A Novel Genetic Grey Wolf Optimizer for Global Optimization and Feature Selection

Badra Khellat Kihel [1,2]   Samira Chouraqui[1]

[1]Department of Informatics, Faculty of Mathematic and Informatics
Université des Sciences et de la Technologie d'Oran Mohamed-Boudiaf USTO-MB
B.P. 1505 EL M'NAOUER 31000 Oran, Algeria
[2]Department of Economics, Oran2 University Oran, Algeria
badra.khellat@univ-usto.dz ; khellat_badra@yahoo.fr

Abstract— **In this paper, a new stochastic search strategy inspired by the Grey Wolf optimization theory is proposed for feature subset selection. Grey Wolf Optimization algorithm (GWO) is a new metaheuristic optimization technique. Its principle is to reproduce the behavior of grey wolves in nature to hunt in a cooperative way. In this work, we have used the Grey Wolf Optimizer and Genetic algorithm to select the most relevant features in a dataset. Then we have proposed a new Genetic Grey Wolf Optimization algorithm. In our proposed strategy, feature selection algorithm is formulated as an optimization problem that searches an optimum with less number of features in a feature space and a good accuracy. The goal of our study is to achieve a balance between the classification accuracy and the size of the feature subsets selected. Our proposed approach has been evaluated on 10 standard datasets taken from UCI repository and validated on 02 big datasets used in literature. The experimental results show the superiority of GWO algorithm in classification performance and dimensionality reduction.**

Keywords— Grey Wolf Optimizer; Genetic Algorithm; feature selection; classification; K nearest neighbors.

I. INTRODUCTION

Metaheuristic algorithms have been successfully applied for global optimization [1]. Different models, techniques and evolutionary approaches have been explored for the aim of feature selection.

The problem of Feature Selection (FS) has been widely investigating due to its importance to a number of disciplines such as pattern recognition and knowledge discovery. Feature selection allows the reduction of feature space, which is crucial in reducing the training time and improving the prediction accuracy. This is achieved by removing irrelevant, redundant, and noisy features (i.e. selecting the subset of features that can achieve the best performance in terms of accuracy and computational time) [2-3].

Recently, many FS based on new optimizers were proposed in the literature including Whale Optimization Algorithm (WOA) [4], Butterfly Optimization Algorithm (BOA) [5], Dragonfly algorithm (DA) [6], Firefly Algorithm (FFA) [7] and Selfish Herd Optimizer (SHO) [8] that have been successfully employed for solving FS problems. Another recent algorithm is the Ant Lion Optimizer (ALO). ALO algorithm has been proposed for feature selection in [9-10]. In [9], two incremental hill-climbing are hybridized with the Binary Ant Lion Optimizer in a model called HBALO.

In [11] PSO and ACO were combined for FS task. In another work, the Ant Colony Optimization (ACO), Artificial Bee Colony (ABC) and Firefly Algorithms (FFA) were used to select the most relevant features in a dataset. The populations generated by the three algorithms (ACO, ABC and FA) were used instead of a random one for Genetic Algorithm (GA) [12].

In this work, we were particularly attracted by the use of bio-inspired methods and there hybridization for feature selection. We have used two algorithms: Grey Wolf Optimizer and Genetic Algorithm to select the most relevant features, then we used a new system which consists in hybridization between genetic and grey wolf algorithm with the use of k-nearest neighbors in each iteration to evaluate the error rate of the selected features.

Grey Wolf Optimizer is a recent metaheuristics, which offers better performance in feature selection task [13]. However, the new positions of wolves are mostly based on the experience of three solutions or leaders: Alpha, Beta and Delta in the position update. In this way, the wolves will gradually become the same as the leaders. To avoid premature convergence, we proposed a new Genetic Grey Wolf Optimizer (GGWO), using crossover operator between Alpha, Beta and Delta wolves to update the wolves' position, to select the best feature subset which include a small number of features and achieve a lower classification error rate than using all available features. The proposed algorithms are tested on 12 well-known datasets and show a very good performance when comparing to other algorithms in the literature.

The rest of this paper is organized as follows: section 2 reviews feature selection algorithm. Section 3 introduces bio-inspired algorithms for feature selection: grey wolf optimizer and genetic algorithm. Section 4 presents the proposed approach. The experimental results obtained are presented and discussed in section 5and compared with previous works in section 6. Finally, section 7 concludes the paper.

