Journal of The Gujarat Research Society

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

Grey Wolf Optimization for Data Mining

Dr Umesh Sharma

SOEIT, Sanskriti University, Mathura, Uttar Pradesh, India

Email Id-umeshsharma.mgmt@sanskriti.edu.in

ABSTRACT: In the last several decades, the development of innovative optimization techniques that were successfully utilized to handle such stochastic mining challenges sparked a lot of interest in data mining optimization. Data mining is the method of identifying data trends with the help of specialized systems. Many research has already demonstrated a range of data mining techniques, but none has indicated which optimization strategy delivers superior data mining results. Researchers offer the grey wolf optimization approach for data mining in this study. The flowchart and mathematical equation for the Grey Wolf Optimizer (GWO) algorithm with training Multi-layer Perceptron (MLP) neural network are presented in this study. This study also explains why GWO is the best optimization approach for data mining jobs, using factors like as run time, Mean Square Error (MSE), and classification rate to support its claim. Hybrid algorithms, which integrate two distinct optimization techniques, can help Data mining jobs in the future.

KEYWORDS: Algorithm, Data Mining, Grey Wolf Optimizer (GWO), Multi-layer Perceptron (MLP), Techniques.

1. INTRODUCTION

In previous years, the term "data mining" was considered a trendy issue. One of the most common job titles in the United States today is "data scientist." Finding necessary trends and patterns in data has been classed under numerous titles in recent years, including knowledge extraction, data mining, and pattern processing. Data mining is the process of using specialized algorithms to identify data patterns. Data mining may uncover hidden trends and patterns that aren't visible through simple query analysis, thus it can play an important role in knowledge discovery and decision-making[1].

Recommendation systems, Nave Bayes classifiers, and text mining are all examples of data mining. Data mining tries to build a model that is either descriptive or predictive, and it works with large amounts of data to find information and hidden patterns or trends[2]. This emphasizes the need of developing optimal model parameters for data mining issues. Traditional mathematical approaches cannot address such problems, and many daily situations involve several conflict metrics for performance, or objectives, that must be optimized simultaneously in order to achieve a true tradeoff [3], [4].

Nature-inspired algorithms are still a hot topic with a wide range of applications in business. Evolutionary algorithms, such as evolutionary feed-forward neural networks, evolutionary deep neural networks, and evolutionary multi-layer perceptron, were also employed in neural networks, and there are many nature-inspired algorithms that are used for optimization[5]. Evolutionary algorithms (EAs) are often used to solve problems with a single goal. However, excellent performance in one area does not always convert into the same degree of success in another[6]. This implies the need for an agreement. Mirjalili et al. proposed a novel meta-heuristic algorithm in their study named “Grey wolf Optimization”. It imitates the behaviour of grey wolves or hunters. The algorithm has been utilized in a variety of disciplines, including

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

those involving neural network training, power systems, time-series classification, clustering, and other topics.

Gray wolf optimization (GWO) is utilized for data mining in this study. It was proposed in 2014 by scholar Mirjalili, and it was based on the natural leadership hierarchy and hunting mechanism of grey wolves. The optimization process includes social hierarchy, prey encirclement, and prey assault. In comparison to other optimization approaches, grey wolf optimization is the best methodology for data mining, according to this article. In this experiment, the researcher generates an MLP dataset and then uses the grey wolf optimization method to determine different parameters like as runtime, classification rate, and MSE.

2. LITERATURE REVIEW

Amir Safari et al. discussed an analysis on Nature-Inspired Optimization Approaches to 2D Geometric Modelling[7]. In recent years, a wide range of population-based optimization approaches have been developed for use in various engineering applications, the majority of which are inspired by natural processes occurring in our environment. However, these algorithms' mathematical and statistical analysis is still missing. This study compares the performance of some of the most prominent nature-inspired optimization methods for complicated high-dimensional curve/surface fitting issues using a new foundation. The point cloud of a gas turbine compressor blade in-hand recorded by touch trigger probes is best fitted using B-spline curves as a case study. Five different population-based evolutionary and swarm optimization approaches are used to identify the optimal number/location of a set of Bezier/NURBS control points for all segments of the airfoil profiles. Before planning an experiment, parametric and nonparametric statistical assessments as well as mathematical research are provided to thoroughly examine and fairly compare the acquired data. The findings reveal a number of benefits and disadvantages of each optimization approach for parameterizing such complicated geometries from a variety of perspectives.

