Acta Scientific COMPUTER SCIENCES

Volume 3 Issue 4 April 2021

Research Article

Classification of Thalassemia Data Using a Modified Fuzzy Inference System Model by Black Hole Algorithm

Fatima Hashim Najim* and Omar Saber Qasim

Department of Mathematics, University of Mosul, Mosul, Iraq

***Corresponding Author:** Fatima Hashim Najim, Department of Mathematics, University of Mosul, Mosul, Iraq.

Received: February 26, 2021

Published: March 24, 2021

Abstract

Recently, many methods and algorithms have been created for solving various classification problems in order to obtain efficient and accurate datasets classification. Many algorithms inspired by nature are used in building a classification method, and one of these important algorithms is the black hole algorithm (BHA). In this study a proposed method was presented to improve the fuzzy inference system (FIS) using the BHA, where the BHA modifies the parameters of the membership functions used in the fuzzification step and finds the best positions for them that increase the accuracy of the classification in thalassemia dataset. The proposed BHA-FIS algorithm demonstrated efficacy and efficiency compared to the standard algorithm FIS.

Keywords: Fuzzy Inference System (FIS); Black Hole Algorithm (BHA); Classification

Introduction

Classification is one of the most important implementation of machine learning, which purpose to classify the datasets into different classes, and when classify a set of data using modern methods, we expect to obtain a better result according to the available classifiers. Fuzzy Inference system (FIS) method is one of the best methods tools that are widely used in several applications such as bioinformatics, and medical datasets, and it contains a number of important operations that influence the performance of the classification [1,2].

Recently, metaheuristics algorithms have been shown a competitive result when solving optimization problems including parameters determination of membership functions in FIS. black hole algorithm (BHA) is one of these algorithms which was used to select optimum parameters of membership functions which leading to improve the classification performance. Our proposed algorithm, BHA-FIS, tunes the parameters values through the BHA, which performs to perform classification accuracy in FIS.

The remainder of this paper is organized as follows. FIS is presented in Section 2. BH is presented in Section 3. The BH-AFIS algorithm is explained in Section 4. Section 5 covers the experimental results. Finally, in Section 6, the conclusions are mentioned.

Fuzzy inference system (FIS)

This concept was originated in (1965) by the scientist of Azerbaijani origin, Lotfi Zadeh, from the University of California, who developed it to use it in a better way to process data [3]. Fuzzy logic is based on the existence of a function whose value at a certain element is a real value located between [0,1], which represents the percentage of belonging to a group, and if the value of the function is (1), then this means that the element belongs completely to the group and if the value is (0) It means that the element does not belong to the group, but if the value of the function lies between [0,1], this indicates the partial belonging of the element in this group [4].

Given that the fuzzy logic system does not deal with the boundary (Crisp), it is considered a powerful and simple system at the same time, as it is not limited by the number of inputs to the system, but rather by the nature of understanding the system's behavior through which the user can set the conditions for the applied problem in a way that is consistent with the state of the problem, which gives ease and flexibility in Improve the performance of this Fuzzing system [5].

If X represents the comprehensive set, then the fuzzy set A, the tribute of X, is the set of ordered pairs, defined by the following formula [6]:

$$A = \{x, \mu_A(x)\} \ldots\ldots (1)$$

For each element x in X, where $\mu_A(x)$ is the membership function of element x in A and that:

$$\mu_A(x) \in [0,1] \ldots\ldots (2)$$

Membership functions

Membership functions are one of the foundations on which the fuzzy logic system is built, and through these functions the link between the linguistic issue, which is represented by these functions and the mathematical degrees called Membership degrees, symbolized by μ (x) for the values x, and there is a group of Membership functions Which are dealt with in fuzzy logic and are selected accurately and not randomly depending on the type of issue dealt with, and the most important of these functions [7].

$$\mu(x) = \begin{cases} 0 \; ; \; x \le a \\ \dfrac{x-a}{b-a} \; ; \; a \le x \le b \\ \dfrac{c-x}{c-b} \; ; \; b \le x \le c \\ 0 \; ; \; c \le x \end{cases} \ldots\ldots (3)$$

where a , b, and c are parameters of the function.

Gaussian function

This function is expressed as follow:

$$\mu(x) = e^{-\frac{(x-c)^2}{2\sigma^2}} \ldots\ldots (4)$$

Where c and σ represent, respectively, the upper vertex of the figure and the distance of the edges from the center of the figure.

$$\mu(x) = \begin{cases} 0 \; ; \; x \le a \\ \dfrac{x-a}{b-a} \; ; \; a \le x \le b \\ 1 \; ; \; b \le x \le c \\ \dfrac{d-x}{d-c} \; ; \; c \le x \le d \\ 0 \; ; \; d \le x \end{cases} \ldots\ldots\ldots (5)$$

where a, b, c and d are parameters of the function.

