



Feature Selection Algorithm Based on Multi Strategy Grey Wolf Optimizer

Guangyue Zhou, Kewen Li, Guoqiang Wan, Hongtu Ji

► To cite this version:

Guangyue Zhou, Kewen Li, Guoqiang Wan, Hongtu Ji. Feature Selection Algorithm Based on Multi Strategy Grey Wolf Optimizer. 11th International Conference on Intelligent Information Processing (IIP), Jul 2020, Hangzhou, China. pp.35-45, 10.1007/978-3-030-46931-3_4 . hal-03456986

HAL Id: hal-03456986

<https://inria.hal.science/hal-03456986>

Submitted on 30 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Feature selection algorithm based on Multi Strategy grey wolf optimizer^{*}

Guangyue Zhou¹, Kewen Li¹, Guoqiang Wan², and Hongtu Ji²

¹ College of Computer Science and Technology, China University of Petroleum,
Qingdao 266580, China

guangyuezhou@qq.com, likw@upc.edu.cn

² Shengli Oilfield Company, SINOPEC, Dongying 257015, China
{wanguoqiang.slyt, jihongtu.slyt}@sinopec.com

Abstract. Feature selection is an important part of data mining, image recognition and other fields. The efficiency and accuracy of classification algorithm can be improved by selecting the best feature subset. The classical feature selection technology has some limitations, and heuristic optimization algorithm for feature selection is an alternative method to solve these limitations and find the optimal solution. In this paper, we proposed a Multi Strategy Grey Wolf Optimizer algorithm (MSGWO) based on random guidance, local search and subgroup cooperation strategies for feature selection, which solves the problem that the traditional grey wolf optimizer algorithm (GWO) is easy to fall into local optimization with a single search strategy. Among them, the random guidance strategy can make full use of the random characteristics to enhance the global search ability of the population, and the local search strategy makes grey wolf individuals make full use of the search space around the current best solution, and the subgroup cooperation strategy is very important to balance the global search and local search of the algorithm in the iterative process. MSGWO algorithm cooperates with each other in three strategies to update the location of grey wolf individuals, and enhances the global and local search ability of grey wolf individuals. Experimental results show that MSGWO can quickly find the optimal feature combination and effectively improve the performance of the classification model.

Keywords: Feature Selection · Heuristic Optimization Algorithm · Multi Strategy Grey Wolf Optimizer Algorithm.

1 Introduction

Feature selection plays an important role in machine learning, data mining and other classification applications. Its goal is to remove the noise in the original

^{*} Supported by organization from the National Natural Science Foundation of China (No.61673396), and the Natural Science Foundation of Shandong Province, China (No.ZR2017MF032).

data and select the best identification feature. In addition, feature selection can improve the efficiency of classification by reducing the dimension of the original data. In recent years, more and more heuristic search algorithms are used for feature selection. Heuristic search algorithm can get a group of solutions at a time, which can get good results with less time and calculation cost. Many experts and scholars have done a lot of research on heuristic search algorithm: Genetic Algorithm (GA) is an evolutionary algorithm, which can search randomly and find the optimal solution based the evolution law of nature [1]. Particle Swarm Optimization (PSO) is a classical swarm intelligence optimization algorithm, which is based on the research of birds' predatory behavior. Each solution is regarded as a particle with a specific position, fitness and speed vector, and its motion direction and speed are adjusted according to the global optimal solution and the optimal solution found by the particle itself, and gradually approach the optimal solution [2]. The Whale Optimization Algorithm (WOA) is a heuristic optimization algorithm, which simulates the predatory behavior of humpback whales in nature. Compared with other swarm optimization algorithms, the main difference is that WOA simulates the bubble net attack of whales by following the best or random individuals and using the spiral mechanisms [3]. The Grey Wolf Optimizer (GWO) is a new evolutionary algorithm, which mainly simulates the predatory behavior of grey wolf group, through the process of wolves tracking, encircling, chasing and attacking to achieve the purpose of optimized search [4]. As GWO has the advantages of simple principle, few parameters to be adjusted, easy to realize and strong global search ability, its research has made remarkable progress. Emary and others first applied GWO to feature selection in 2015, and proposed two binary GWO feature selection methods based on different update mechanisms [5].

