Genetic and Hybrid Gray Wolf Optimization Algorithm

Noor Muhammed Noori[1], Omar Saber Qasim[2]

[1]M.Sc. Student, [2]Assist. Prof., Department of Mathematics, University of Mosul, Mosul, Iraq

Abstract

Much of the data in the classification issue contains a number of additional attributes that do not affect the accuracy of the classification. There are many evolutionary algorithms that are used to define the feature and reduce dimensional patterns such as the gray wolf algorithm (GWO) after converting it from a continuous space to a discrete space. In this research, a method of feature selection was proposed through two consecutive stages in the first stage, the mutual information (MI) method is used to determine the most important feature selection. In the second stage, the binary gray wolf optimization (BGWO) algorithm is used to lessen and determine a specific number of features affecting the process of classification, which came from the first stage. The proposed algorithm MI_BGWO is efficient and effective by obtaining higher classification accuracy and a small number of specific features compared to other competing algorithms.

Keywords: *Gray wolf optimization; classification; feature selection; mutual information.*

Introduction

Gray wolf optimization (GWO) is a swarm intelligent technique developed by Mirjalili et al., 2014[1]. The hunting techniques and the social hierarchy of wolves are mathematically modeled in order to develop GWO and perform optimization. The GWO algorithm is tested with the standard test functions that indicate that it has superior exploration and exploitation characteristics than other swarm intelligence techniques [2].Scientists have developed the basic algorithm GWO by converting the search algorithm from a continuous search space to a discrete search space. This modified algorithm is called a binary gray wolf optimization (BGWO) algorithm, which works in binary search areas and uses binary values equal to 0 or 1[3].Mutual information (MI) is one of the filtering method,which is calculated between two random variables using entropy. Entropy measures an average random variable the amount of information required to describe the random variable[7].In this study, a new MI_BGWO algorithm is suggested to determine the best-feature selection. Our proposed algorithm can efficiently achievement the powerful points of both MI and BGWO algorithms in finding the most important feature. The experimental results show the excellent performance of the proposed algorithm when the number of feature is high, and the sample size is low. The GWO is presented in Section 2. In Sections 3 and 4, a brief description of feature selection and MI respectively. In Section 5, the proposed algorithm is explained. Section 6 covers the obtained results and their discussion. Finally, the most important general conclusions are mentioned, in section 7.

Grey wolf optimization (GWO): The gray wolf (Canis lupus) is part of the Canidae family. Gray wolves are types with a very rigorous social dominant hierarchy of leadership. Males and females in the pack are leaders, and they are called alpha[1]. Alpha wolf is also known as the dominant wolf because all its orders must be followed by all wolves in the pack. Only alpha wolves are allowed to mate in the pack. This means that the order of the pack and its regularity is more important than its strength. Beta is the second level in the hierarchy of gray wolves. Beta are wolves that help alpha wolves make decisions or other things about the pack. The wolf beta is the best

Corresponding Author:
Noor Muhammed Noori
Department of Mathematics, University of Mosul, Mosul, Iraq
e-mail: noor muhammednoori@uomosul.edu.iq

candidate to be alpha in the case if one of the wolves of alpha becomes too big or in the case of the death of someone in command. Omega is less gray wolves. The omega wolf always offers all the other dominant wolves on it. Omega acts as a scapegoat. The last wolves allowed to eat are omega. We may apparently notice that omega is not an important member of the group, but in fact, if omega is lost, the entire pack will face problems and internal fighting. Delta is the fourth class in the pack. Delta wolves control omega, but they must undergo alpha and beta orders. Delta shall be responsible for guards, scouts, hunters, elders, and patients and each has its own specific responsibility[8].

Mathematical Modelling: The first best solution in GWO is alpha (α), the second-best solution is beta (β), and the third-best solution is Delta (δ). While the rest of the solutions are Omega (ω), which follows the rest of the first three solutions.

