Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 65 (2015) 623 – 632

Procedia Computer Science

International Conference on Communication, Management and Information Technology (ICCMIT2015)

Multi-Objective Gray-Wolf Optimization for Attribute Reduction

E.Emary[a], Waleed Yamany[b,*],Aboul Ella Hassanien[a] and Vaclav Snasel[c]

[a]*Faculty of Computers and Information, Cairo University, Egypt*
[b]*Fayoum University, Faculty of Computers and Information, Fayoum, Egypt*
[c]*Faculty of Electrical Engineering and Computer Science, VÅB-Technical University of Ostrava*
**Scientific Research Group in Egypt (SRGE), http: // www. Egyptscience.net*

Abstract

Feature sets are always dependent, redundant and noisy in almost all application domains. These problems in The data always declined the performance of any given classifier as it make it difficult for the training phase to converge effectively and it affect also the running time for classification at operation and training time. In this work a system for feature selection based on multi-objective gray wolf optimization is proposed. The existing methods for feature selection either depend on the data description; filter-based methods, or depend on the classifier used; wrapper approaches. These two main approaches lakes of good performance and data description in the same system. In this work gray wolf optimization; a swarm-based optimization method, was employed to search the space of features to find optimal feature subset that both achieve data description with minor redundancy and keeps classification performance. At the early stages of optimization gray wolf uses filter-based principles to find a set of solutions with minor redundancy described by mutual information. At later stages of optimization wrapper approach is employed guided by classifier performance to further enhance the obtained solutions towards better classification performance. The proposed method is assessed against different common searching methods such as particle swarm optimization and genetic algorithm and also was assessed against different single objective systems. The proposed system achieves an advance over other searching methods and over the other single objective methods by testing over different UCI data sets and achieve much robustness and stability.

Peer-review under responsibility of Universal Society for Applied Research

Keywords: Gray-Wolf ; Multi-Objective ; Attribute Reduction.

1. Introduction

Classification problems one of the essential task in data mining and machine learning, which target classify every object in data set into various collections based on the information depicted by its attributes. It is complicated to separate the attributes which are beneficial, without previous knowledge. Sometimes the dataset containing relevant,

Peer-review under responsibility of Universal Society for Applied Research
doi:10.1016/j.procs.2015.09.006

irrelevant, or redundant attributes[1]. The redundant and irrelevant attributes are slow down the classifier performance and they might even minimize the classification accuracy because the search space become huge[2,3]. Attribute reduction could handle this problem by choosing only relevant attribute for classification. The reduct set will improving the classifier performance and providing a faster and more cost effective classification, which leads to obtain comparable or even best classification accuracy from using all attributes[3].

Attribute reduction is a complicated mission because there exist complex interaction between attributes. A single redundant (relevant) attribute might become relevant (redundant) when working with other attribute[1]. So that, the optimal attribute collection (subset) will be a collection of integrated attributes that span over the diverse properties of the classes to properly discriminate them. The attribute reduction mission is challenging because of the huge search space. In search space the size is exceeds exponentially with respect to the number of attributes in the data set[2]. So, in practice the exhaustive search is impossible in almost cases. A diversity of search technique have been utilized to solve attribute reduction issue, such as greedy search based on sequential forward selection (SFS)[4] and sequential backward selection (SBS)[5]. However, these attribute reduction approaches still suffer from several of issues, such as stagnation in local optima and increasing in the cost of computational[1].

So as to improve the attribute reduction issues, an efficient global search algorithm is needed[6]. Evolutionary computation (EC) algorithms are well-known for their global search capability. Gray wolf optimization (GWO)[7] is a comparatively recent EC algorithm, that is computationally less expensive than some another EC techniques.

Generally, attribute reduction is a multi-objective issue. It has two main objectives, which are to minimize the size of attributes and to maximize the classification accuracy. Usually, these two objectives are contradictious and the optimal solution needs to be made in the presence of a tradeoff between them. Treating attribute reduction as a multi-objective issue can obtain a set of non-dominated attribute subsets to meet different requirements in real-world applications. Although GWO, multi-objective optimization, and attribute reduction have been individually investigated frequently, there are very few studies on multi-objective attribute reduction. Meanwhile, existing attribute reduction algorithms suffer from the issues of high computational cost, and GWO is argued computationally less expensive than other EC techniques. In addition, the utilizing of GWO for multi-objective attribute reduction has not been investigated[1].