GWO algorithm can not effectively find the global optimal feature combination due to its single search strategy and insufficient global search ability [6]. Therefore, in order to improve the effectiveness of GWO for feature selection, this paper proposes a Multi Strategy Grey Wolf Optimizer algorithm (MSGWO), which solves the problem caused by a single search strategy and improves the accuracy of the original GWO. MSGWO includes three different search strategies—— Random guidance strategy, local search strategy and sub group cooperation strategy. The grey wolf optimizer algorithm with three strategies can further improve the search efficiency and find the optimal feature combination.

2 Grey Wolf Optimizer (GWO)

GWO is an intelligent optimization algorithm proposed by Mirjalili[4] in 2014. Due to its simple principle, fewer parameters to be adjusted, simple implementation and strong global search ability, the method is becoming more and more popular. Many research have been carried out using GWO [7–11].GWO algorithm is inspired by the predatory behavior of grey wolves, and it optimizes search through hunting, searching for prey, encircling prey, and attacking prey. There is a strict hierarchy between them. α , β , δ and ω represent different grades

of grey wolves, and the dominance rate decreases from top to bottom. In order to model the grey wolf's social system mathematically, α is regarded as the optimal solution, β and δ are regarded as the suboptimal solution and the third optimal solution, respectively. They lead other wolves toward the possible position of prey. ω is regarded as the rest of the solutions, which is updated according to the positions of α, β and δ . Three definitions of the algorithm [4] are given below.

Definition 1 Distance between Grey Wolf and Prey

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \quad (1)$$

Where t indicates the current iteration, \vec{X}_p represents position vector of prey, $\vec{X}(t)$ represents current position vector of grey wolf.

$$\vec{C} = 2 \cdot \vec{r}_1 \quad (2)$$

Where r_1 is random vector in $[0, 1]$, \vec{C} is coefficient vector. We can explore and exploit search space by randomly enhancing ($C > 1$) or weakening ($C < 1$) the distance between prey and grey wolf.

Definition 2 Update position of Grey Wolf

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (3)$$

$$\vec{A} = 2\vec{a} \cdot \vec{r}_2 - \vec{a} \quad (4)$$

Where components of \vec{a} are linearly decreased from 2 to 0, r_2 is random vector in $[0, 1]$. As A decreases, half of the iterations are used for exploring ($|A| > 1$), and the rest for exploiting ($|A| < 1$).

Definition 3 Determine position of prey

In the abstract search space, the exact position of the prey (optimal solution) is not known. According to the hierarchy of grey wolves, hunting is usually guided by α, β and δ . Therefore, it is assumed that α (optimal candidate solution), β (suboptimal candidate solution), and δ (third optimal candidate solution) have a better acquaintance of the position of prey. It is known that grey wolves α, β and δ are closest to prey. By preserving the obtained three optimal solutions during each iteration, the orientation of prey can be determined according to the positions of the three optimal solutions, and other grey wolf individuals are forced to update their positions according to the three optimal solutions. The mathematical descriptions of grey wolf individuals tracking prey orientation are as follows:

$$\vec{D}_\alpha = |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}|, \vec{D}_\beta = |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}|, \vec{D}_\delta = |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}| \quad (5)$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha, \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta, \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta \quad (6)$$

$$\vec{X}(t+1) = 1/3(\vec{X}_1 + \vec{X}_2 + \vec{X}_3) \quad (7)$$

The distances between grey wolf individuals and α, β and δ are calculated in terms of formulas (5) and (6). Then the direction of grey wolf individuals moving towards prey are judged in terms of formula (7). Where $\vec{X}_\alpha, \vec{X}_\beta, \vec{X}_\delta$ represent the positions of α, β, δ respectively, $\vec{C}_1, \vec{C}_2, \vec{C}_3$ are random vectors and \vec{X} is the position of the current solution.

