note that in the equations, a uniform distribution will be used to generate random numbers.

3.1.1. Exploitation phase (encircling prey/bubble-net attacking method)

To hunt a prey, humpback whales first encircle it. Eqs. (1) and (2) can be used to mathematically model this behavior [29].

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

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

where t indicates the current iteration, X^* represents the best solution obtained so far, X is the position vector, $|\ |$ is the absolute value, and $\cdot$ is an element-by-element multiplication. In addition, A and C are coefficient vectors that are calculated as in Eqs. (3) and (4), respectively:

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

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

where a decreases linearly from 2 to 0 over the course of iterations (in both exploration and exploitation phases) and r is a random vector generated with uniform distribution in the interval of [0,1]. According to Eq. (2) the search agents (whales) update their positions according to the position of the best known solution (prey). The adjustment of the values of A and C vectors control the areas where a whale can be located in the neighborhood of the prey.

The Shrinking encircling behavior is achieved by decreasing the value of a in Eq. (3) according to Eq. (5).

$$a = 2 - t\frac{2}{MaxIter} \tag{5}$$

where t is the iteration number and *MaxIter* is the maximum number of allowed iterations. To simulate the spiral-shaped path, the distance between a search agent (X) and the best known search agent so far (X^*) is calculated (see Fig. 1), then a spiral equation is used to create the position of the neighbor search agent as in Eq. (6).

$$\vec{X}(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X^*}(t) \tag{6}$$

where $D' = |\vec{X^*}(t) - \vec{X}(t)|$ and indicates the distance of the ith whale and the prey (best solution obtained so far), b is a constant for defining the shape of the logarithmic spiral, and l is a random number in $[-1,1]$.

To model the two mechanisms, shrinking encircling and the spiral-shaped path, a probability of 50% is assumed to choose between them during the optimization process as in Eq. (7).

$$\vec{X}(t+1) = \begin{cases} \text{Shrinking Encircling } (Eq.(2)) & \text{if}(p < 0.5) \\ \text{spiral}-\text{shaped path } (Eq.(6)) & \text{if}(p \geq 0.5) \end{cases} \tag{7}$$

where p is a random number in [0,1].

3.1.2. Exploration phase (search for prey)

In order to enhance the exploration in WOA, instead of updating the positions of the search agents according to the position of the best one so far, a random search agent is selected to guide the search. So, a vector A with the random values greater than 1 or less than -1 is used to force search agent to move far away from the

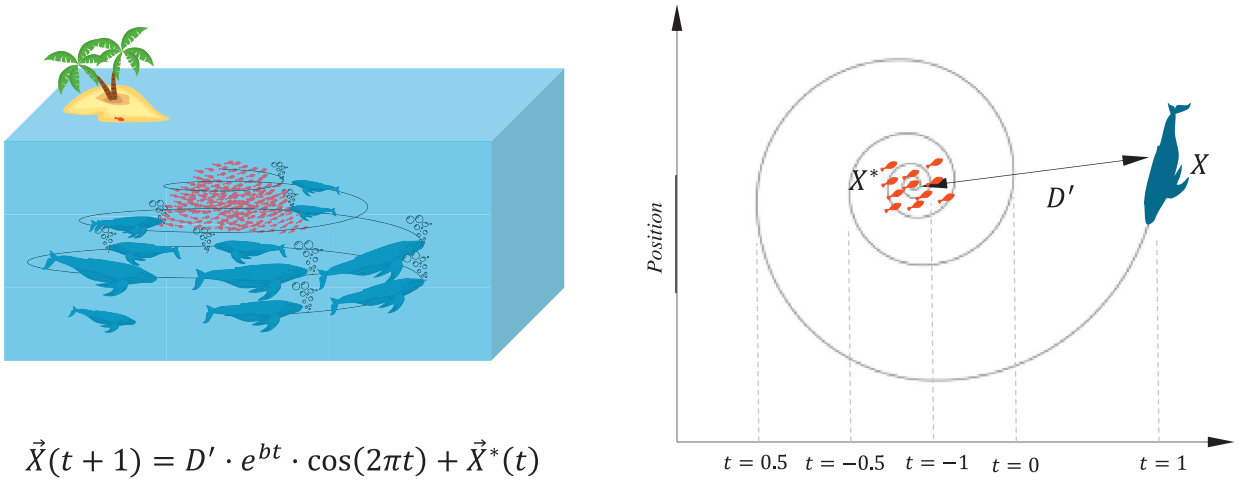

Fig. 1. Unique bubble-net feeding methods of humpback whales and the mathematical model.

Algorithm 1
Pseudo-code of the native WOA algorithm.

```
Generate Initial Population X_i (i = 1, 2, ..., n)
Calculate the fitness of each solution
X* = the best search agent
while (t < Max_Iteration)
    for each solution
        Update a, A, C, l, and p
        if 1 (p < 0.5)
            if 2 (|A| < +1)
                Update the position of the current solution by Eq. (2)
            else if 2 (|A| > +1)
                Select a random search agent ()
                Update the position of the current search agent by the Eq. (9)
            end if 2
        else if 1 (p ≥ 0.5)
            Update the position of the current search by the Eq. (6)
        end if 1
    end for
Check if any solution goes beyond the search space and amend it
Calculate the fitness of each solution
Update X* if there is a better solution t = t + 1
end while
return X*
```

best known search agent. This mechanism can be mathematically modeled as in Eqs. (8) and (9).

$$\vec{D} = \left| \vec{C} . \overrightarrow{X_{\text{rand}}} - \vec{X} \right| \tag{8}$$

$$\vec{X}(t+1) = \overrightarrow{X_{\text{rand}}} - \vec{A} . \vec{D} \tag{9}$$

where $\overrightarrow{X_{\text{rand}}}$ is a random whale chosen from the current population.

Algorithm 1 shows the pseudo code of WOA algorithm. It may be seen that WOA creates a random, initial population and evaluates it using a fitness function once the optimization process starts. After finding the best solution, the algorithm repeatedly executes the following steps until the satisfaction of an end criterion. Firstly, the main coefficients are updated. Secondly, a random value is generated. Based on this random value, the algorithm updates the position of a solution using either Eqs. (2)/(9) or Eq. (6). Thirdly, the solutions are prevented from going outside the search landscape. Finally, the algorithm returns the best solution obtained as an approximation of the global optimum.