This paper represents the first time that GWO has been applied to multi-objective attribute reduction. This will require novel methods to be introduced as there is no longer a single basis global solution but a set of solutions to meet different requirements.

The overall goal of this paper is to develop a GWO-based multi-objective attribute reduction approach to classification which include a small number of attributes and achieve a lower classification error rate than using all available attributes.

This goal is achieved by using the new gray-wolf inspired algorithm that exploits mutual information index as a fitness function to find solutions with minor redundancy that are passed to a second phase of optimization with different objective which is the classification performance and initialized with the past obtained solutions.

The remainder of this paper is organized as follows. Section II provides background information. Section III describes the GWO-based multi-objective attribute reduction algorithms. Section IV presents the experimental results with discussions. Section V provides the conclusion and future work.

2. Related Work

Greedy search based on sequential backward selection (SBS)[5] and sequential forward selection (SFS)[4] are two model wrapper techniques. SBS (SFS) starts with all attributes (no attributes), then candidate attributes are consecutively removed to (added from) the subset till the further removal (addition) does not rise the classification accuracy. But, these two techniques suffer from the issue of so-called nesting effect, that means once an attribute is eliminated (chosen) it could not be chosen (eliminated) later. This issue could be resolved by merging both SFS and SBS into one technique.

Thus, Stearns[8] proposes a plus-l-take away-k technique, which perform l times forward selection followed by k times backward elimination. However, it is hard to detect the best magnitudes of (l, k).

FOCUS[9] is a filter attribute reduction technique, which exhaustively examines all potential attribute subsets and then chooses the minimal attribute subset. But, the FOCUS technique was not computationally efficient due to the exhaustive search. Relief[10], also a filter technique specifies a weight to every attribute to indicate the relevance of the attribute to the aim concept. But, Relief does not transact with redundant attributes because it tries to obtain all relevant attributes regardless of the redundancy among them.

EC algorithms have been used to attribute reduction issues, such as genetic algorithm (GA), genetic programming (GP), ant colony optimization (ACO), and particle swarm optimization (PSO). Zhu[11] propose an attribute reduction technique using a mimetic technique which is a component of local search and GA. In this technique, individual attributes are firstly ranked according to a filter measure. GA utilizes the accuracy of classification as the fitness function and eliminates or adds an attribute according to the ranking information. The experiments prove that this technique outperforms GA individually and another techniques.

3. Preliminaries

3.1. Gray Wolf Optimization

Gray wolf optimization is presented in the following subsections based on the work in[7].

- Inspiration

Gray wolves are species with very strict social dominant hierarchy of leadership. The leaders are a male and a female, called alphas. The alpha is mostly responsible for making decisions about hunting, sleeping place, time to wake, and so on. The alphas decisions are dictated to the pack.

The second level in the hierarchy of grey wolves is beta. The betas are subordinate wolves that help the alpha in decision-making or other pack activities. The beta wolf is the best candidate to be the alpha in case one of the alpha wolves passes away or becomes very old to lead.

The lowest ranking grey wolf is omega. The omega plays the role of scapegoat. Omega wolves always have to submit to all the other dominant wolves. They are the last wolves that are allowed to eat.

The fourth class is called subordinate (or delta in some references). Delta wolves have to submit to alphas and betas, but they dominate the omega. *Scouts, sentinels, elders, hunters*, and *caretakers* belong to the delta category and each has its own defined responsibilities.

- Mathematical Modelling

The GWO the fittest solution is called the alpha (α) while the second and third best solutions are named beta (β) and delta (δ) respectively. The rest of the candidate solutions are assumed to be omega (ω). The hunting is guided by α, β, and δ and the ω follow these three candidates.

In order for the pack to hunt a prey them first encircling it. In order to mathematically model encircling behavior the following equations are used 1.

$$\vec{X}(t+1) = \vec{X}_p(t) + \vec{A}.\vec{D} \tag{1}$$

Where $\vec{D}$ is as defined in 2 and t is the number of iteration, $\vec{A}$, $\vec{C}$, are coefficient vectors, $\vec{X}_p$ is the prey position and $\vec{X}$ is the gray wolf position.