3 Multi Strategy Grey Wolf Optimizer algorithm

3.1 Random guidance strategy

In GWO, α, β and δ lead ω to the promising region to search for the optimal solution, but only following the optimal solution to update is easy to lead to premature convergence in the current optimal position, making GWO fall into the local optimum. In this paper, we randomly select a grey wolf position $\overrightarrow{X_{rand}}$, and make other individuals update the position according to $\overrightarrow{X_{rand}}$. The mathematical description is as follows:

$$\overrightarrow{D_{rand}} = |\overrightarrow{C_g} \cdot \overrightarrow{X_{rand}} - \overrightarrow{X}| \quad (8)$$

$$\overrightarrow{X_g} = \overrightarrow{X_{rand}} - \overrightarrow{A_g} \cdot \overrightarrow{D_{rand}} \quad (9)$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X_g} \quad (10)$$

Where $\overrightarrow{D_{rand}}$ is the distance between grey wolf individual and $\overrightarrow{X_{rand}}$, $\overrightarrow{X}(t+1)$ is the updated position, and $\overrightarrow{C_g}$ is the random vector.

This strategy can make the gray wolf individuals that converge too early jump out of the local optimum, expand the global search range of the population, and increase the possibility of finding the global optimum solution.

3.2 Local search strategy

Because the whale optimization algorithm approaches the prey according to the shrinking encirclement mechanism, and moves along the spiral path according to the spiral renewal mechanism, it can expand the local search range of whales. Inspired by the whale optimization algorithm, this paper improves the single update mechanism of grey wolf position, which makes the grey wolf individuals explore the surrounding solutions while moving towards the optimal solution, greatly expanding the range of local search. The mathematical description of the strategy is as follows:

$$\overrightarrow{X_{loc}} = \overrightarrow{D} * e^{bl} * \cos(2\pi l) + \overrightarrow{X} \quad (11)$$

$$\overrightarrow{D} = 1/3(\overrightarrow{D_\alpha} + \overrightarrow{D_\beta} + \overrightarrow{D_\delta}) \quad (12)$$

Where \overrightarrow{D} is the mean distance between grey wolf individual and α, β, δ , b is constant defining the shape of a logarithmic spiral, and l is the random number between $[-1, 1]$.

The local search strategy is to make the grey wolf move along the spiral path as well as within the shrinking circle. In order to simulate the two simultaneous behaviors, we assume that there is a 0.5 probability of the contraction encirclement mechanism or spiral update mechanism to choose to update the location of the grey wolves. The mathematical description of the strategy is as follows:

$$\overrightarrow{X}(t+1) = \begin{cases} 1/3(\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}) & p < 0.5 \\ \overrightarrow{X_{loc}} & p \geq 0.5 \end{cases} \quad (13)$$

Where p is the random number between $[0, 1]$.

3.3 Sub group cooperation strategy

In order to give full play to the advantages of random guidance strategy and local search strategy, expand the search space of the algorithm as much as possible, and guide algorithm jump out of the local optimum, this paper proposes a sub group cooperation strategy, the basic idea is as follows:

In the evolutionary process, the population is divided into three subgroups A, B and C according to the fitness values. A represents the subgroup with large fitness value, B represents the subgroup with medium fitness value, and C represents the subgroup with poor fitness value. The fitness value of grey wolf individuals in subgroup A is large, which indicates that the convergence degree is high and it is easy to fall into local optimum. In this case, grey wolf individuals update their positions according to formula (13), so that the individual can search around the extreme point more finely, find the position with better fitness than before, enhance the local search ability. The fitness value of grey wolf individuals in subgroup B is medium, which is updated according to the formulas (5) - (7) in the standard grey wolf algorithm. The fitness value of grey wolf individuals in subgroup C is poor, which is updated according to the formulas (8) - (10), and this subgroup can cover all possible solutions as much as possible by the random guidance strategy, so as to enhance the global search ability. The grey wolf individuals in subgroups A, B and C evolve according to their own update strategies, and each grey wolf individual migrates to the corresponding subgroup according to the new fitness value after every iteration until the termination condition is satisfied.