Athraa Jasim Mohammed et al. discussed a review on Nature Inspired Data Mining Algorithm for Document Clustering in Information Retrieval[8]. Document clustering is an essential approach in information retrieval that has been widely used (IR). Various clustering algorithms have been described, however most of them rely on the initial value of k clusters for their efficacy. Such an approach may not be appropriate since we may not have prior information of the document collection. Various swarm-based clustering approaches have been presented to solve this challenge to date, including this work, which investigates the application of the Firefly Algorithm (FA) in document clustering. We expand on the Gravitation Firefly Algorithm (GFA) work by providing a relocate mechanism that, if necessary, relocates allocated documents. The newly suggested clustering method, GFA R, is then put to the test on a 20Newsgroups benchmark dataset. The GFA R's experimental findings on external and relative quality metrics are compared to those achieved with the conventional GFA and Bisect K-means. It is discovered that by expanding GFA to become GFA R, improved clustering may be achieved.

J.M Johnson et al. discussed about Genetic algorithms and its various application related to designing of antenna [9]. Many writers have looked at the synthesis of antenna designs using iterative optimization approaches. To be effective, these approaches to pattern synthesis must be confined to relatively basic arrays or need a careful, intelligent selection of optimization

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

starting points based on the nature of the optimization techniques and functions to be optimised. Because traditional functional-optimization approaches rely on local, greedy optimization methods like random walk solution or gradient methods space searches, this is the case. This study explains how genetic algorithm (GA) optimization, a fundamentally diverse and relatively new methodology for functional optimization, overcomes the above-mentioned difficulties of previous methodologies and is applied to one-D and Two-D designing of antenna.

Nitin Uniyal et al. discussed an overview of Nature inspired algorithms and its applications[10]. Due to its inevitability in the creation of new algorithms in nearly every applied field of mathematics, optimization has become a hot issue. Regardless matter how wide optimization approaches are in study domains, there is always room for improvement. We give an overview of nature-inspired optimization with grounding in basics and categorization, as well as their dependability applications. An attempt has been made to demonstrate the differences between multi-objective optimization and single-objective optimization. Though there are a variety of approaches for achieving optimality in optimization issues, nature-inspired algorithms have shown to be quite effective and have received a lot of attention in recent study. The goal of this paper is to provide an interested researcher with the basis of a few nature-inspired optimization approaches and their dependability applications.

Research Questions:

- How grey wolf optimization is better than the other optimization techniques?
- Explain the flowchart of training of MLP with GWO algorithm?
- How Nature-Inspired Optimization Techniques helps in data mining process?

3. METHODOLOGY

3.1. Design:

3.1.1. Grey wolf optimizer (GWO):

This algorithm is inspired by the natural leadership organization and hunting mechanism of grey wolves. In addition, three crucial phases of hunting have been optimized: looking for prey, surrounding, and attacking prey. It has the same behaviour as grey wolves or hunters. Neural network training, clustering, time-series classification, and power systems are just a few of the domains where the approach has been applied. The flowchart for training a multi-layer perceptron (MLP) with GWO is shown in Figure 1. Based on the MLP theorem[11], no-one meta-heuristic can be stated to be best suited for solving all optimization problems; therefore various variants of the method, such as Multi-objective and Hybrid GWO, have been developed.