$$\vec{D} = \left|\vec{C}.\vec{X_p}(t) - \vec{X}(t)\right| \tag{2}$$

The $\vec{A}$, $\vec{C}$ vectors are calculated as in equations 3 and 4

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

$$\vec{C} = 2\vec{r_2} \tag{4}$$

Where components of $\vec{a}$ are linearly decreased from 2 to 0 over the course of iterations and r_1, r_2 are random vectors in $[0,1]$. The hunt is usually guided by the alpha. The beta and delta might also participate in hunting occasionally. In order to mathematically simulate the hunting behavior of grey wolves, the alpha (best candidate solution) beta, and delta are assumed to have better knowledge about the potential location of prey. The first three best solutions obtained so far and oblige the other search agents (including the omegas) to update their positions according to the position of the best search agents. So the updating for the wolves positions is as in equations 5, 6 and 7.

$$\vec{D_\alpha} = \left|\vec{C_1}.\vec{X_\alpha} - \vec{X}\right|, \vec{D_\beta} = \left|\vec{C_2}.\vec{X_\beta} - \vec{X}\right|, \vec{D_\delta} = \left|\vec{C_3}.\vec{X_\delta} - \vec{X}\right| \quad (5)$$

$$\vec{X_1} = \left|\vec{X_\alpha} - \vec{A_1}.\vec{D_\alpha}\right|, \vec{X_2} = \left|\vec{X_\beta} - \vec{A_2}.\vec{D_\beta}\right|, \vec{X_3} = \left|\vec{X_\delta} - \vec{A_3}.\vec{D_\delta}\right| \quad (6)$$

$$\vec{X}(t+1) = \frac{\vec{X_1} + \vec{X_2} + \vec{X_3}}{3} \quad (7)$$

A final note about the GWO is the updating of the parameter $\vec{a}$ that controls the tradeoff between exploitation and exploration. The parameter $\vec{a}$ is linearly updated in each iteration to range from 2 to 0 according to the equation 8.

$$\vec{a} = 2 - t.\frac{2}{Max_{iter}} \quad (8)$$

Where t is the iteration number and Max_{iter} is the total number of iteration allowed for the optimization.

Input: N number of wolves (agents) used $NIter$ number of iterations for optimization.
Output: Optimal wolf position
Initialize a population of N wolves' positions at random
While Stopping criteria not met **do**
Evaluate individual wolves' positions using the given fitness function.
Find the best wolf position; called α solution.
Find the best wolf position excluding α solution; Called β solution.
Find the best wolf excluding α and β solutions; Called δ solution.
Calculate the $\vec{a}$ parameter given the current iteration and the maximum number of iterations using equation 8
Foreach $Wolf_i$ **do**
Update the $Wolf_i$ position given the α, β and δ solutions, $\vec{a}$, and the $Wolf_i$ current position as in equation 7
End

Alg 1: GWO Search Algorithm

4. The Proposed Multi- Objective GWO Feature Selection Algorithm (MO-GWO)

Attribute reduction in general can be categorized into filter and wrapper methods. Filters perform attribute reduction based on the characteristics of data itself. Filters and it is performed independently of the learning algorithm by estimating the *usefulness* of attributes. Attributes that are not expected to provide valuable information for classification are filtered out of the dataset before training starts.

In the wrapper approach, the attribute space is explored to find an attribute (feature) subset guided by classification performance of individual attribute subsets. Hence intelligent exploration of search space is always a challenge as the single evaluation of fitness function is always time consuming. This approach may be slow since the classifier must be retrained on all candidate subsets of the attribute set and its performance must also be measured.

Filter methods always have poor performance in attribute reduction as it depends only on measuring the importance of attributes based on the characteristics of data regardless of the classifier used. On the other hand wrapper approach searches a very large space of attribute combinations which may be inefficient but it is much classifier guided and hence; if efficiently used, can have better performance.

The proposed algorithm is a wrapper attribute reduction that is guided by filter-based principles so that it exploits the classification performance of wrapper-based methods and the efficiency of the filter-based ones. A two-stage