3.4 Multi Strategy Grey Wolf Optimizer algorithm (MSGWO)

In this paper, we combine the random guidance strategy, local search strategy, sub group cooperation strategy and the standard grey wolf algorithm to propose a Multi Strategy Grey Wolf Optimizer algorithm. In the continuous MSGWO, each individual can change its position to any point in the space. The purpose of this paper is to use MSGWO for feature selection, so the value of each dimension in the individual position is limited to 0 or 1. 0 means that the feature in the corresponding position is not selected, and 1 means that the feature in the corresponding position is selected. Thus, the updating formula of MSGWO for feature selection is as follows:

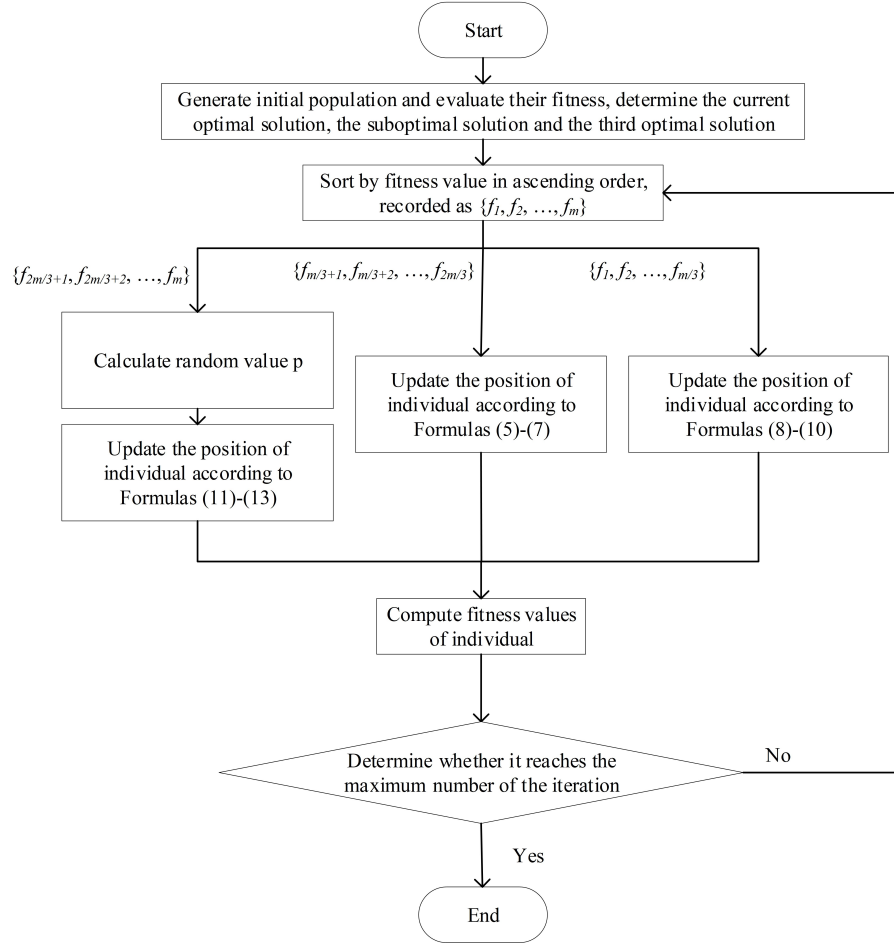
$$x_{in}^{k+1} = \begin{cases} 1 & \text{if } \text{sigmoid}(x_{in}^{k+1}) \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

sigmoid is defined as follows:

$$\text{sigmoid}(a) = 1/(1 + e^{-10(a-5)}) \quad (15)$$

In order to provide the more intuitive description of the MSGWO algorithm, we draw an algorithm flowchart of it. It is shown in Fig.1.