WOA is a population-based stochastic algorithm as mentioned above. What guarantees the convergence of this algorithm is the use of the best solution obtained so far to update the position of the rest of solutions. However, this mechanism might lead solutions to local optima. This is the reason of using random variables to switch between three equations to update the position of solutions. The balance between local optima avoidance (exploration) and convergence (exploitation) is done with the adaptive parameter *a*. This parameter smoothly reduces the magnitude of changes in the solutions and promotes convergence/exploitation proportional to the number of iterations.

3.2. Simulated annealing

Simulated annealing, proposed by Kirkpatrick et al. [28], is a single-solution metaheuristic algorithm based on the hill climbing method that. To overcome the problem of stagnating in local optima, SA utilizes a certain probability to accept a worse solution. The algorithm starts with a randomly generated solution (*initial solution*); in each iteration, a neighbor solution to the best one so far is generated according to a predefined neighborhood structure and evaluated using a fitness function. The improving move (the neighbor is fitter than the original solution) is always accepted, whilst worse neighbor is accepted with a certain probability determined by the Boltzmann probability, $P = e^{-\theta/T}$ where θ is the difference between the fitness of the best solution (*BestSol*) and the generated neighbor (*TrialSol*). Moreover, T is a parameter (called the temperature) which periodically decreases during the search process according to some cooling schedule. In this work, the initial temperature is set to $2^*|N|$, where $|N|$ represents the number of attributes for each dataset, and the cooling schedule is calculated as $T = 0.93^* T$ (as adopted in [58]). Algorithm 2 shows the pseudo code of SA algorithm.

3.3. Tournament selection

Tournament selection is a simple and easy to implement selection mechanism that was stated by Goldberg et al. [59]. It is one of the most well-regarded selection mechanisms used in evolutionary algorithms [60]. In tournament selection, n solutions are to be selected randomly from the population, these solutions are compared against each other and a tournament will be placed to determine the winner. The tournament involves generating a random number between 0 and 1, then it is compared with a selection probability that provides a convenient mechanism for adjusting the selection pressure (usually set to 0.5), if the random is greater, then the solution with the highest fitness value will be selected and, otherwise, the weak solution is selected. This property in tournament

Please cite this article as: M.M. Mafarja, S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.04.053

Algorithm 2
Pseudocode of SA algorithm.

```
T0 = 2*|N| where |N| is the number of attributes for each dataset
BestSol ← Si'
δ (BestSol) ← δ (Si) // δ indicates the quality of the solution
while T > T0
    generate at random a new solution TrialSol in the neighbor of Si'
    calculate δ(TrialSol)
    if (δ(TrialSol) > δ(BestSol))
        Si' ← TrialSol;
        BestSol ← TrialSol;
        δ(Si') ← δ(TrialSol);
        δ(BestSol) ← δ(TrialSol);
    else if ((δ(TrialSol) = δ(BestSol))
        Calculate |TrialSol| and | BestSol |;
        if (|TrialSol| < | BestSol |)
            Si' ← TrialSol;
            BestSol ← TrialSol;
            δ(Si') ← δ(TrialSol);
            δ(BestSol) ← δ(TrialSol);
        end if
    else // accepting the worse solution
        Calculate θ = δ(TrialSol) – δ(BestSol))
        Generate a random number, P = [0,1];
        if (P ≤ e^(−θ/T))
            Si' ← TrialSol; δ(Si) ← δ(TrialSol);
        end if
    end if
    T = 0.93 * T; // update temperature
end while
Output BestSol
```

selection gives the chance to most solutions to be selected which preserves the diversity of the selected solutions [60].

4. The proposed approach

Feature selection is a binary optimization problem, where solutions are restricted to the binary {0, 1} values. For the WOA algorithm to be used with the feature selection problem, a binary version should be developed. In this work, a solution is represented in one dimensional vector, where the length of the vector is based on the number of attributes of the original dataset. Each value in the vector (cell) is represented by "1" or "0". Value "1" shows that the corresponding attribute is selected; otherwise the value is set to "0".

Feature selection can be considered as a multi-objective optimization problem where two contradictory objectives are to be achieved; minimal number of selected features and higher classification accuracy. The smaller is the number of features in the solution and the higher the classification accuracy, the better the solution is. Each solution is evaluated according to the proposed fitness function, which depends on the KNN classifier [61] to get the classification accuracy of the solution and on the number of selected features in the solution. In order to balance between the number of selected features in each solution (minimum) and the classification accuracy (maximum), the fitness function in Eq. (10) is used in both WOA and SA algorithms to evaluate search agents.

$$Fitness = \alpha\gamma_R(D) + \beta\frac{|R|}{|N|}, \tag{10}$$

where $\gamma_R(D)$ represents the classification error rate of a given classier (the *K*-nearest neighbor (KNN) classifier is used here). Furthermore, $|R|$ is the cardinality of the selected subset and $|N|$ is the total number of features in the dataset, α and β are two parameters corresponding to the importance of classification quality and subset length, $\alpha \in [0, 1]$ and $\beta = (1-\alpha)$ adopted from [62].

As can be noted in the WOA algorithm, exploitation (as in Eqs. (2) and (6)) depends on calculating the distance between the search agent and the best known whale so far. We believe that employing an efficient local search algorithm to search the neighborhood around the best known solution will improve the results. Moreover, since the exploration in the native WOA algorithm (as in Eq. (9)) depends on changing the position of each search agent according to a randomly selected solution we believe that using a different selection mechanism like tournament selection may improve the exploration ability within the algorithm. This is simply because the tournament selection gives more chance to the weak solutions to be selected during the search process depending on the selection pressure which improves the diversity capability of WOA algorithm.