978-1-7281-7503-4/20/$31.00 ©2020 IEEE

II. FEATURE SUBSET SELECTION

The identification of useful and informative attributes for a given dataset, broadly referred to as Feature Selection (FS), is an attractive and challenging research topic for several domains including predictive data mining, pattern recognition, machine learning and information retrieval. One of the fundamental motivations for feature selection is the curse of dimensionality. In fact, the presence of useless features may not only deteriorate the performance of learning algorithms but also obscure information behind data. Considered as a fundamental problem in machine learning, the role of FS is critical, especially in a context deemed with irrelevant features (i.e. redundant and noisy features) [14].

In literature, the authors state a list of three objectives of using feature selection for classification:

- To reduce the task of extraction of characteristics;
- To improve the precision of the module of classification;
- To improve the reliability of the performance estimation.

Existing feature selection algorithms can be broadly classified into two categories: wrapper approaches and filter approaches. Wrapper approaches include a learning/classification algorithm in the evaluation procedure, while filter approaches do not include such algorithm. Filter approaches are argued to be computationally less expensive and more general, while wrapper approaches can usually achieve better results. In recent years, a large number of feature selection algorithms have been proposed [15].

Two principal approaches are used in the literature for feature selection: approach by extraction (creation of new variables by combination of those existing) and that based on the selection of relevant attributes (choose only an optimal subset of attributes for a given criterion) [2-3].

III. BACKGROUND

A. Feature selection based GWO

Optimization problem is one of the most difficult and challenging problems that has received considerable attention over the last decade. Researchers have been constantly investigating better ways to solve it. Recently, one optimization technique called grey wolf optimizer algorithm has gained the interest of many researchers. This algorithm is a type of swarm intelligence algorithm based on the position change of wolves according to the experience of leaders: Alpha, Beta and Delta [13] [16].

In the feature selection problem, a representation for candidate feature subset must be chosen and encoded. In most studies, wolf is a binary string of length equal to the total number of features so that each bit encodes a single feature. A bit of ‘1’ (‘0’) implies the corresponding feature is selected (excluded).

For modeling the social hierarchy of wolves, the leaders of the park are considered as the alpha (α).the main responsibility of alpha is making decisions. Beta (β) known to assist alpha in making decisions and the main responsibility of beta is the feedback suggestions. Delta (δ) performs as scouts, and controls omega (ω) wolves by obeying alpha and beta wolves. The omega wolves must obey every other wolf.

In GWO, α, β, and δ, guides the hunting process and ω wolves follows them. The encircling behavior for the pack to hunt a prey can be expressed as

$$X(t+1) = X_p(t) - A.D \quad (1)$$

Where X_p is the position of prey, A , C are coefficient vectors, and D is defined as

$$D = \left|C.X_p(t) - X(t)\right| \quad (2)$$

The leaders are guiding the omega wolves to move toward the optimal position. Mathematically, the new position of wolf is:

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

Where X_1 , X_2 and X_3 are the three best wolves in the swarm at a given iteration t.

B. Feature selection based GA

Genetic algorithms (GA) are adaptive search techniques, based on the analogy with biology, in which a set of possible solutions evolves via natural selection. In recent years, some researchers addressed the feature selection problem using genetic algorithms in a wrapper model approach**.** In these works, genetic algorithms were used to explore the space of all possible subsets of feature so to obtain a set of features which maximizes the predictive accuracy of a specific learning algorithm [17].

Genetic based feature selection algorithms that have emerged in the literature are usually implemented as follows:

- Individuals represent subsets of features by means of binary strings. Each binary digit (gene) stands for the presence (1) or the absence (0) of a given feature.
- Standard genetic operators, crossover and mutation, are applied without any modification.
- The predictive accuracy of a given learning algorithm is used to measure fitness of individual.

The objective function is defined by the error rate calculated using k-Nearest Neighbors (KNN). In proposed algorithms, the classifiers are used to evaluate the subsets, which are selected with classification accuracies. Therefore, they play an important role in the algorithm: the classifiers with superior performance will improve the final results [14-15].

IV. PROPOSED APPROACH GGWO

Recently, another kind of hybridization was explored. The latest does not perform on combination of wrappers and filters or the use of local search to enhance exploitation performance but it is extended to metaheuristics combination. Several bio-inspired hybrid methods were proposed to perform FS problems.