Encircling Prey:: Gray wolves surrounded prey while hunting. The following equations will represent a mathematical model about the surrounding behavior:

$$\vec{D} = \left|\vec{C} \cdot \overrightarrow{X_p}(t) - \vec{X}(t)\right| (1)$$

$$\vec{X}(t+1) = \overrightarrow{X_p}(t) - \vec{A} \cdot \vec{D} \quad (2)$$

where t is the iteration, $\overrightarrow{X_p}$ is the prey position, $\vec{X}$ is the gray wolf position, the operator indicates vector entry wise multiplication.

where $\vec{A}$ and $\vec{C}$ are coefficient vectors calculated as follows:

$$\vec{A} = 2\vec{a} \cdot \overrightarrow{r_1} - a \quad (3)$$

$$\vec{C} = 2\overrightarrow{r_2} \quad (4)$$

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].

Hunting: Gray wolves have the ability to encircle prey and locate them. Hunting is performed by a complete pack based on, Information from the alpha, beta and delta wolves,So the updating for the wolves positions is as in the following equations:

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

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

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

The updating of the parameter $\vec{a}$ through the following equation:

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

where *t* is the iteration number and Max_{iter} is the total number of iteration allowed for the optimization 9. The pseudo code of the GWO algorithm is displayed in Figure 1.

Start

Generate an initial population of the gray wolf $\boldsymbol{X_i}$ *(i=1.....n)*

Generate an initial value A, a, and c

Find the fitness function of each search agent

x_α *= the first best search agent*

x_β *= the second-best search agent*

x_δ *= the third best search agent*

while *(t <* $\boldsymbol{Max_{iter}}$ *of iterations).*

for *each search agent*

Update the position of the current search agent by equation (7)

end for

Update A, a, and c

Find the fitness function of each search agent

Update x_α, x_β *and* x_δ

set t = t + 1

end while

return x_α

End

Figure 1: The pseudo code of the GWO algorithm.

Binary gray wolf optimization (BGWO): The positions of gray wolves constantly change in space to any point. Solutions are limited to binary values such as 0 or 1 in some special problems such as feature selection. In our research, we suggested feature selection through a binary GWO algorithm. Updating the equations of wolves are a function of the positions of three vectors is X_α, X_β and X_δ, which represents the best three solutions and which attracts each wolf towards it.At any given time, all solutions are at a corner of a hypercube and the solutions are grouped in binary form.Based on the basic GWO algorithm, the given wolf positions are updated, a binary restriction must be maintained according to equation (9).

The main update equation in the GWO algorithm is written as[3]:

$$X_i^{t+1} = Crossover(X_1, X_2, X_3) \quad (9)$$

where (X_1, X_2, X_3) they are binary vectors representing wolves in bGWO and $crossover(a, b, c)$ are an appropriate intersection between solutions (a, b, c) and (X_1, X_2, X_3), and the wolves calculate alpha, beta and delta in order using equations (10), (13) and (16).

$$X_1^d = \begin{cases} 1 & if \ (X_\alpha^d + bstep_\alpha^d) \geq 1 \\ 0 & other\ wise \end{cases} \quad (10)$$

where $bstep_\alpha^d$ is a binary step in dimension d, X_α^d is a vector representing the position of the alpha wolf in dimension d . $bstep_\alpha^d$ is calculated by the following equation:

$$bstep_\alpha^d = \begin{cases} 1 & if \quad cstep_\alpha^d \geq rand \\ 0 & other\ wise \end{cases} \quad (11)$$

where $cstep_\alpha^d$ is the continuous-valued step size for dimension d, $rand$ is a random number in the closed period $[0,1]$ derived from the uniform distribution . $cstep_\alpha^d$ can be calculated using the sigmoidal function by the following equation:

$$cstep_{\alpha}^{d} = \frac{1}{1 + e^{-10(A_1^d D_{\alpha}^d - 0.5)}} \quad (12)$$

where A_1^d and D_{α}^d are calculated using equations (3), and (5) in the dimension *d*.