4.1. Hybrid WOA–SA methods

WOA algorithm is a recent optimization algorithm that shows superior results in many optimization problems. The native algorithm uses a blind operator to play the role of exploitation regardless of the fitness value of the current solution and the operated one. We replaced this operator with a local search which considers a solution as its initial state, work on it, and replace the original solution by the enhanced one. This approach represents a hybridization between global search (WOA) and local search algorithm (SA). In the proposed approach, two hybridization models between the two algorithms are considered, (1) *Low-Level Teamwork Hybrid (LTH) and* (2) *High-Level Relay Hybrid (HRH).* On one hand, in LTH, SA algorithm is embedded in WOA algorithm to search for the best solution in the neighbour of both the randomly selected solution (to replace Eq. (9)) and the neighbour of the best known solution (to replace Eq. (2)) and replace the original one. This approach is called WOASA-1. This process improves the exploitation ability of WOA algorithm. SA algorithm in this approach acts as an operator in WOA algorithm. On the other hand, HRH model uses SA algorithm after applying WOA algorithm and finding the best solution, then SA is used to enhance that final solution. This approach is called WOASA-2. In both WOASA-1 and WOASA-2, WOA uses random selection mechanism to select the random solution that enables the algorithm to explore the feature space. Moreover, the two approaches are tested when WOA uses the Tournament Selection mechanism. The two techniques are named WOASAT-1 and WOASAT-2 respectively.

5. Experiments

5.1. Datasets

The implementation of the proposed algorithm is done using Matlab. In order to assess the performance of the proposed approaches, the experiments are performed on 18 FS benchmark datasets from the UCI data repository [63]. Table 1 shows the details of the used datasets such as number of attributes and instances in each dataset.

5.2. Parameter settings

A wrapper approach-based on the KNN classifier (where $K = 5$ [62]) with the Euclidean distance matric is used to generate the best reduct. In the proposed approach, each dataset is divided in cross validation in a same manner to that in [64] for evaluation. In *K*-fold cross-validation, $K-1$ folds are used for training and validation and the remaining fold is used for testing. This process is repeated *M* times. Hence, individual optimizer is evaluated $K*M$ times for each data set. The data for training, validation are equally sized. In all experiments in this section, the parameters are set as follows. The maximum number of iterations is 100 and the population size is 10. Furthermore, each algorithm is run 5 times with random seed on an Intel Core i5 machine, 2.2 GHz CPU and 4GB of

Please cite this article as: M.M. Mafarja, S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.04.053

Table 1
List of Datasets datasets Used used in the Experimentsexperiments.

	Dataset	No. of attributes	No. of objects
1.	Breastcancer	9	699
2.	BreastEW	30	569
3.	CongressEW	16	435
4.	Exactly	13	1000
5.	Exactly2	13	1000
6.	HeartEW	13	270
7.	IonosphereEW	34	351
8.	KrvskpEW	36	3196
9.	Lymphography	18	148
10.	M-of-n	13	1000
11.	PenglungEW	325	73
12.	SonarEW	60	208
13.	SpectEW	22	267
14.	Tic-tac-toe	9	958
15.	Vote	16	300
16.	WaveformEW	40	5000
17.	WineEW	13	178
18.	Zoo	16	101

RAM. Please note that the parameters of SA are identical to those used in the previous subsection.

5.3. Results and discussion

All results obtained from the proposed approaches are reported in this section. The native WOA, WOASA-1 and WOASA-2 are compared to assess the effect of hybridizing SA algorithm with the native WOA. Then the three approaches that use the Tournament Selection (WOAT, WOASAT-1, and WOASA-2) are compared to assess the effect of using Tournament selection instead of the random mechanism. To find out the best approach among the proposed approaches, all approaches are compared together in one table. The proposed approaches are also compared to state-of-the-art feature selection methods including Ant Lion Optimizer (ALO), PSO and GA [62,65] based on the following criteria:

- Classification accuracy by using the selected features on the test dataset. The average accuracy gained from 5 runs is calculated.
- The average selection size is the second comparison that is presented in this section.
- Then the fitness values obtained from each approach is reported, the mean, min and max fitness values are compared.

5.3.1. Comparison of WOA, WOASA-1 and WOASA-2

The performance of WOA, WOASA-1 and WOASA-2 over the two objectives (average selection size and classification accuracy) in addition to the computational time is outlined in this section. It is worth to remind that WOASA-1 embedded SA algorithms in WOA algorithm to act as an internal operator, while in WOASA-2, SA is employed on the final solution after WOA algorithm terminated.

Inspecting Table 2, it is evident that the hybrid algorithms are much better than the native one for both objective: classification accuracy and number of selected attributes. The native algorithm does not outperform any hybrid approach over all datasets in terms of number of selected attributes, whereas it performs better than others on only two datasets with an insignificant difference. WOASA-1 outperforms all approaches in 8 and 10 datasets in both classification accuracy and number of selected attributes respectively. For the classification accuracy, it outperforms WOA over sixteen datasets and the difference between the results of the two approaches varies from 0.03% to 23%. In Exactly dataset, WOASA-1 provides 100% accuracy by using 6 attributes only, at the same time WOA show 77.2% accuracy and 9.2 attributes. The rest of results in Table 2 show that WOASA-1 outperforms WOA in both classification accuracy and selected attributes for the same dataset. A similar pattern can be seen in the results of WOASA-2, it performs better than WOA on almost all datasets. When comparing WOASA-1 and WOASA-2, it may be seen that WOASA-2 performs better than WOASA-1 over 8 datasets for classification accuracy. At the same time WOASA-1 obtained better higher classification accuracy than WOASA-2 over 9 datasets. In terms of number of selected attributes, WOASA-2 outperforms WOASA-1 over 8 datasets. By contrast, WOASA-1 performs better than WOASA-2 on six datasets.

Generally speaking, the performance of both proposed approaches is very close over both objectives and both of them perform better than the native algorithm. From the reported results we can conclude that the two hybrid models (LTH and HRH) significantly enhanced the performance of the native algorithm, but there is no significant discrepancy between the two models proposed. This motivated us to keep the proposed approaches and study the influence of enhancing the exploration property by employing tournament selection mechanism to make more balance between exploitation and exploration. The results of employing tournament selection are represented in Table 3.

Table 2
Classification accuracy and average selected attributes obtained from WOA, WOASA-1 and WOASA-2.