The GGWO's basic idea is to increase the algorithm's capability to exploit GA with the ability to explore GWO to achieve good results. Initially, the population of grey wolves is randomly initialized (either bit 1 or 0). Afterward, the fitness of each wolf is evaluated using Knn. The best, second, and third best solutions are defined as alpha, beta, and delta. Then, the position of wolf is updated by applying the crossover between X_1, X_2, and X_3.

$$X(t+1) = \mathrm{Crossover}(X_1, X_2, X_3) \quad (4)$$

Next, the fitness of each wolf is evaluated and the positions of alpha, beta, and delta are updated. The algorithm is repeated until the stopping criterion is satisfied (number of iterations). Finally, the alpha solution is selected as the optimal feature subset.

```
GGWO pseudo code
    Begin
    Generate initial population of wolves
    Initialize GA and GWO parameters
    Evaluate the fitness of wolves
    Select α , β and δ
    Repeat
    |  For each wolf
    |   |Compute X1, X2 and X3
    |   |Compute Xnew = crossover (X1, X2 , X3)
    |  Evaluate the fitness of wolves
    |  Update position of α , β and δ
    Until iteration =T
    Return ( α )
```

V. EXPERIMENTAL RESULTS

A. Datasets used

In order to make our evaluation results comparable to the most of the published results in feature selection evaluations, we have chosen datasets from the UCI machine learning repository.

TABLE I. DATASETS USED

Data set	Number of Features	Number of Classes	Number of instances
Spect Heart (Binary)	23	02	187
Spect Heart	44	02	187
Parkinson'Diseace	22	02	96
Parkinson 2	26	02	1040
Glass dataset	10	06	214
Breast tissue	09	06	106
Ionosphere	34	02	351
Musk1	166	02	476
Wisconsin Breast Cancer	09	02	699
Wisconsin Breast Cancer Diagnostic	31	02	569
Madelon	500	02	2000
ORL	1024	40	400

B. Results and discussion

In this section we present the experimental results obtained for the different approaches on datasets. Figure.1 demonstrates the convergence curve of feature selection methods for different datasets.

Table 2 and figure 1 presents the results of the GA, GWO and GGWO proposed algorithms for feature selection in terms of number of selected features and classification error for the twelve databases.

The results demonstrate that the proposed approach produced a very good error rate using KNN classifier with less number of features. Features selected by the GWO algorithm can accomplish the goal of achieving higher accuracy with smaller size of features. Results show that the proposed methods are able to produce good performance on reducing the effects of the outliers and the noises and improve classification accuracy.

The results of the proposed algorithm show that the manner of combination between Alpha, Beta and Delta wolves influence the robustness and the algorithm convergence, by using the crossover operator to calculate position that ensures diversity of the population

TABLE II. RESULTS FOR ALGORITHMS USED

Data set	GWO			GA			GGWO	
	Number of features	Error	Time CPU	Number of features	Error	Time CPU	Number of features	Error
Dataset1	10	0.3250	147.68	09	0.2875	240.33	17	0.3000
Dataset2	16	0.1500	74.11	22	0.1500	229.63	31	0.1750
Dataset3	11	0.0208	74.10	11	0.0104	245.55	17	0.0208
Dataset4	14	0.3413	188.73	17	0.3356	300.64	19	0.3346
Dataset5	05	0.2336	147.43	07	0.2336	215.74	07	0.2477
Dataset6	04	0.2642	146.50	06	0.2736	238.78	06	0.2642
Dataset7	07	0.0826	144.33	11	0.0769	252.70	22	0.0969
Dataset8	50	0.0987	167.75	85	0.0924	233.24	136	0.1092
Dataset9	06	0.0258	162.88	06	0.0272	278.39	07	0.0258
Dataset10	07	0.0422	162.40	16	0.0615	248.86	28	0.0598
Dataset11	82	0.1885	761.96	237	0.2545	2729.97	390	0.3135
Dataset12	405	0.0775	160.98	512	0.0775	289.17	732	0.0500