X_2^d and X_3^d Can be found by the following equations:

$$X_2^d = \begin{cases} 1 & if \ (X_{\beta}^d + bstep_{\beta}^d) \geq 1 \\ 0 & other\ wise \end{cases} \quad (13)$$

$$bstep_{\beta}^{d} = \begin{cases} 1 & if \ cstep_{\beta}^d \geq rand \\ 0 & other\ wise \end{cases} \quad (14)$$

$$cstep_{\beta}^{d} = \frac{1}{1 + e^{-10(A_1^d D_{\beta}^d - 0.5)}} \quad (15)$$

$$X_3^d = \begin{cases} 1 & if \ (X_{\delta}^d + bstep_{\delta}^d) \geq 1 \\ 0 & other\ wise \end{cases} \quad (16)$$

$$bstep_{\delta}^{d} = \begin{cases} 1 & if \ cstep_{\delta}^d \geq rand \\ 0 & other\ wise \end{cases} \quad (17)$$

$$cstep_{\delta}^{d} = \frac{1}{1 + e^{-10(A_1^d D_{\delta}^d - 0.5)}} \quad (18)$$

In the following equations, we will apply the intersection to each of the solutions *a*, *b*, *c*:

$$X_d = \begin{cases} a_d & if \ rand < 1/3 \\ b_d & if \ 1/3 \leq rand < 2/3 \\ c_d & other\ wise \end{cases} \quad (20)$$

where $rand$ is a random number derived from the uniform distribution in the closed period $[0,1]$, and a_d, b_d and c_d represents the binary values of the first, second and third parameters of the dimension d, X_d is the result of the exchange in dimension d. The pseudo code of the BGWO algorithm is displayed in Figure 2 [10]:

Start

Generate an initial population of the grey wolf X_i *(i=1.....n)*

Generate an initial value A, a, and c

Find the fitness function of each search agent

x_{α} *= the first best search agent*

x_{β} *= the second-best search agent*

x_{δ} *= the third best search agent*

while (t < Max_{iter} *of iterations).*

for each search agent

Calculate $x_1; x_2; x_3$ *using equations (10), (13), and (16)*

Apply the crossover method among $x_1; x_2; x_3$ *using the equation (9)*

end for

Update A, a, and c

Find the fitness function of each search agent

Update x_{α}, x_{β}, *and* x_{δ}

set t = t + 1

end while

return x_{α}

End

Figure 2: The pseudo code of the BGWO algorithm.

Feature Selection: The feature selection method improves the performance of the algorithm by reducing the number of attributes used to describe a data set[13]. The purpose of features selecting in an algorithm is to reduce the number of genes used to improve classification and increase classification accuracy[14]. The feature identification algorithms consist of three parts:

1. **Search algorithm:** A subset of properties (features), which are part of the original features.
2. **Fitness function:** These input and digital assessment modes. The goal of the search algorithm is to draw attention to this function.
3. **Classifier:** It represents the required algorithm that uses the latest subset of genes (i.e., an algorithm that selects the most important features required)[15,16].

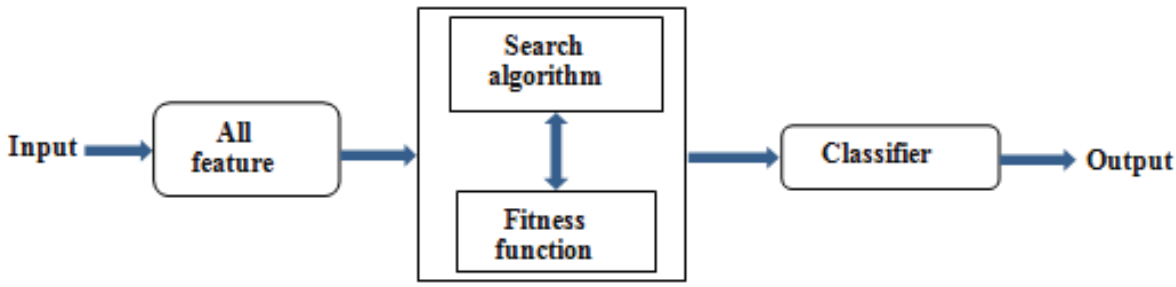

Figure 3: Filter control strategy

One of the ways to select a feature on the predictive performance of a predefined algorithm is the wrapper method, which evaluates the quality and efficiency of certain features[17]. This is based on two-steps, based on a specific algorithm: (1) Searching for a subset of properties and features (2) Evaluate these specific features and attributes. In order to obtain learning performance or to reach some stop criteria we repeat (1) and (2)[11].