Dataset	Accuracy			Attributes		
	WOA	WOASA-1	WOASA-2	WOA	WOASA-1	WOASA-2
Breastcancer	0.96	**0.97**	0.96	6.4	5.6	**5.2**
BreastEW	0.93	0.96	**0.97**	23.8	13.6	**12.6**
CongressEW	0.93	**0.97**	**0.97**	10	**4.4**	5.2
Exactly	0.77	**1.00**	**1.00**	9.2	**6**	**6**
Exactly2	**0.74**	0.73	0.72	4.8	**1**	1.4
HeartEW	0.79	0.79	**0.84**	9.4	**6.2**	7.2
IonosphereEW	0.87	0.92	**0.96**	22.4	**11.4**	11.8
KrvskpEW	0.93	**0.98**	**0.98**	24.2	19.4	**17**
Lymphography	0.78	**0.90**	0.87	10.8	**6.8**	7.6
M-of-n	0.91	**1.00**	**1.00**	8.6	**6**	**6**
PenglungEW	0.84	0.85	**0.91**	188.4	138	**128.8**
SonarEW	0.86	0.94	**0.95**	46.4	26.6	**26.4**
SpectEW	0.81	0.82	**0.84**	**9.4**	9.6	**9.4**
Tic-tac-toe	0.76	**0.79**	0.76	8.4	**5.8**	**5.8**
Vote	0.92	**0.97**	0.96	9.4	**3.8**	5.8
WaveformEW	**0.71**	0.69	0.68	33.6	21.6	**19.4**
WineEW	0.95	0.99	**0.99**	7.4	**6.8**	**6.8**
Zoo	0.96	**0.99**	0.97	8.8	5.8	**5.4**

Please cite this article as: M.M. Mafarja, S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.04.053

Table 3
Classification accuracy and average selected attributes obtained from WOAT, WOASAT-1 and WOASAT-2.

Dataset	Accuracy			Attributes		
	WOA	WOASAT-1	WOASAT-2	WOA	WOASAT-1	WOASAT-2
Breastcancer	0.96	0.96	**0.97**	6.4	**4.0**	4.2
BreastEW	0.95	0.97	**0.98**	21.2	**11.2**	11.6
CongressEW	0.94	**0.98**	**0.98**	7.2	**4.0**	6.4
Exactly	0.78	**1.00**	**1.00**	10.2	**6.0**	**6.0**
Exactly2	0.70	0.72	**0.75**	5.4	**1.8**	2.8
HeartEW	0.80	0.80	**0.85**	9.2	5.8	**5.4**
IonosphereEW	0.88	0.90	**0.96**	17.6	**10.2**	12.8
KrvskpEW	0.93	**0.98**	**0.98**	28.0	**16.8**	18.4
Lymphography	0.79	0.85	**0.89**	13.6	8.0	**7.2**
M-of-n	0.88	**1.00**	**1.00**	10.4	**6.0**	**6.0**
PenglungEW	0.83	0.90	**0.94**	173.8	144.8	**127.4**
SonarEW	0.87	0.95	**0.97**	43.0	28.8	**26.4**
SpectEW	0.79	0.83	**0.88**	12.8	**8.0**	9.4
Tic-tac-toe	0.75	0.77	**0.79**	6.8	**6.0**	**6.0**
Vote	0.93	0.96	**0.97**	9.8	6.0	**5.2**
WaveformEW	0.72	0.67	**0.76**	34.6	21.2	**20.6**
WineEW	0.97	0.99	**0.99**	9.6	**6.0**	6.4
Zoo	0.94	0.95	**0.97**	9.2	**5.0**	5.6

Table 4
Average computational time (in seconds) for random selection based approaches and the tournament selection based approaches.

	Random selection based approaches			Tournament selection based approaches		
Dataset	WOA	WOASA-1	WOASA-2	WOAT	WOASAT-1	WOASAT-2
Breastcancer	**2.77**	15.10	43.24	2.80	14.06	41.74
BreastEW	3.26	14.04	43.18	**3.22**	15.54	44.30
CongressEW	**2.24**	14.10	37.34	2.25	14.10	35.67
Exactly	**4.50**	25.30	57.39	4.67	20.52	51.79
Exactly2	4.25	24.90	55.85	**3.67**	20.41	54.88
HeartEW	1.87	10.85	28.67	**1.86**	11.15	29.79
IonosphereEW	2.23	12.86	30.79	**2.10**	11.51	30.84
KrvskpEW	**57.53**	235.37	641.01	60.49	209.67	589.56
Lymphography	1.68	10.48	27.15	**1.66**	10.87	26.17
M-of-n	**4.43**	20.62	52.32	4.64	20.51	51.54
PenglungEW	2.24	11.07	28.47	**2.10**	10.12	30.49
SonarEW	2.02	10.00	28.02	**1.98**	10.80	27.76
SpectEW	**1.86**	12.52	30.34	**1.86**	10.58	31.38
Tic-tac-toe	5.08	19.16	51.93	**4.25**	23.65	56.89
Vote	**1.90**	10.72	30.63	1.96	12.76	30.79
WaveformEW	180.89	615.35	1770.48	**173.72**	617.10	1633.27
WineEW	**1.77**	10.33	29.59	**1.77**	10.86	26.33
Zoo	1.75	9.56	27.55	**1.71**	11.52	27.02

Table 5
Comparison between the proposed approaches and the state-of-the-art approaches in terms of classification accuracy.

Dataset	WOASAT-2	ALO	GA	PSO	Full
Breastcancer	**0.97**	0.96	0.96	0.95	0.94
BreastEW	**0.98**	0.93	0.94	0.94	0.96
CongressEW	**0.98**	0.93	0.94	0.94	0.92
Exactly	**1.00**	0.66	0.67	0.68	0.67
Exactly2	0.75	0.75	**0.76**	0.75	0.74
HeartEW	**0.85**	0.83	0.82	0.78	0.82
IonosphereEW	**0.96**	0.87	0.83	0.84	0.87
KrvskpEW	**0.98**	0.96	0.92	0.94	0.92
Lymphography	**0.89**	0.79	0.71	0.69	0.68
M-of-n	**1.00**	0.86	0.93	0.86	0.85
PenglungEW	**0.94**	0.63	0.70	0.72	0.66
SonarEW	**0.97**	0.74	0.73	0.74	0.62
SpectEW	**0.88**	0.80	0.78	0.77	0.83
Tic-tac-toe	**0.79**	0.73	0.71	0.73	0.72
Vote	**0.97**	0.92	0.89	0.89	0.88
WaveformEW	0.76	**0.77**	**0.77**	0.76	**0.77**
WineEW	**0.99**	0.91	0.93	0.95	0.93
Zoo	**0.97**	0.91	0.88	0.83	0.79
Average	**0.92**	0.83	0.83	0.82	0.81

5.3.2. Comparison of WOAT, WOASAT-1 and WOASAT-2

The performance of the TS-based approaches is presented in Table 3. In terms of classification accuracy, one can easily observe that WOASAT-2 shows superior performance since it provides the highest classification accuracy on all datasets. At the same time, its performance for the second objective, number of selected attributes, is competitive on some datasets. In addition, it is worth noting that the WOA algorithm does not outperform the hybrid approaches over any dataset.