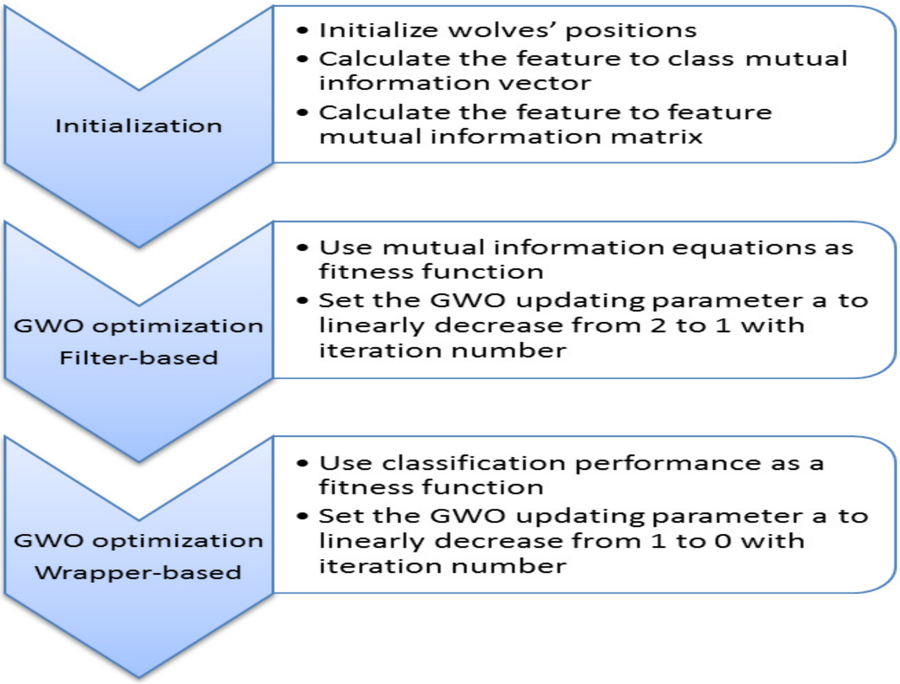

Fig. 1. The proposed Feature Selection Algorithm

gray wolf optimization is used to find attribute combination that both exploit filter-methods principles and wrapper-based methods principles; see Fig 1.

Initially GWO is used to search for a attribute combination that maximizes the following fitness function based on mutual information index:

$$\theta = V - P \tag{9}$$

Where V is the average mutual information between the selected attributes and the class labels and P is the average mutual information among the selected attributes. V and P are calculated as:

The used fitness function represents the predictability of attributes from each other and the predictability between individual features. Hence the goodness of a attribute combination is estimated as how much the selected attributes can correctly predict the output class labels and how much are they dependent. The convergence speed for GWO is ensured for it efficient searching capability and for the simplicity of the used fitness function that mainly depend on precalculated attribute to class mutual information vector and the precalculated feature to feature mutual information. This step of optimization is stopped at a predetermined number of iterations. The range for the parameter $\vec{a}$ that controls the tradeoff between exploitation and exploration is limited to the range from 2 to 1 rather than from 2 to 0 to keep solution diversity and to tolerate stagnation[12].

The obtained population by the end of the first stage is a set of solutions that maximizes the mutual information equation in 9. The obtained population is used as initial solutions for the second level optimization that used GWO to maximize classification performance as follows:

$$Fitness = CCR(D) \tag{14}$$

Where *CCR(D)* is the correct classification ratio at feature set *D*. The optimization in this second phase is much guided towards enhancing the classification accuracy given a preselected classifier; *K-nearest* neighbor in the current case, but the individual evaluation is much time consuming than the one used in the first stage. So, the first stage is used to motivate the search agents to regions with expected promising regions in the attribute space while the second level optimization uses exploitation to intensively find the solution with best classification performance[13].

A note worth mentioning is that the parameter $\overrightarrow{a}$ used by the GWO to control the diversification and intensification is set in this second level of optimization to the range from 1 to 0 to enhance the intensification of the solutions. This parameter choice allows for less deviation from the initial solutions to this second stage of optimization and allows also for fine tuning to find classification-performance guided solutions[14].

5. Experimental Results

5.1. Datasets and Parameters Used

Table 1 summarizes the 8 used data set for further experiments. The data set are drawn from the UCI data repository[15]. The data is divided into 3 equal parts one for *training*, the second part is for *validation* and the third part is for *testing*. GWO algorithm is compared with the particle swarm optimization (PSO)[16] and genetic algorithms (GA)[17] which are common for space searching. The parameter set for the GWO algorithm is outlined in table 2. Same number of agents and same number of iterations are used for GA and PSO.

Table 1. Description of The Data sets Used in Experiments.