1. **Mutual information (MI):** You can get some information about the variable *Y* that is included in the mutual information *I*(*X*, *Y*) which is considered to be in the information shared by two random variables.

$$I(X,Y) = \frac{H(X)}{H(\frac{X}{Y})} \quad (23)$$

If there is a close association between the two variables *X* and *Y*, the mutual information *I*(*X*, *Y*) will be large[18]. If the variables *X* and *Y* are not completely associated, then (*X*, *Y*) = 0. The mutual information was applied to filter the feature set to see the relationship between certain features and category classifications. In the theory of information, there is a classic use of many measures of feature classification. Note that these statistics work to build a relationship with the classroom is consist of data and information register each feature and feature in F_i[19]. In the theory of information, one of the most important contributions to the research is the selection of the feature where it works to use the information exchanged in the evaluation of the feature and in the following formula *F* will be indicated by a set of features is referred to and the class naming[20].

$$I(F,C) = \iint P(F,C) log \frac{P(F,C)}{P(F) * P(C)} dfdc \quad (24)$$

The evaluation of mutual information (MI) is in some method between class naming and one feature[17]. The problem is not in this measure. When evaluating entire feature sets will have difficulties arise. Possible interactions between features are the reason for evaluating entire feature sets in a multivariate way. The combination of two features may in some cases provide important information, because in some cases two individual features may not provide sufficient information about the chapter[11]. Between the variable Y and N variables X1, X2 …XN, the chain rule is

$$I(X1,X2,\ldots..Xn;Y) = \sum_{i=1}^{N} I(Xi,{}^{Y}/_{Xi-1,Xi-2,\ldots..X1}) \quad (25)$$

$$fitness(Xi) = I(Xi,C) \quad (26)$$

In the measurement of entropy, there is a way to calculate and replace mutual information in the form of mutual information.There is an appropriate criterion for selecting items through information exchange. It is possible to define the information exchanged is a measure to reduce the uncertainty about category labels, because of knowledge of characteristics and characteristics of the data set, where the fitness function is maximizing the mutual information value[12] **Experimental results**

The proposed algorithm MI_BGWO is evaluated, and its interest is compared with the other competitor algorithms.

Data Sets: We have selected (3) different classification problems from the literature to verify the effectiveness of the proposed algorithm for classification problems. From a repository UCI, a data set was obtained[21].

The target variable is a binary variable that includes a set of data where the binary variable represents the condition of the sick person who has good = 1 and bad = 0.

The following table shows the overall description of the data set:

Table 1: Description of the used datasets.

Dataset	# Samples	# Features	Target class
Prostate	102	12601	2
DLBCL	77	7130	2
Colon	62	2001	2

Discussion

To correctly evaluate the proposed algorithm MI-BGWO, the results were compared with the BGWO algorithm for feature selection. The training and testing dataset for the proposed algorithm, MI-BGWO, achieved the best results for the classification. For instance, in the prostate dataset, the CA of the testing dataset is 94% by the MI-BGWO which is higher than 65% by BGWO.

Table 2: Classification performance on average of the algorithms used over 20 partitions (wherethe number in parentheses is the standard error).

Datasets	Method	Training dataset				Testing dataset
		CA	SE	SP	MCC	CA
Prostate	MI-BGWO	0.9426 (0.0176)	0.9663 (0.0211)	0.9207 (0.0324)	0.8864 (0.0336)	0.9324 (0.0417)
	BGWO	0.6529(0.0348)	0.7901(0.0474)	0.6029(0.0512)	0.3483(0.0478)	0.7382 (0.0765)
DLBCL	MI-BGWO	0.9192 (0.0199)	0.9914 (0.0138)	0.7674 (0.0398)	0.8158 (0.0447)	0.9400 (0.0432)
	BGWO	0.9000(0.0469)	0.9009(0.1060)	0.9058(0.0590)	0.7231(0.1396)	0.9640(0.0515)
Colon	MI-BGWO	0.8619 (0.0623)	0.7692(0.1079)	0.9361 (0.0431)	0.7245 (0.1179)	0.9100 (0.0699)
	BGWO	0.7429(0.0446)	0.5833(0.0548)	1(0)	0.5918(0.0571)	0.8750(0.1161)

When comparing the BGWO algorithm with the proposed algorithm MI-BGWO we note that the proposed algorithm MI-BGWO has a clear advantage in terms of accuracy and classification capacity and BGWO worse than MI-BGWO through the three datasets.