5.3.3. Comparison between all proposed approaches

This subsection compares the six proposed approaches to analyze the impact of enhancing the exploration, exploitation and both of them combined. By analyzing the results in Tables 2 and 3, we can say that using TS plays a complementary role in enhancing exploration in WOA algorithm besides SA which enhances exploitation. The performance of the proposed approaches is enhanced gradually: WOA < WOAT < WOASA-1 < WOASAT-1 < WOASA-2 < WOASAT-2. It can be stated thatWOA is a robust algorithm that balances exploration and exploitation when searching for the global optimum, and SA enhances the exploitation in WOA algorithm Then Tournament Selection employment enhanced

Please cite this article as: M.M. Mafarja, S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.04.053

Table 6
Comparison between the proposed approaches and the state-of-the-art approaches in terms of average number of selected attributes.

Dataset	Attributes	Instances	WOASAT-2	ALO	GA	PSO
Breastcancer	9	699	**4.20**	6.28	5.09	5.72
BreastEW	30	569	**11.60**	16.08	16.35	16.56
CongressEW	16	435	**6.40**	6.98	6.62	6.83
Exactly	13	1000	**6.00**	6.62	10.82	9.75
Exactly2	13	1000	**2.80**	10.70	6.18	6.18
HeartEW	13	270	**5.40**	10.31	9.49	7.94
IonosphereEW	34	351	12.80	**9.42**	17.31	19.18
KrvskpEW	36	3196	**18.40**	24.70	22.43	20.81
Lymphography	18	148	**7.20**	11.05	11.05	8.98
M-of-n	13	1000	**6.00**	11.08	6.83	9.04
PenglungEW	325	73	**127.40**	164.13	177.13	178.75
SonarEW	60	208	**26.40**	37.92	33.30	31.20
SpectEW	22	267	**9.40**	16.15	11.75	12.50
Tic-tac-toe	9	958	**6.00**	6.99	6.85	6.61
Vote	16	300	**5.20**	9.52	6.62	8.80
WaveformEW	40	5000	**20.60**	35.72	25.28	22.72
WineEW	13	178	**6.40**	10.70	8.63	8.36
Zoo	16	101	**5.60**	13.97	10.11	9.74
Average			**15.99**	22.68	21.77	21.65

the exploration in WOA algorithm which complemented the role of SA, so we observed that WOASAT approaches outperformed all other approaches on almost all datasets in classification accuracy and produced better results in terms of number of selected attributes on many datasets. These findings prove that WOASAT properly and efficiently balances exploration and exploitation, which is due to the employed local search and selection mechanism.

It seems that using SA after WOA algorithm performs better than embedding SA in WOA. This shows the capability of WOA in locating the high-performance regions in the feature space of the problem in hand. It can also be stated that a sequential hybrid WOA and SA does not damage the exploration of WOA, which is essential when solving challenging problems.

Table 4 presents the average computational time (in seconds) required by each approach to get (near) optimal solution. All approaches use the same parameter settings and are tested on the same datasets, so we used the computational time to compare between the performances of the proposed approaches. As can be seen in Table 4, WOAT has the best computational time in comparison to other approaches on twelve datasets. WOA also performs better than other approaches on six datasets. When comparing the two hybrid models (LTH and HRH), we found that HRH based approaches (WOASA-1 and WOASAT-1) require nearly 38% of the total time required by LTH based approaches (WOASA-2 and WOASAT-2) to find the best results. This can be interpreted easily since the number of times of running the embedded SA algorithm in LTH approaches is much higher than that for the HRH approaches. Although HRH-based approaches require only one third of the time of LTH-based approaches, they significantly outperform that approaches in terms of classification accuracy and minimal reducts.

5.3.4. Comparison with the state-of-the-art approaches

In the previous section, after analysing the results we found that WOASAT-2 approach outperforms other approaches in terms on classification accuracy on all datasets and has competitive performance compared to other approaches in terms of number of selected attributes. In this section, we compare the performance of the best approach among the proposed approaches against the most related approaches in the feature selection literature. Table 5 presents the results of WOASAT-2, ALO, GA, and PSO. To show the importance of using feature selection in classification problem, the classification accuracy using full feature set is reported as well.

As per the results in Table 5, the accuracy of using full feature set is worse than selecting features using the wrapper approaches proposed. Moreover, WOASAT-2 outperforms all approaches on all datasets, except for two datasets where GA performs better than other approaches with a slight difference from WOASAT-2, and WOASAT-2 comes in the second place. In Table 6, the average of selected attributes using WOASAT-2 and other approaches are reported. WOASAT-2 shows much better performance than other approaches on 90% of the datasets where WOASAT-2 produces average accuracy of 92.2%, while the closest approach produces accuracy of 82.9%. This can be interpreted by the enhanced exploration and exploitation capability of WOASAT-2 searches intensively the high-performance regions of the feature space.

Table 6 shows the average number of selected attributes. WOASAT-2 performs better than other optimizers employed in this paper. Inspecting the results in Tables 5 and 6, a substantial discrepancy might be seen in the classification accuracy and the number of the elected attributes when comparing WOASAT-2 and other approaches. For example, in exactly dataset, WOASAT-2 is better than other approaches with nearly 34% in classification accuracy and only 6 selected features. The significant superiority of WOASAT-2 in selecting less number of attributes is justifiable since it uses the TS mechanism that gives more chances to the weak solutions to be selected, which decreases the probability of trapping in local optima. This mechanism allows the algorithm to extensively explore regions in the feature space and use SA to intensify these regions. This approach leverages the strengths of using a global search algorithm, which is efficient in exploration, and a local search, which is efficient in exploitation, to solve feature selection problems. Selecting less number of attributes means eliminating the irrelevant/redundant attributes in a dataset and reduces the search space.