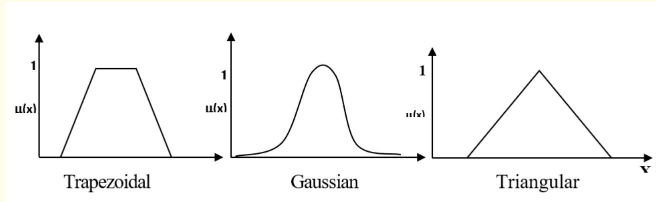

Figure 1: Represents a diagram of all membership functions.

The implementation of the type of function depends on the type of problem and the datasets used [8].

The basic stages of building the fuzzy model

The fuzzy model, like any branch of artificial intelligence, passes through several stages in order to complete its construction and formation, which can be summarized as follows [6].

Fuzzification

It is the first stage and the operations of fuzzy logic, as it is the basis stage through which the processes of converting the problem of the linguistic formula to the mathematical formula are carried out by dropping the input variable x on the organic functions and choosing the appropriate organic degree, meaning that during this stage the marginal inputs are converted (fragile). To fuzzy entries with a degree of membership ranging from zero to one.

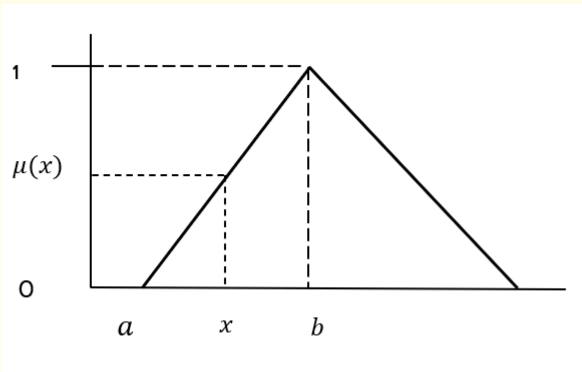

Figure 2: Represents the input fuzzing process.

Rule evaluation

The second stages of the fuzzy logic representation represent the evaluation of the rule that represents the possible possibilities of the issue or application, as it is developed by an expert in the applied problem and is in the form of easy-to-understand linguistic expressions, as it represents the heart of the fuzzy logic and is represented by the following general formula:

IF condition THEN result

Whereas, the expressions following IF represent the condition and are the parameters of Membership functions or (input after converting them into mathematical expressions) and the expressions that follow THEN represent the result or (output), and the representation of the result is in one of the two cases based on the inference mechanism used.

There are two types of inference mechanism, the first is called Mamdani and the second is called Sugeno, as these two methods are not chosen randomly, but rather depending on the issue used, Where in this paper, Sugeno type inference was used [9].

Defuzzification

It represents the last stage of the stages and processes of fuzzy logic, in which the values of the problem are converted to the crisp state after the completion of the fuzzing and programming it language.

Black Hole algorithm (BH)

The BH algorithm is a population-based method that has some advantages in common with other population-based methods. As with other population-based algorithms, a set of candidate solutions to a specific problem are generated and distributed randomly in the search space. Population-based algorithms work to optimize the solution-generated population via specific mechanisms [10].

In the BH algorithm, the community is developed by moving all of the candidates towards (stars) the best candidate in each iteration which is the black hole (best candidate) and replacing those candidates who enter the scope of the black hole by newly created candidates in the search space [10].

Where the fitness value is calculated for the black hole algorithm

$$f_i = \sum_{i=1}^{pop\_size} eval(p(t)) \quad \dots (6)$$

$$f_{BH} = \sum_{i=1}^{pop\_size} eval(p(t)) \quad \dots (7)$$

where f_i and f_{BH} are the fitness values of black hole and ith star in the initialized population. The population is estimated and the best candidate in the population, which has the best fitness value, $f_I$is selected to be the black hole and the remaining form the normal stars. The black hole has the capability to absorb the stars that surround it. After initializing the first black hole and stars, the black hole starts absorbing the stars around it and all the stars start moving towards the black hole [11].

Where the movement of the stars is as in the following equation:

$$x_i^{k+1} = x_i^k \times \theta\left(x_{BH} - x_i^k\right), i = 1,2,3, \dots \dots . N \quad \dots (8)$$

x_i^{k+1} and x_i^k positions of i stars in the frequency k and $k+1$

x_{BH} : Location (BH)in space , θ: random number located in the period [0.1]

and N: number of stars.

After the star moves to the new position (towards the black hole), if its fitness value is better than Black hole, the star is chosen as the black hole. Then this algorithm continues with the black hole in the new location and the stars begin to move towards this new black hole.