Dataset	No. of Features	No. of Samples
Breast cancer	9	699
Exactly	13	1000
Exactly2	13	1000
Lymphography	18	148
M-of-N	13	1000
Tic-tac-toe	9	958
Vote	16	300
Zoo	16	101

5.2. Results and Discussion

Table 3 and 4 summarizes the result of running the different optimization algorithms for 10 different runs. Mean fitness function obtained by the GWO achieves remarkable advance over PSO and GA using the different fitness functions over the different data sets used which ensures the searching capability of GWO. By, remarking standard deviation of the solution obtained on the different runs of individual algorithms we can see that GWO has comparable or minimum variance value which proves the capability of convergence to global optima regardless of

Table 2. Parameter Setting for Gray- Wolf Optimization

Parameter	value(s)
No of wolves	5
No of iterations	100
Problem dimension Search domain	same as number offeatures in any given database18

the initial solutions which proves the stability of the algorithm. Also, on the level of best and worst solutions obtained at the different runs we can see advance on the fitness value obtained by GWO over PSO and GA over almost all the test data sets.

Table 3. Experiments Results of Different Runs for GA, PSO and GWO of Fitness Function.

Dataset	Breast Cancer								
	KNN			MI			MI and KNN		
	GA	PSO	GWO	GA	PSO	GWO	GA	PSO	GWO
Mean fitness	0.024034	0.030043	0.025751	- 0.37176	0.91588	-0.41202	0.024034	0.028326	0.027468
Std fitness	0.006509	0.009104	0.00607	0.06582	0.129056	0.031957	0.006509	0.008366	0.006509
Best fitness	0.017167	0.021459	0.021459	-0.42633	-0.42633	-0.42633	0.017167	0.021459	0.021459
Worst fitness	0.034335	0.042918	0.034335	-0.27369	-0.12947	-0.35466	0.034335	0.042918	0.038627
Dataset	Exactly								
	KNN			MI			MI and KNN		
	GA	PSO	GWO	GA	PSO	GWO	GA	PSO	GWO
Mean fitness	0.269461	0.297006	0.188024	0.003356	0.003414	0.003416	0.291617	0.308982	0.101198
Std fitness	0.051118	0.017381	0.141883	0.00013	0.000366	0.000314	0.0105	0.019905	0.120123
Best fitness	0.179641	0.278443	0.017964	0.003231	0.003049	0.003049	0.275449	0.290419	0.01497
Worst fitness	0.034335	0.042918	0.034335	-0.27369	-0.12987	-0.354866	0.034335	0.042918	0.038627
Dataset	Exactly2								
	KNN			MI			MI and KNN		
	GA	PSO	GWO	GA	PSO	GWO	GA	PSO	GWO
Mean fitness	0.232934	0.245509	0.234731	0.0033	0.003578	0.003244	0.24012	0.241317	0.232934
Std fitness	0.012235	0.009468	0.010285	0.000208	0.000356	0.00012	0.014729	0.012453	0.012235
Best fitness	0.218563	0.236527	0.224551	0.003157	0.003157	0.003157	0.227545	0.233533	0.218563
Worst fitness	0.248503	0.260479	0.248503	0.003613	0.004108	0.003375	0.263473	0.263473	0.248503
Dataset	Lymphography								
	KNN			MI			MI and KNN		
	GA	PSO	GWO	GA	PSO	GWO	GA	PSO	GWO
Mean fitness	0.167347	0.151616	0.127041	-0.10050	-0.09185	-0.194105	0.147534	0.179592	0.131633
Std fitness	0.039256	0.033997	0.050777	0.006798	0.008537	0	0.071252	0.033534	0.02286
Best fitness	0.122449	0.102041	0.061224	-0.11134	-0.10271	-0.194105	0.061224	0.142857	0.10204
Worst fitness	0.22449	0.183673	0.204082	-0.09255	-0.08419	-0.194105	0.244898	0.22449	0.163265
Dataset	M-of-N								
	KNN			MI			MI and KNN		
	GA	PSO	GWO	GA	PSO	GWO	GA	PSO	GWO
Mean fitness	0.097006	0.12515	0.073653	-0.04805	-0.04906	-0.053143	0.108982	0.584615	0.028144
Std fitness	0.041302	0.036815	0.064305	0.005811	0.00546	0.000101	0.03767	0.08503	0.010712
Best fitness	0.035928	0.086826	0.017964	-0.05318	-0.05318	-0.053189	0.065868	0.029533	0.017964
Worst fitness	0.149701	0.176647	0.146707	-0.04133	-0.03997	-0.052962	0.164671	0.035928	0.041916

Table 4. Experiments Results of Different Runs for GA, PSO and GWO of Fitness Function.