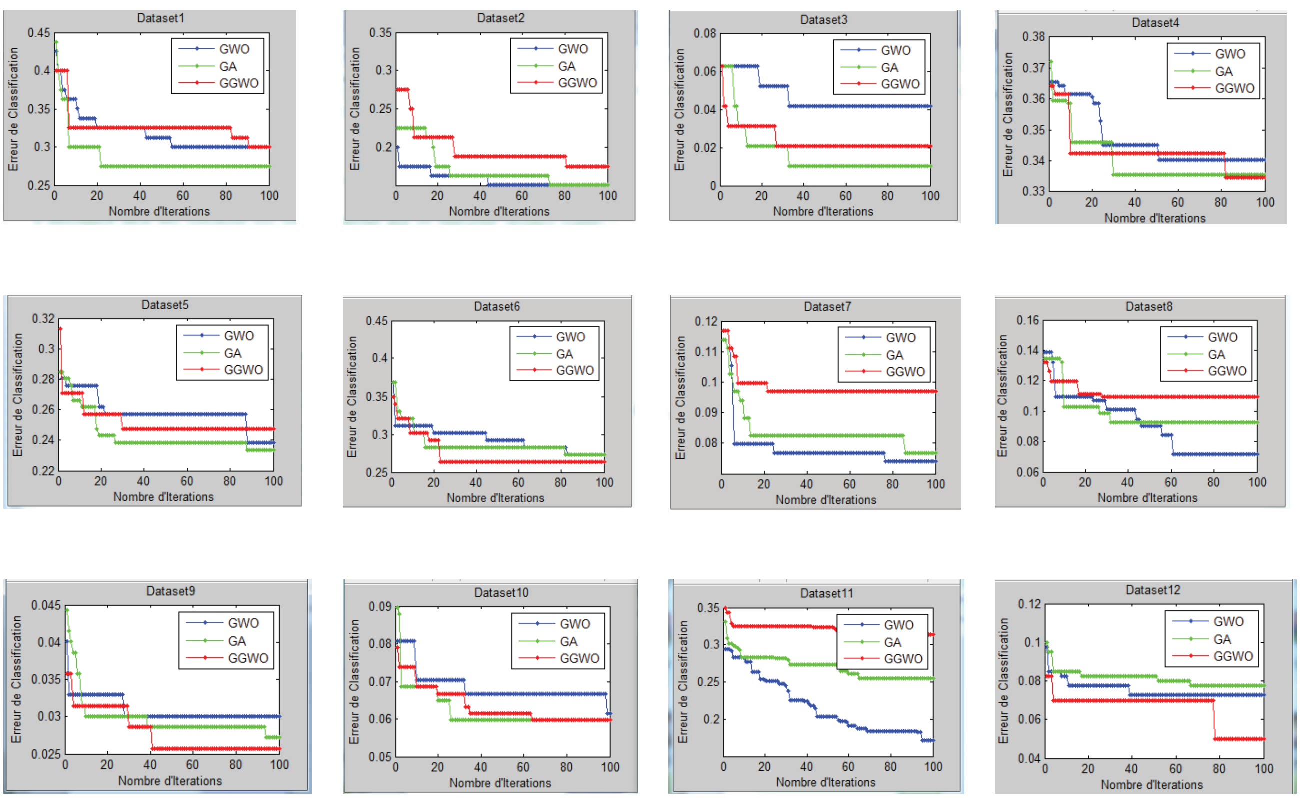

Fig. 1. The convergence curve of feature selection methods for different datasets

VI. COMPARATIVE STUDY

In this section, the results of the proposed methods are compared with some of the proposed methods in the literature.

TABLE III. PREVIOUS WORKS

Algorithm	Dataset 1		Dataset 2		Dataset 3		Dataset 4		Dataset 5		Dataset 6		Dataset 7		Dataset 8		Dataset 9		Dataset 10	
	Error	NF	Error	NF	Error	NF	Error	NF	Error	NF	Error	NF	Error	NF	Error	NF	Error	NF	Error	NF
GWO	0.325	10	**0.150**	**16**	0.020	11	0.341	14	**0.233**	**5**	**0.264**	**4**	**0.082**	**7**	**0.098**	**50**	0.025	6	**0.042**	**7**
GA	0.287	09	0.150	22	0.010	11	0.335	17	0.233	7	0.273	6	0.076	11	0.092	85	0.027	6	0.061	16
GGWO	0.300	17	0.175	31	0.020	17	0.334	19	0.247	7	0.264	6	0.096	22	0.109	136	0.025	7	0.059	22
IFA 1 [18]	0.112	14	0.056	28	**0.023**	**10**	**0.171**	**13**	0.510	7	0.325	7	0.026	21	0.030	94	0.017	7	0.008	18
IFA 2 [18]	0.099	11	0.056	24	0.029	12	0.168	16	0.396	8	0.315	5	0.032	22	0.024	83	**0.017**	**6**	0.008	21
ACO-ABC-FA/GA [12]	**0.094**	**8**	0.060	20	0.020	11	0.157	17	0.392	8	0.282	7	0.030	19	0.029	84	0.016	8	0.008	20

Table 3 depicts the error classification rate and number of selected features of the proposed system. These obtained results were compared with the existing methods. The results obtained by our approach show a good performance comparing to the existed algorithms.