Also, there is the possibility of crossing an event horizon (the star's distance with the black hole) while the stars are moving Toward a black hole.

The filtered solution (star) crossing the event horizon of the black hole will be swallowed up.

Black hole. Then a new star is created that follows the swallowing star and is randomly distributed in the search space [10,12].

The radius of the event horizon (R) is formulated as follows [10]:

$$R = \frac{f_{BH}}{\sum_{i=1}^{N} f_i} \quad ...(9)$$

Where f_{BH} Is the fitness value of a black hole.

f_i It is the fitness value of a star, and N: number of stars.

When the distance between the filtered solution and the black hole (best filter) is less than R, then this filter is evanescent and a new filter is created and randomly distributed in the search space [10].

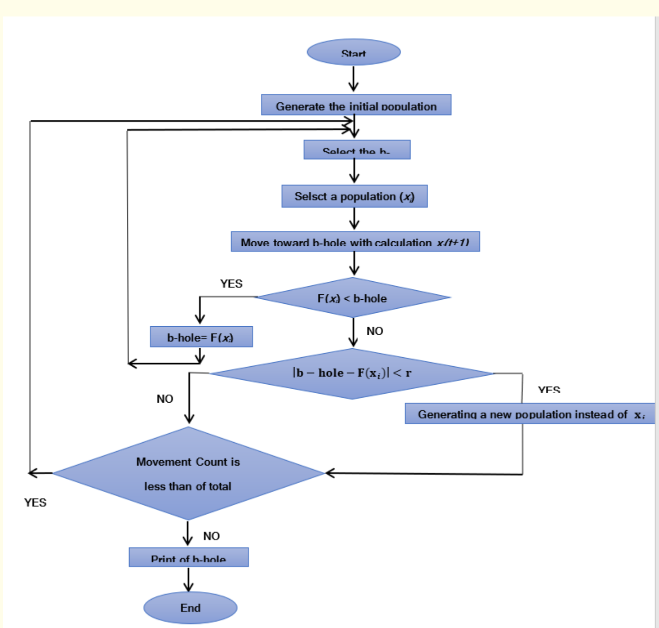

Figure 3: The Flowchart of BHA.

The proposed algorithm is BHA-FIS

The proposed BHA-FIS method consisted of two basic stages. In the first stage, a black hole (BH) algorithm is used to determine the optimum parameters of the fuzzy inference system (FIS) in the Sugeno model. In the second stage, the FIS model is constructed according to the optimal parameters found by the BH algorithm, where these optimal parameters are entered into the inference system and the ideal model is formed according to the data set used. The proposed BH-AFIS method was tested on a set of Thalassemia data and classified according to mean square error (MSE) criteria.

Step1: Set the initial parameters of FIS and BH algorithm. **Step2:** Create FIS of Sugeno model using optional parameters.
Step 3: Evaluate dataset by fitness function
Step 4: Set the location and radius of the event horizon using, Eq. (6) and Eq. (7)
Step 5: Set iteration i from 1 to max of iteration.
Step 6: Update location and radius of the event horizon according to Eq. (6) and Eq. (7).
Step 7: When i≤ Max_iteration stop satisfied and return get the best parameters of FIS. **Step 8:** Insertion of the optimal parameters of FIS into the Sugeno model.
Step 9: Calculate the MSE criterion for the proposed BHA-FIS method.

Algorithm 1: Steps of the proposed BHA-FIS algorithm.

Experimental results

The proposed BH-FIS algorithm is evaluated, and its results compared with the FIS model. In order to verify the effectiveness of the BH-FIS algorithm for solving classification problems, we selected datasets of thalassemia patients that are based on actual age measured by month, erythroid , and number of blood units.

Dataset	# Samples	# Features
Data=Thalassemia	150	3

Table 1: Description of the used data.

The training and test dataset of the BH-FIS algorithm achieved the best MSE, which are as follows:

Data	Methods	MSE of Training	MSE of Testing
Thalassemia	BH-FIS	771.8766	779.0312
	FIS	780.3089	1.0341e+03

Table 2: A comparison between the FIS and BH-FIS in the MSE when used the Gaussian membership function.

The execution comparison shows that, compared to the FIS model, the proposed algorithm, BH-FIS, has an obvious advantage in terms of MSE of the classification and the FIS is worse than BH-FIS through the datasets.