The obtained statistical measures on the different runs of the optimizers on all the data sets are summarized in Table 7. Here, WOASAT-2 approach is used in comparison with other approaches. As may be observed in this table, we can see that WOASAT-2 outperforms ALO, PSO and GA in Mean and Best Fitness criteria on fifteen datasets, and it is not worse than any other approach on fifteen datasets. The high-performance of the enhanced WOA based approaches improves its capability.

This high performance of the WOA algorithm proves its capability to balance between the exploration and exploitation throughout iterations of the optimization. As per the results obtained, the performance of the WOA algorithm is proved on the large datasets

Please cite this article as: M.M. Mafarja, S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.04.053

Table 7
Mean, best and worst fitness values obtained from the different optimizers.

Dataset	Mean				Best				Worst			
	WOASAT-2	ALO	GA	PSO	WOASAT-2	ALO	GA	PSO	WOASAT-2	ALO	GA	PSO
Breastcancer	0.04	**0.02**	0.03	0.03	0.03	**0.02**	**0.02**	0.03	**0.04**	0.03	0.04	0.03
BreastEW	**0.03**	0.03	0.04	0.03	**0.02**	0.03	0.02	0.02	0.04	0.04	0.05	**0.05**
CongressEW	**0.03**	0.05	0.04	0.04	**0.02**	0.03	0.03	0.03	0.05	**0.06**	0.06	0.04
Exactly	**0.01**	0.29	0.28	0.28	**0.01**	0.28	0.27	0.21	0.01	0.29	0.31	**0.32**
Exactly2	0.25	**0.24**	0.25	0.25	0.23	0.23	**0.22**	**0.22**	0.27	0.25	0.30	**0.31**
HeartEW	0.16	**0.12**	0.14	0.15	0.13	**0.11**	0.12	0.13	**0.18**	0.13	0.14	0.18
IonosphereEW	**0.04**	0.11	0.13	0.14	**0.03**	0.10	0.09	0.12	0.05	0.12	0.16	**0.17**
KrvskpEW	**0.02**	0.05	0.07	0.05	**0.02**	0.03	0.03	0.03	0.02	0.07	**0.13**	0.07
Lymphography	**0.11**	0.14	0.17	0.19	0.09	**0.08**	0.12	0.14	0.14	0.16	**0.27**	**0.27**
M-of-n	**0.01**	0.11	0.08	0.11	**0.01**	0.09	0.02	0.06	0.01	0.12	0.15	**0.16**
PenglungEW	**0.06**	0.14	0.22	0.22	0.03	**0.00**	0.13	0.13	0.11	0.21	**0.29**	**0.29**
SonarEW	**0.03**	0.18	0.13	0.13	**0.01**	0.13	0.07	0.07	0.05	**0.26**	0.23	0.22
SpectEW	0.13	**0.12**	0.14	0.13	0.11	**0.09**	0.12	0.10	0.15	0.15	0.15	**0.16**
Tic-tac-toe	**0.21**	0.22	0.24	0.24	**0.20**	0.20	0.21	0.21	0.23	0.24	0.26	**0.27**
Vote	**0.04**	0.04	0.05	0.05	**0.02**	0.03	0.03	0.03	0.04	0.05	**0.08**	**0.08**
WaveformEW	0.25	0.21	**0.20**	0.22	0.23	0.19	**0.19**	0.21	**0.26**	0.22	0.21	0.23
WineEW	**0.01**	0.02	0.01	0.02	0.00	**0.00**	**0.00**	**0.00**	0.03	**0.03**	**0.03**	**0.03**
Zoo	**0.04**	0.07	0.08	0.10	0.00	0.04	**0.00**	0.03	0.10	0.12	0.18	**0.21**
Total	**1.83**	2.15	2.30	2.39	**1.57**	1.70	1.69	1.77	2.14	2.54	3.03	**3.08**

as well as small size data sets. The three datasets; penglungEW, krvskpEW and ionosphereEW are relatively large datasets and the fitness value of the proposed approach is clearly less than all other approaches. We can also see that WOA performs better than ALO, PSO and GA in terms of the best and worst obtained solution.

6. Conclusion

Feature selection is one of the key factors in enhancing the classifier abilities in the classification problem. In this paper four variant hybrid metaheuristic algorithms based on WOA algorithm were proposed. The proposed approaches integrate SA algorithm with the global search of WOA. SA was employed in the proposed approaches following two hybrid models; low-level teamwork hybrid model (LTH), and high-level relay hybrid model (HRH).

In LTH, SA was used as a local search operator around the selected search agents in two approaches (WOASA-1 and WOASAT-1). By contrast, SA was used to search the neighborhood of the best found solution after each iteration of WOA in HRH (WOASA-2 and WOASAT-2). To give more chances to the weak solutions to be selected in order to enhance the diversity of WOA, Tournament Selection mechanism was used to selected the search agents instead of random selection mechanism (WOASAT-1 and WOASAT-2).

The performances of the proposed approaches were assessed and compared against three recent wrapper feature selection methods including ALO, PSO, and GA. Two criteria were reported to evaluate each approach: classification accuracy, average selection size. The proposed approaches were compared against each other and the native WOA algorithm. We found that WOASAT-2, which used SA to intensify the neighboring region of the best solution found in each iteration of WOA, and the tournament selection to select the search agents show the best performance among all proposed approaches in terms of classification accuracy They also show competitive results over the second objective; selecting the minimal reducts. Since the performance of the proposed approaches follows this order: WOA < WOAT < WOASA-1 < WOASAT-1 < WOASA-2 < WOASAT-2, we can conclude that the proposed approaches managed to successfully balance the main two objectives of a metaheuristic algorithm: exploration and exploitation.

For future studies, it would be interesting to hybridize WOA algorithm with another population-based metaheuristic algorithm like ALO. In addition, employing a hybrid WOA algorithm for solving real world problem like medical data is another possible future work.