Dataset	Tic-tac-toe								
	KNN			MI			MI and KNN		
	GA	PSO	GWO	GA	PSO	GWO	GA	PSO	GWO
Mean fitness	0.22625	0.25	0.235625	-0.01815	-0.01076	-0.01815	0.245	0.241875	0.22125
Std fitness	0.021264	0.013258	0.029365	0	0.004195	0	0.019838	0.027828	0.017315
Best fitness	0.203125	0.23125	0.203125	-0.01815	-0.01815	-0.018158	0.21875	0.203125	0.203125
Worst fitness	0.253125	0.26875	0.265625	-0.01815	-0.00833	-0.018158	0.26875	0.271875	0.240625
Dataset	Vote								
	KNN			MI			MI and KNN		
	GA	PSO	GWO	GA	PSO	GWO	GA	PSO	GWO
Mean fitness	0.054 -	0.056	0.054	-0.20943	-0.19771	-0.444235	0.06	0.056	0.058
Std fitness	0.011402	0.008944	0.020736	0.136339	0.143631	0	0.023452	0.019494	0.019235
Best fitness	0.04	0.04	0.03	-0.44423	-0.44423	-0.444235	0.04	0.04	0.04
Worst fitness	0.07	0.06	0.08	-0.1025	0.093762	0.444235	0.1	0.09	0.09
Dataset	Zoo								
	KNN			MI			MI and KNN		
	GA	PSO	GWO	GA	PSO	GWO	GA	PSO	GWO
Mean fitness	0.076471	0.076471	0.076471	-0.31006	-0.27985	-0.528763	0.094474	0.112132	0.082531
Std fitness	0.049215	0.049215	0.053429	0.14243	0.024575	0.077866	0.023973	0.051896	0.043355
Best fitness	0	0	0	-0.56358	-0.30750	-0.563585	0.060606	0.03125	0.030303
Worst fitness	0.117647	0.117647	0.147059	-0.22423	-0.24388	-0.389472	-0.117647	0.176471	0.147059

Table 5 describes the average selected feature size by the different optimizers using different fitness functions over the different data sets. We can see that the proposed multi-objective function outputs solutions with minor feature size in comparison to the other single objective fitness functions thanks to the exploitation of mutual information in feature selection process. Also, we can see that GWO is still performing better for feature reduction.

Table 5. Experiments Results of Mean Attribute Reduction.

Dataset	KNN			MI			MI and KNN		
	GA	PSO	GWO	GA	PSO	GWO	GA	PSO	GWO
Breast cancer	0.644444	0.644444	0.511111	0.111111	0.177778	0.111111	0.577778	0.666667	0.622222
Exactly	0.615385	0.6	0.492308	0.892308	0.892308	0.892308	0.784615	0.507692	0.507692
Exactly2	0.292308	0.523077	0.234731	0.953846	0.892308	0.969231	0.353846	0.523077	0.261538
Lymphography	0.488889	0.377778	0.127041	0.266667	0.233333	0.055556	0.488889	0.488889	0.366667
M-of-N	0.630769	0.569231	0.461538	0.323077	0.369231	0.292308	0.692308	0.584615	0.461538
Tic-tac-toe	0.6	0.6	0.488889	0.111111	0.222222	0.111111	0.55555	0.622222	0.533333
Vote	0.375	0.525	0.4125	0.175	0.2125	0.0625	0.4625	0.3625	0.35
Zoo	0.5875	0.5625	0.5625	0.2625	0.225	0.0625	0.5875	0.5625	0.45

Table 6. Experiments Results of Mean Classification Accuracy.

Dataset	KNN			MI			MI and KNN		
	GA	PSO	GWO	GA	PSO	GWO	GA	PSO	GWO
Breast cancer	0.95279	0.958798	0.946781	0.886695	0.91588	0.924464	0.949356	0.959657	0.959657
Exactly	0.725526	0.673273	0.804805	0.664264	0.670871	0.675676	0.684685	0.669069	0.90991
Exactly2	0.747748	0.733333	0.748949	0.732733	0.727327	0.73093118	0.738739	0.724324	0.754955
Lymphography	0.74702	0.726776	0.776	0.754939	0.74135	0.6007699	0.770939	0.688	0.776
M-of-N	0.868468	0.836036	0.914114	0.806006	0.837237	0.81801816	0.855255	0.921321	0.972372
Tic-tac-toe	0.731661	0.712853	0.723511	0.673354	0.673354	0.673354	0.736677	0.748589	0.7360
Vote	0.916	0.93	0.912	0.93	0.952	0.966	0.94	0.924	0.932
Zoo	0.854545	0.842424	0.872727	0.654545	0.727273	0.6	0.884848	0.848485	0.854545