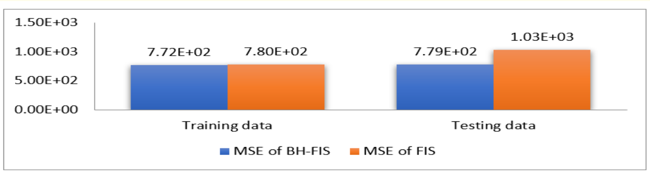

Figure 4: Represents the comparison of MSE between BHA-FIS and FIS in Thalassemia dataset using the Gaussian membership function.

In table 2, a Gaussian membership function is used, and each variable has three membership function when the number of membership function for each variable change and the type of function used changes, the implementation comparison shows that compared to the FIS model, the proposed algorithm, BH-FIS, has a clear advantage in terms of MSE for classification and that FIS is worse than BH-FIS through data sets. As shown in tables 3 and 4.

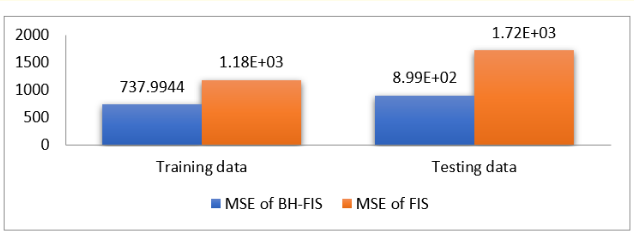

Figure 5: Represents the MSE comparison between BHA-FIS and FIS in thalassemia data using the Z-shaped membership function.

Data	Methods	MSE of Training	MSE of Testing
Thalassemia	BH-FIS	737.9944	899.3675
	FIS	1.1794e+03	1.7227e+03

Table 3: A comparison between the FIS and BH-FIS in the MSE when used the Z-shaped membership function.

Data	Methods	MSE of Training	MSE of Testing
Thalassemia	BH-FIS	518.0301	655.5197
	FIS	894.1779	1.0406e+03

Table 4: A comparison between the FIS and BH-FIS in the MSE when used the S-shaped membership function.

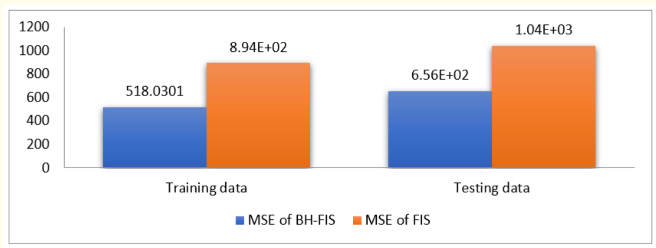

Figure 6: Represents the MSE comparison between BHA-FIS and FIS in thalassemia data using the S-shaped membership function.

Conclusion

In this paper, we proposed a BHA-FIS method between a black hole algorithm (BHA) and a fuzzy inference system (FIS) of the Sugeno model. The proposed BHA-FIS was used to improve the classification performance of a dataset of thalassemia patients. The results from the proposed method were compared with the FIS results through Tables (2-4) and Figure (4-6). The experimental results of the dataset in Tables (2-4) indicate that the proposed BH-FIS method has a higher classification performance than FIS through the MSE criteria.

Bibliography

1. Qasim OS., *et al*. "Feature selection based on chaotic binary black hole algorithm for data classification". *Chemometrics and Intelligent Laboratory Systems* 204 (2020): 104104.
2. Al-Thanoon NA., *et al*. "Tuning parameter estimation in SCAD-support vector machine using firefly algorithm with application in gene selection and cancer classification". *Computers in Biology and Medicine* 103 (2018): 262-268.
3. Atanassov KT. "Type-1 fuzzy sets and intuitionistic fuzzy sets". *Algorithms* 10.3 (2017): 106.
4. Mendel JM. "Type-2 fuzzy sets, in Uncertain rule-based fuzzy systems". (2017): 259-306.
5. Sepúlveda R., *et al*. "Experimental study of intelligent controllers under uncertainty using type-1 and type-2 fuzzy logic". *Information Sciences* 177.10 (2007): 2023-2048.
6. Nguyen HT., *et al*. "A first course in fuzzy logic". CRC press (2018).
7. Klir GJ., *et al*. "Fuzzy set theory: foundations and applications". Prentice-Hall, Inc (1997).
8. Castillo O. "Type-2 fuzzy logic in intelligent control application". 272 (2012).
9. Shill PC. "Fuzzy Logic Controllers: Optimization Issues on Design and Rule Base Reduction Algorithms". (2013).
10. Hatamlou A. "Black hole: A new heuristic optimization approach for data clustering". *Information Sciences* 222 (2013): 175-184.
11. Kumar S., *et al*. "Black Hole Algorithm and Its Applications". 575 (2015): 147-170.
12. Chen SM and YC. Chang. "Multi-variable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques". *Information Sciences* 180.24 (2010): 4772-4783.