Conclusion

In this study, the MI_BGWO method was proposed between the MI technique and the BGWO to improve the classification performance (when the datasets are big), relying on subsets of important features of the dataset. After the feature selection process, datasets were submitted to the Naïve Bayes classifier, and the results of the MI_BGWO method were compared with the BGWO method by the criteria shown in Table 2. Experimental results from the feature selection dataset indicate that the MI_BGWO method has a better classification performance than the BGWO method and has few features compared to the default method.

Financial Disclosure: There is no financial disclosure.

Conflict of Interest: None to declare.

Ethical Clearance: All experimental protocols were approved under the University of Mosul and all experiments were carried out in accordance with approved guidelines.

References

1. S Mirjalili, SMMirjalili, ALewis, "Grey wolf optimizer," Adv. Eng. Softw., 2014;69:46–61.
2. AMadadi, MMMotlagh, "Optimal control of DC motor using grey wolf optimizer algorithm," Tech J Eng Appl Sci, 2014;4(4): 373–379.
3. EEmary, HMZawbaa, AE. Hassanien, "Binary grey wolf optimization approaches for feature selection," Neurocomputing, 2016;172: 371–381.
4. MZaffar,S Iskander, MAHashmani, "A study of feature selection algorithms for predicting students academic performance," Int. J. Adv. Comput. Sci. Appl, 2018;9(5): 541–549.
5. L-Y Chuang, C-HYang, J-C. Li, "Chaotic maps based on binary particle swarm optimization for feature selection," Appl. Soft Comput., 2011;11(1):239–248,.
6. MDash, HLiu, "Feature selection for classification," Intell. data Anal., 1997;1(1–4) :131–156.
7. PA Estévez, MTesmer, CAPerez, JMZurada, "Normalized mutual information feature selection," IEEE Trans. Neural Networks, 2009;20(2):189–201.
8. N Singh, HHachimi, "A new hybrid whale optimizer algorithm with mean strategy of grey wolf optimizer for global optimization," Math. Comput. Appl., 2018;23(1):14.
9. EEmary, HMZawbaa, CGrosan, "Experienced gray wolf optimization through reinforcement learning and neural networks," IEEE Trans. neural networks Learn. Syst., 2017;29(3):681–694.
10. SPManikandan, RManimegalai, MHariharan, "Gene Selection from microarray data using binary grey wolf algorithm for classifying acute leukemia," Curr. Signal Transduct. Ther., 2016;11(2):76–83.
11. J Chen, HHuang, S Tian, YQu, "Feature selection for text classification with Naïve Bayes," Expert Syst. Appl., 2009;36(3): 5432–5435.
12. BLiu, EBlasch, Y Chen, DShen, G Chen, "Scalable sentiment classification for big data analysis using naive bayes classifier," in 2013 IEEE international conference on big data, 2013;99–104.
13. P Ghamisi, JA Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," IEEE Geosci. Remote Sens. Lett., 2014; 12(2):309–313.
14. AJain, DZongker, "Feature selection: Evaluation, application, and small sample performance," IEEE Trans. Pattern Anal. Mach. Intell., 1997; 19(2): 153–158.
15. AG Karegowda, MA Jayaram, AS Manjunath, "Feature subset selection problem using wrapper approach in supervised learning," Int. J. Comput. Appl., 2010;1(7): 13–17.
16. L-YChuang, H-WChang, C-J. Tu, C-H. Yang, "Improved binary PSO for feature selection using gene expression data," Comput. Biol. Chem., v;32(1,):29–38.
17. S-BKim, K-SHan, H-C Rim, S HMyaeng, "Some effective techniques for naive bayes text classification," IEEE Trans. Knowl. Data Eng., 2006;18(11): 1457–1466.