Table 6 summarizes the average testing performance of the different optimizers over the different data set. We can see that the performance of GWO is better than GA and PSO for all the used fitness functions over the test data sets. By comparing the performance of different fitness functions used; namely mutual information, classification performance and the multi-objective fitness functions, we can see the advance of the proposed multi-objective function on performance. This advance can be interpreted by the good description of data with minimal redundancy and classifier guidance by the second objective of the fitness function.

6. Conclusions and Future Work

In this work a system for attribute reduction was proposed based on multi-objective gray wolf optimization. The proposed method tolerates the problems that are common on both wrapper-based feature selection as well as filter-based ones. The proposed fitness function exploits the capabilities of mutual information index as measure to ensure data dependence and classification performance as a second objective to grantee classification performance. The gray-wolf optimization in comparison to PSO and GA proves good performance in reaching global minima and robustness against different initial starting solutions. In future we will try use three initialization methods depend on the forward and backward selection to enhancement the performance of algorithm.

7. References

1. Xue, Bing, Mengjie Zhang, and Will N. Browne. Particle swarm optimization for feature selection in classification: a multi-objective approach. IEEE transactions on cybernetics; 2013; 43: 1656-1671.
2. I. Guyon, A. Elisseeff. An introduction to variable and attribute selection. The Journal of Machine Learning Research; 2003; 3: 1157-1182.
3. M. Dash and H. Liu. Feature selection for Classification. Intelligent Data Analysis; 1997; 1: 131-156.
4. A. Whitney. A direct method of nonparametric measurement selection. IEEE Trans. Comput; 1971; 20: 1100-1103.
5. T. Marill and D. Green.On the effectiveness of receptors in recognition systems. IEEE Trans. Inf. Theory; 1963; 9: 11-17.
6. Bing Xue, Mengjie Zhang, Will N. Browne. Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. Applied Soft Computing; 2014; 18: 261-276.
7. Seyedali Mirjalili, Seyed Mohammad Mirjalili, Andrew Lewis. Grey Wolf Optimizer. Advances in Engineering Software; 2014; 69: 46-61.

8. S. Stearns. On selecting features for pattern classifier. Proceedings of the 3rd International Conference on Pattern Recognition, IEEE, Coronado, CA; 2003: 71-75.
9. H. Almuallim, T.G. Dietterich. Learning Boolean concepts in the presence of many irrelevant features. Artificial Intelligence; 1994; 69: 279-305.
10. K. Kira, L.A. Rendell. A practical approach to feature selection. Assorted Conferences and Workshops; 1992: 249-256.
11. Z.X. Zhu, Y.S. Ong, M. Dash. Wrapper-filter feature selection algorithm using a memetic framework. IEEE Transactions on Systems, Man, and Cybernetics. Part B: Cybernetics; 2007; 37: 70-76.
12. Emary, E., Waleed Yamany, and Aboul Ella Hassanien. New approach for feature selection based on rough set and bat algorithm. In Computer Engineering & Systems (ICCES), 9th International Conference on IEEE; 2014: 346-353.
13. Yamany, Waleed, E. Emary, and Aboul Ella Hassanien. Wolf search algorithm for attribute reduction in classification. In Computational Intelligence and Data Mining (CIDM), IEEE Symposium; 2014 :351-358.
14. Yamany, Waleed, Nashwa El-Bendary, Hossam M. Zawbaa, Aboul Ella Hassanien, and Václav Snášel. "Rough Power Set Tree for Feature Selection and Classification: Case Study on MRI Brain Tumor." In Innovations in Bio-inspired Computing and Applications. Springer International Publishing; 2014: 259-270.
15. Bache, K. Lichman, M. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science; 2013.
16. Kennedy, J.; Eberhart, R. Proceedings of IEEE International Conference on Neural Networks; 1995: 1942-1948.
17. Eiben, A. E. et al. Genetic algorithms with multi-parent recombination". PPSN III: Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature; 1994: 540-584.