As can be seen from Fig.1, the local search strategy enables grey wolf individuals to make full use of the search space around the current best solution, which

**Fig. 1.** The flow chart of MSGWO

has good local search ability and helps to find more accurate solutions. For the random guidance strategy, we make full use of the random characteristics to improve the diversity of the population, thus enhancing the global search ability. In addition to the above two strategies, the sub group cooperation strategy is very important to balance the global search and local search in the iterative process. It not only guarantees the convergence speed of the algorithm, but also expands the search range of the population, and prevents the algorithm from falling into local stagnation in the later stage of the iteration. Due to the slow convergence speed in the early stage and the fast convergence speed in the later stage, the time complexity of the MSGWO is the same as that of the GWO.

4 Experiment

In order to prove the effectiveness of MSGWO proposed for feature selection in this paper, we evaluated the proposed algorithm using the Vehicle, Wine, Glass, Zoo, Landsat and Segment public datasets from UCI repository. Table 1 lists the details of the six datasets used in the evaluation. In order to prove the superi-

Table 1. Details of the six datasets used in the evaluation

No	Dataset	Number of samples	Dimension of samples	Number of categories
1	Vehicle	946	18	4
2	Wine	178	13	3
3	Glass	214	9	6
4	Zoo	101	16	7
5	Landsat	2000	36	6
6	Segment	2100	18	7

ority of MSGWO in feature selection, we used the classical K-Nearest-Neighbor (KNN) classifier and Support Vector Machine (SVM) as the benchmark classifier, and compared the performance on six public datasets of our proposed MSGWO method and several classical feature selection methods including analysis of variance (ANOVA) [12], principal component analysis (PCA) [13], PSO [2], WOA [3] and GWO, in which no feature selection is recorded as NFS. In the experiment, KNN and SVM adopt default parameters. For KNN, k -value is set to 5. For SVM, penalty parameter c is set to 1 and the gauss kernel is set to the kernel function. We carried on experiments by using five-folds cross validation for 5 times, and set the number of grey wolves in GWO, particles in PSO and killer whales in WOA were all set to 15. Table 2 and table 3 list the comparison of the average values of accuracy and F1 of five-folds cross validation on SVM and KNN, respectively.

From table 2 and table 3, we can see that the classification effect of NFS is very poor, which shows the necessity of feature selection. The performance of ANOVA and PCA is not high, because the filter selection method needs to

Table 2. Comparison of classification performance between various feature selection algorithms on SVM

Datasets	Metrics	NFS	ANOVA	PCA	PSO	WOA	GWO	MSGWO
Vehicle	Accuracy	0.633	0.709	0.639	0.825	0.819	0.825	0.846
	F1	0.584	0.695	0.592	0.822	0.827	0.836	0.849
Wine	Accuracy	0.613	0.894	0.736	0.993	0.992	0.992	1
	F1	0.554	0.892	0.711	0.994	0.996	0.997	1
Glass	Accuracy	0.784	0.846	0.776	0.921	0.927	0.920	0.940
	F1	0.747	0.829	0.738	0.915	0.918	0.924	0.943
Zoo	Accuracy	0.733	0.940	0.733	0.978	0.984	0.974	0.988
	F1	0.685	0.923	0.685	0.974	0.985	0.993	0.997
Landsat	Accuracy	0.696	0.822	0.710	0.868	0.870	0.874	0.885
	F1	0.673	0.819	0.689	0.863	0.868	0.877	0.893
Segment	Accuracy	0.885	0.926	0.879	0.950	0.953	0.955	0.970
	F1	0.889	0.926	0.881	0.949	0.951	0.956	0.964

Table 3. Comparison of classification performance between various feature selection algorithms on KNN