CONCLUSION

In this work a system to evaluate Feature Selection Algorithms was proposed for the aim of understanding their general behaviour on the particularities of relevance, irrelevance, redundancy and sample size of synthetic data sets. In addition, new methods for feature selection based on grey wolf optimizer and genetic algorithms were proposed. The proposed algorithms were tested on 12 well-known UCI datasets. Simulation results show the performance of the grey wolf optimizer algorithm that outperforms the other algorithms. As future contribution, this work can be extended in many ways to carry up richer evaluations and to combine different evaluation measures.

REFERENCES

[1] M. Esseghir. “Metaheuristics for The Feature Selection Problem : Memetic, Adaptive and Swarm Approaches”. Thèse de Doctorat Université d’Artois, 2011.

[2] B. Xue, “Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach”, IEEE transactions on cybernetics, Vol.43,No.6, pp.1656–1671, 2013.

[3] H. Liu and L. Yu, “Toward Integrating Feature Selection Algorithms for Classification and Clustering”, inKnowledge and Data Engineering, Vol. 17, No. 4, pp. 491-502,2005.

[4] M.Mafarja and M,Mirjalili, “Whale optimization approaches for wrapper feature selection”. Applied Soft Computing, Vol.62, pp.441–453, 2018.

[5] S. Arora and P. Anand, “Binary butterfly optimization approaches for feature selection”, Expert Systems with Applications, Vol.116, 2018.

[6] S. Mirjalili, “Dragonfly algorithm: a new metaheuristic optimization technique for solving single objective, discrete, and multi-objective problems”, Neural Computing and Applications, vol. 27, pp. 1053-1073, 2016.

[7] S. Mukherjee and L. Bhaumik, “Simultaneous Clustering and Feature Selection Using Nature-Inspired Algorithm”, Lecture Notes in Networks and Systems, pp. 545 – 550, 2019.

[8] P. Anand and S. Arora, “A novel chaotic selfish herd optimizer for global optimization and feature selection”, Artificial Intelligence Review, Vol.51, pp.1-46, 2019

[9] S. Mirjalili, “Hybrid Binary Ant Lion Optimizer with Rough Set and Approximate Entropy Reducts for Feature Selection”, Soft Computing, pp. 1-17, 2018.

[10] E. Emary and H. M. Zawbaa, “Feature selection via Levy Antlion optimization”, Pattern Analysis and Applications, Vol.21, pp. 1–20, 2018.

[11] K. Menghour and L. Meslati, “Hybrid ACO-PSO Based Approaches for Feature Selection”, International Journal of Intelligent Engineering and Systems, Vol.9, No.3, pp.65-79, 2016.

[12] B. Khellat and M. Benyettou, “Hybrid Bio-inspired Approach for Feature Subset Selection”, International Journal of Applied Engineering Research , Vol. 13, No.10, pp. 7895-7902, 2018.

[13] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer”, Advances in Engineering Software, Vol. 69, pp. 46-61, 2014.

[14] M. Anusha and J. Sathiaseelan, “Feature Selection Using K-Means Genetic Algorithm for Multi-objective Optimization”, Procedia Computer Science, Vol. 57, pp. 1074-1080, 2015.

[15] T. Mohammed , S.Alhayali , O.Bayat and O. Uçan, “Feature Reduction Based on Hybrid Efficient Weighted Gene Genetic Algorithms with Artificial Neural Network for Machine Learning Problems in the Big Data”, Scientific Programming, Vol. 2018, 2018.

[16] Q. Al-Tashi, H. Rais and S. Jadid, “Feature Selection Method Based on Grey Wolf Optimization for Coronary Artery Disease Classification”, International Conference of Reliable Information and Communication Technology, Kuala Lumpur, Malaysia, pp. 257–266, 2018.

[17] A. Al-ani, “Ant Colony Optimization for Feature Subset Selection”, World academy of science, engineering and technology, Vol.1, No. 4, 2007.

[18] B. Khellat and S. Chouraqui, “Firefly Optimization Using Artificial Immune System for Feature Subset Selection”, International Journal of Intelligent Engineering and Systems , Vol. 12, No.4, pp. 377-347, 2019.