References

[1] J. Han, J. Pei, M. Kamber, Data Mining: Concepts and Techniques, Elsevier, 2011.
[2] S.F. Crone, S. Lessmann, R. Stahlbock, The impact of preprocessing on data mining: an evaluation of classifier sensitivity in direct marketing, Eur. J. Oper. Res. 173 (3) (2006) 781–800.
[3] H. Liu, H. Motoda, Feature Selection for Knowledge Discovery and Data Mining, Kluwer Academic Publishers, Boston, 1998.
[4] Z. Zhu, Y.S. Ong, M. Dash, Wrapper-filter feature selection algorithm using a memetic framework, IEEE Trans. Syst. Man Cybern. 37 (1) (2007) 70–76.
[5] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artif. Intel. 97 (1) (1997) 273–324.
[6] H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (4) (2005) 491–502.
[7] A. Zarshenas, K. Suzuki, Binary coordinate ascent: an efficient optimization technique for feature subset selection for machine learning, Knowl. Based Syst. 110 (2016) 191–201 in press.
[8] S. Paul, M. Magdon-Ismail, P. Drineas, Column selection via adaptive sampling, Adv. Neural Inf. Process. Syst. (2015).
[9] C. Boutsidis, P. Drineas, M. Magdon-Ismail, Near-optimal column-based matrix reconstruction, SIAM J. Comput. 43 (2) (2014) 687–717.
[10] P. Drineas, M.W. Mahoney, S. Muthukrishnan, Relative-error CUR matrix decompositions, SIAM J. Matrix Anal. Appl. 30 (2) (2008) 844–881.
[11] C. Boutsidis, P. Drineas, M.W. Mahoney, Unsupervised feature selection for the *k*-means clustering problem, in: Proceeding of the 2009 Advances in Neural Information Processing Systems, 2009.
[12] S. Paul, M. Magdon-Ismail, P. Drineas, Feature selection for linear SVM with provable guarantees, Pattern Recogit. 60 (2016) 205–214.
[13] S. Paul, M. Magdon-Ismail, P. Drineas, Feature selection for linear SVM with provable guarantees, in: Proceeding of the 2015 Artificial Intelligence and Statistics, 2015.
[14] S. Paul, P. Drineas, Feature selection for ridge regression with provable guarantees, Neural Comput. 28 (4) (2016) 716–742.
[15] E.G. Talbi, Metaheuristics From Design to Implementation, Wiley Online Library, 2009.
[16] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[17] C. Lai, M.J. Reinders, L. Wessels, Random subspace method for multivariate feature selection, Pattern Recognit. Lett. 27 (10) (2006) 1067–1076.
[18] A.-R. Hedar, J. Wang, M. Fukushima, Tabu search for attribute reduction in rough set theory, Soft Comput. Fusion Found. Methodol. Appl. 12 (9) (2006) 909–918.
[19] R. Jensen, Q. Shen, Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches, IEEE Trans. Knowl. Data Eng. 16 (12) (2004) 1457–1471.
[20] M. Mafarja, S. Abdullah, A fuzzy record-to-record travel algorithm for solving rough set attribute reduction, Int. J. Syst. Sci. 46 (3) (2015) 503–512.
[21] M. Mafarja, S. Abdullah, Record-to-record travel algorithm for attribute reduction in rough set theory, J Theor. Appl. Inf. Technol. 49 (2) (2013) 507–513.
[22] M.M. Kabir, M. Shahjahan, K. Murase, A new local search based hybrid genetic algorithm for feature selection, Neurocomputing 74 (17) (2011) 2914–2928.

Please cite this article as: M.M. Mafarja, S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.04.053