Datasets	Metrics	NFS	ANOVA	PCA	PSO	WOA	GWO	MSGWO
Vehicle	Accuracy	0.697	0.675	0.697	0.813	0.810	0.805	0.826
	F1	0.670	0.635	0.667	0.802	0.798	0.816	0.833
Wine	Accuracy	0.955	0.961	0.949	0.994	0.994	0.992	0.996
	F1	0.960	0.960	0.949	0.992	0.993	0.992	0.996
Glass	Accuracy	0.841	0.856	0.808	0.945	0.951	0.940	0.962
	F1	0.826	0.846	0.781	0.938	0.960	0.952	0.963
Zoo	Accuracy	0.928	0.962	0.919	0.987	0.997	0.991	1
	F1	0.915	0.953	0.906	0.993	1	0.996	1
Landsat	Accuracy	0.825	0.833	0.826	0.861	0.863	0.869	0.875
	F1	0.820	0.829	0.821	0.861	0.861	0.867	0.876
Segment	Accuracy	0.914	0.922	0.903	0.945	0.949	0.951	0.964
	F1	0.911	0.921	0.900	0.944	0.949	0.951	0.964

manually specify the number of selected features, and the most appropriate number of features is difficult to determine. The results of feature selection by heuristic method are ideal. The performance of standard GWO algorithm for different datasets is sometimes better than PSO and WOA, and sometimes lower than PSO and WOA. The MSGWO proposed in this paper is not only better than GWO, but also better than PSO and WOA for all the datasets listed in this paper in terms of accuracy and F1 value, which proves the effectiveness of MSGWO.

Because the heuristic method is a random search method, the results of each search may be different. In order to verify that the MSGWO method proposed in this paper can not only select the optimal feature combination, but also has good stability. In this paper, only the accuracy is analyzed. Fig.2 and Fig.3 show the box diagram of the accuracy of five times random experiments among four feature selection algorithms on SVM and KNN classifiers. From Fig.2 and Fig.3, it can be concluded that whether SVM or KNN is used as the base classifier, the maximum, minimum and average accuracy of MSGWO algorithm is superior to the other three heuristic methods, which further proves that MSGWO can effectively improve the performance of the classifier. In addition, by introducing the

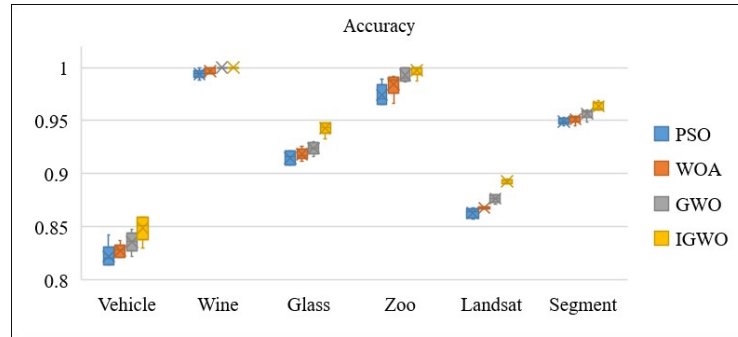


Fig. 2. Comparison of the accuracy of heuristic feature selection methods on SVM

standard deviation to further analyze the stability of the algorithm, the standard deviations of the accuracy of five times random experiments are recorded in Table 4. According to table 4, it can be seen that accuracy of MSGWO is relatively stable and the standard deviation is the smallest in almost all datasets, which proves that MSGWO is stable.

5 Conclusion

In this paper, we propose a Multi Strategy Grey Wolf Optimizer algorithm based on random guidance, local search and sub group cooperation strategies for feature selection. In MSGWO, the search agent updates its position through the

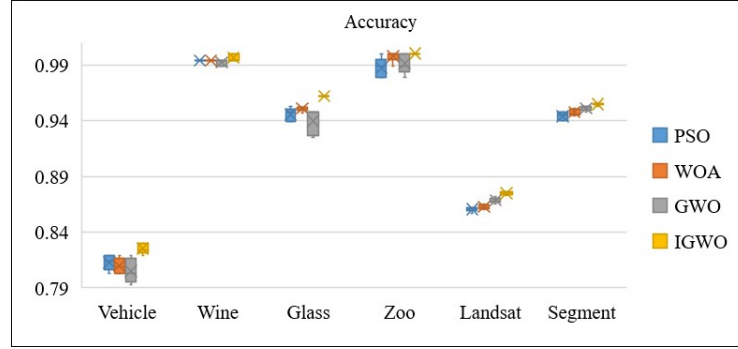


Fig. 3. Comparison of the accuracy of heuristic feature selection methods on KNN

Table 4. Comparison of classification performance between various feature selection algorithms on SVM and KNN

Datasets	SVM				KNN			
	PSO	WOA	GWO	MSGWO	PSO	WOA	GWO	MSGWO
Vehicle	1.03E-2	5.84E-3	8.51E-3	4.40E-3	6.41E-3	6.24E-3	9.88E-3	4.17E-3
Wine	3.93E-3	2.99E-3	0	0	0	0	2.44E-3	2.44E-3
Glass	6.07E-3	6.11E-3	5.17E-3	4.26E-3	5.53E-3	2.00E-3	1.03E-2	0
Zoo	8.90E-3	9.48E-3	6.31E-3	5.11E-3	7.80E-3	4.40E-3	7.91E-3	0
Landsat	4.13E-3	5.85E-3	3.30E-3	1.61E-3	1.73E-3	1.73E-3	2.19E-3	1.16E-3
Segment	3.87E-3	3.28E-3	3.80E-3	2.74E-3	3.40E-3	2.65E-3	1.78E-3	9.80E-4

cooperation of three search strategies, which improves the global and local search ability of the algorithm. MSGWO not only retains the advantages of the fast convergence speed of GWO algorithm, but also makes full use of characteristics of various search strategies, and balances the global and local search ability, which makes it easy to find the optimal feature subset. In this paper, we use a variety of feature selection methods on six public datasets for comparative experiments. The results show that our MSGWO feature selection method can improve the accuracy of search, find the optimal solution, and is an efficient and reliable algorithm.

References

1. J.Holland.: *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI (1975)
2. Bratton,D., Kennedy,J.: Defining a Standard for Particle Swarm Optimization. In: *Proceedings of the IEEE Swarm Intelligence Symposium*, pp. 1–5. Honolulu, HI, USA (2007)
3. Mirjalili., Seyedali., Lewis., Andrew.: The Whale Optimization Algorithm. *Advances in Engineering Software* **95**,51–67 (2016)
4. Mirjalili,S., Mirjalili,S,M., Lewis,A.: Grey Wolf Optimizer. *Advances in Engineering Software* **69**(3),46–61 (2014)
5. E,Emery., H,Zawbaa., A,Hassanien.: Binary grey wolf optimization approaches for feature selection. *Neurocomputing* **172**,371–381 (2016)
6. Tu,Q., Chen,X., Liu,X.: Multi-strategy ensemble grey wolf optimizer and its application to feature selection. *Applied Soft Computing* **76**,16–30 (2019)
7. Emary,E., Yamany,W., Hassanien,A,E.: Multi-objective gray-wolf optimization for attribute reduction. *Procedia Computer Science* **65**,623–632 (2015)
8. Elhariri,E., El-Bendary,N., Hassanien,A,E.: Grey wolf optimization for one-against-one multi-class support vector machines. In: *7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, pp. 7–12. Kyushu Univ, Ohashi Campus, Fukuoka, JAPAN (2015)
9. Heidari,A,A., Pahlavani,P.: An efficient modified grey wolf optimizer with Levy flight for optimization tasks. *Applied Soft Computing* **60**,115–134 (2017)
10. ZHU,A., XU,C., LI,Z.: Hybridizing grey wolf optimization with differential evolution for global optimization and test scheduling for 3D stacked SoC. *Journal of Systems Engineering and Electronics* **26**(2),317–328 (2015)
11. WEI,Z., ZHAO,H., HAN,B.: Grey wolf optimization algorithm with self-adaptive searching strategy. *Computer Science* **44**(3),259–263 (2017)
12. DAI,J., YUAN,J.: The comparison of test methods between single factor analysis of variance and multiple linear regression analysis. *Statistics and Decision* **09**,23–26 (2016)
13. LI,J.: Combination of feature extraction in text classification algorithm based on PCA. *Application Research of Computers* **30**(08),2398–2401 (2013)