[23] R. Bello, et al., Two-step particle swarm optimization to solve the feature selection problem, in: Proceedings of the Seventh International Conference on Intelligent Systems Design and Applications, IEEE Computer Society, 2007, pp. 691–696.
[24] S. Kashef, H. Nezamabadi-pour, An advanced ACO algorithm for feature subset selection, Neurocomputing 147 (2015) 271–279.
[25] E. Zorarpacı, S.A. Özel, A hybrid approach of differential evolution and artificial bee colony for feature selection, Expert Syst. Appl. 62 (2016) 91–103.
[26] J. Wang, T. Li, R. Ren, A real time idss based on artificial bee colony-support vector machine algorithm, in: Proceeding of the 2010 Third International Workshop on Advanced Computational Intelligence (IWACI), IEEE, 2010.
[27] E.-G. Talbi, A taxonomy of hybrid metaheuristics, J. Heuristics 8 (5) (2002) 541–564.
[28] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science 220 (4598) (1983) 671–680.
[29] S. Mirjalili, A. Lewis, The whale optimization algorithm, Adv. Eng. Softw. 95 (2016) 51–67.
[30] H.J. Touma, Study of the economic dispatch problem on IEEE 30-bus system using whale optimization algorithm, Int. J. Eng. Technol. Sci. 5 (1) (2016).
[31] A. Kaveh, M.I. Ghazaan, Enhanced whale optimization algorithm for sizing optimization of skeletal structures, Mech. Based Des. Struct. Mach. (2016) 1–18.
[32] D.P. Ladumor, et al., A whale optimization algorithm approach for unit commitment problem solution, in: Proceeding of the 2016 National Conference on Advancements in Electrical and Power Electronics Engineering (AEPEE-2016), Morbi, 2016.
[33] I.-S. Oh, J.-S. Lee, B.-R. Moon, Hybrid genetic algorithms for feature selection, IEEE Trans. Pattern Anal. Mach. Intell. 26 (11) (2004) 1424–1437.
[34] O.C. Martin, S.W. Otto, Combining simulated annealing with local search heuristics, Ann. Oper. Res. 63 (1) (1996) 57–75.
[35] K. Lenin, B.R. Reddy, M. Suryakalavathi, Hybrid Tabu search-simulated annealing method to solve optimal reactive power problem, Int. Electr. Power Energy Syst. 82 (2016) 87–91.
[36] Y. Lin, Z. Bian, X. Liu, Developing a dynamic neighborhood structure for an adaptive hybrid simulated annealing – tabu search algorithm to solve the symmetrical traveling salesman problem, Appl. Soft Comput. 49 (2016) 937–952.
[37] P. Vasant, Hybrid simulated annealing and genetic algorithms for industrial production management problems, Int. J. Comput. Methods 7 (02) (2010) 279–297.
[38] Z. Li, P. Schonfeld, Hybrid simulated annealing and genetic algorithm for optimizing arterial signal timings under oversaturated traffic conditions, J. Adv. Transp. 49 (1) (2015) 153–170.
[39] Y. Li, et al., A hybrid genetic-simulated annealing algorithm for the location-inventory-routing problem considering returns under E-supply chain environment, Sci. World J. 2013 (2013).
[40] L. Junghans, N. Darde, Hybrid single objective genetic algorithm coupled with the simulated annealing optimization method for building optimization, Energy Build. 86 (2015) 651–662.
[41] M. Mafarja, S. Abdullah, Investigating memetic algorithm in solving rough set attribute reduction, Int. J. Comput. Appl. Technol. 48 (3) (2013) 195–202.
[42] R. Azmi, et al., A hybrid GA and SA algorithms for feature selection in recognition of hand-printed Farsi characters, in: Proceeding of the 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems, 2010.
[43] J. Wu, Z. Lu, A novel hybrid genetic algorithm and simulated annealing for feature selection and kernel optimization in support vector regression, in: Proceeding of the 2012 IEEE Fifth International Conference on Advanced Computational Intelligence (ICACI)„ IEEE, 2012.
[44] K. Manimala, K. Selvi, R. Ahila, Hybrid soft computing techniques for feature selection and parameter optimization in power quality data mining, Appl. Soft Comput. 11 (8) (2011) 5485–5497.
[45] O. Olabiyisi Stephen, et al., Hybrid metaheuristic feature extraction technique for solving timetabling problem, Int. J. Sci. Eng. Res. 3 (8) (2012).
[46] W.C. Tang, Feature Selection For The Fuzzy Artmap Neural Network Using A Hybrid Genetic Algorithm And Tabu Search, USM, 2007.
[47] M. Majdi, S. Abdullah, N.S. Jaddi, Fuzzy Population-based meta-heuristic approaches for attribute reduction in rough set theory, World Acad. Sci. Eng. Technol. Int. J. Comput. Electr. Autom. Control Inf. Eng. 9 (12) (2015) 2289–2297.
[48] P. Moradi, M. Gholampour, A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy, Appl. Soft Comput. 43 (2016) 117–130.
[49] E.-G. Talbi, et al., Comparison of population based metaheuristics for feature selection: application to microarray data classification, in: Proceeding of the 2008 IEEE/ACS International Conference on Computer Systems and Applications, IEEE, 2008.
[50] Z. Yong, G. Dun-wei, Z. Wan-qiu, Feature selection of unreliable data using an improved multi-objective PSO algorithm, Neurocomputing 171 (2016) 1281–1290.
[51] J. Jona, N. Nagaveni, A hybrid swarm optimization approach for feature set reduction in digital mammograms, WSEAS Trans. Inf. Sci. Appl. 9 (2012) 340–349.
[52] M.E. Basiri, S. Nemati, A novel hybrid ACO–GA algorithm for text feature selection, in: Proceeding of the 2009 IEEE Congress on Evolutionary Computation, IEEE, 2009.
[53] R. Babatunde, S. Olabiyisi, E. Omidiora, Feature dimensionality reduction using a dual level metaheuristic algorithm, International Journal of Applied Information Systems (IJAIS) 7 (1) (2014) http://www.ijais.org.
[54] J. Jona, N. Nagaveni, Ant-cuckoo colony optimization for feature selection in digital mammogram, Pakistan J. Biol. Sci. 17 (2) (2014) 266.
[55] M. Nekkaa, D. Boughaci, Hybrid harmony search combined with stochastic local search for feature selection, Neural Process. Lett. 44 (1) (2016) 199–220.
[56] I. BoussaïD, J. Lepagnot, P. Siarry, A survey on optimization metaheuristics, Inf. Sci. 237 (2013) 82–117.
[57] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Comput. Electr. Eng. 40 (1) (2014) 16–28.
[58] R. Jensen, Q. Shen, Finding Rough Set Reducts with Ant Colony Optimization, in: Proceedings of the 2003 UK Workshop on Computational Intelligence, 2003, pp. 15–22.
[59] D. Goldberg, K. Deb, B. Korb, Messy genetic algorithms: motivation, analysis, and first results, Complex Syst. 3 (1989) 493–530.
[60] G. Sanchita, D. Anindita, et al., Evolutionary algorithm based techniques to handle big data, in: P.B.S. Mishra, et al. (Eds.), Techniques and Environments for Big Data Analysis: Parallel, Cloud, and Grid Computing, Springer International Publishing:, Cham, 2016, pp. 113–158.
[61] N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Am. Stat. 46 (3) (1992) 175–185.
[62] E. Emary, H.M. Zawbaa, A.E. Hassanien, Binary ant lion approaches for feature selection, Neurocomputing 213 (2016) 54–65.
[63] Blake, C.L. and C.J. Merz. UCI Repository of machine learning databases. 1998 [cited 2016 1 June]; Available from: ⟨http://www.ics.uci.edu/~mlearn/⟩.
[64] J. Friedman, T. Hastie, R. Tibshirani, The Elements of Statistical Learning, vol. 1, Springer, Berlin, 2001.
[65] H.M. Zawbaa, E. Emary, B. Parv, Feature selection based on antlion optimization algorithm, in: Proceeding of the 2015 Third World Conference on Complex Systems (WCCS), 2015.

Please cite this article as: M.M. Mafarja, S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.04.053

Majdi Mafarja received his B.Sc. in Software Engineering and M.Sc. in Computer Information Systems from Philadelphia University and The Arab Academy for Banking and Financial Sciences, Jordan in 2005 and 2007, respectively. He did his Ph.D. in Computer Science at National University of Malaysia (UKM). He was a member in Datamining and Optimization Research Group (DMO). Now he is an assistant professor at the Department of Computer Science at Birzeit University. His research interests include Evolutionary Computation, Meta-heuristics and Data mining.

Seyedali Mirjalili is a lecturer in Griffith College, Griffith University. He received his B.Sc. degree in Computer Engineering (software) from Yazd University, M.Sc. degree in Computer Science from Universiti Teknologi Malaysia (UTM), and Ph.D. in Computer Science from Griffith University. He was a member of Soft Computing Research Group (SCRG) at UTM. His research interests include Robust Optimization, Multi-objective Optimization, Swarm Intelligence, Evolutionary Algorithms, and Artificial Neural Networks. He is working on the application of multi-objective and robust meta-heuristic optimization techniques in Computational Fluid Dynamic (CFD) problems as well.

Please cite this article as: M.M. Mafarja, S. Mirjalili, Hybrid Whale Optimization Algorithm with simulated annealing for feature selection, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2017.04.053