Journal of The Gujarat Research Society

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

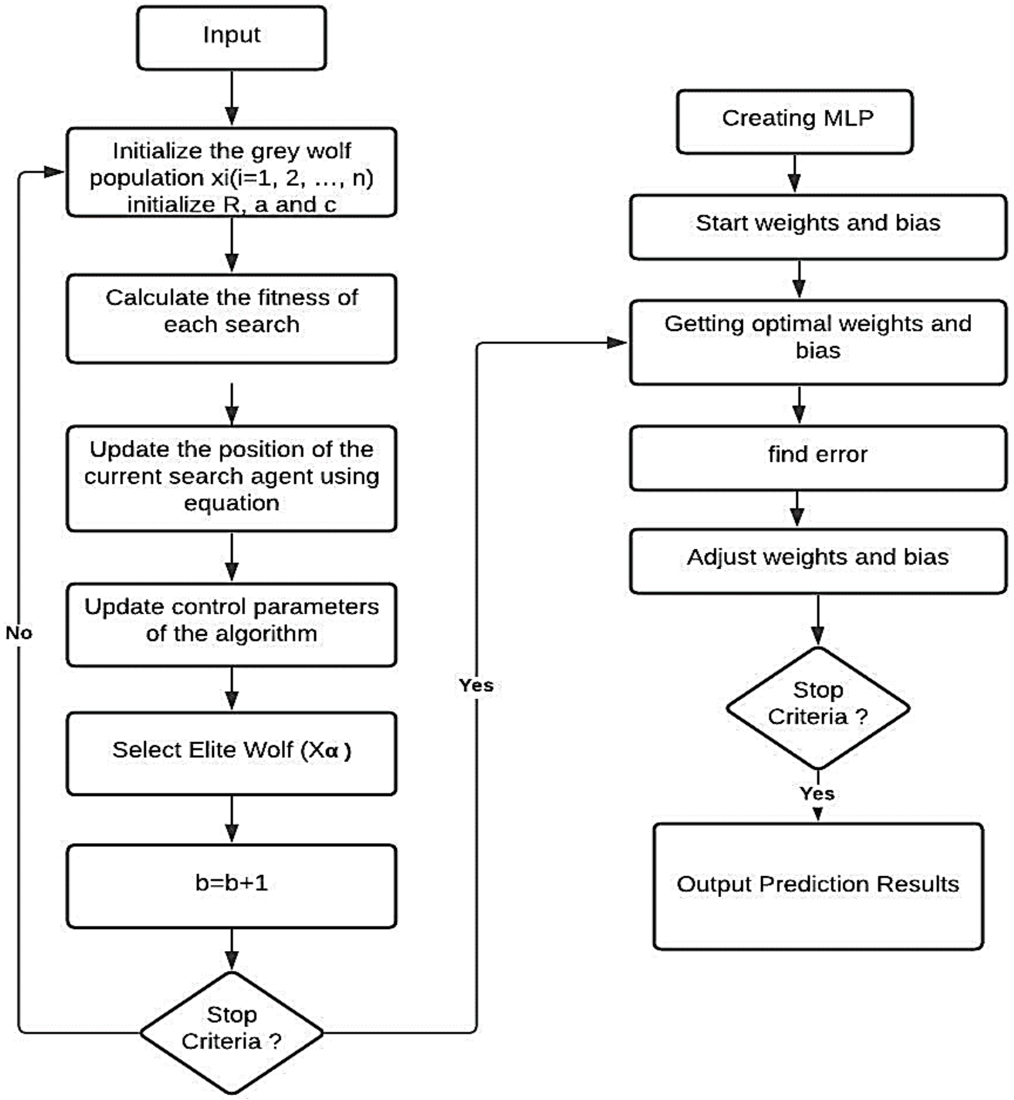

Figure 1: Illustrates the flowchart of training MLP with GWO algorithm.

After that, the grey wolf optimizer is set up, and each search's fitness is calculated. The current search agent's position is determined using equations based on the grey wolf optimizer after the computation is complete. The control settings of the algorithm are then changed. Increase the number of iterations. The stop criteria are then evaluated to determine whether they fulfil, and if they do, optimum weights and bias, as well as weights and bias errors for MLP, are produced. Finally, the errors are corrected to guarantee that the stop requirements are satisfied, and the classification rate and MSE outputs are generated.

3.1.1.1.Prey encircling: Equations involved in this are discussed as:

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

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

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

Here, current iteration is represented by t, C and A vectors are coefficient vector, the position of the prey is shown by X_p vector & X vector the location vector.

The vectors representation of C and A are considered as follow:

$$\vec{A} = 2a \cdot r_1 - a$$

$$\vec{C} = 2 \cdot r_2$$

Here, a is the component which decreased from two to zero over and r_1, r_2 are random-vector in [0, 1].

3.1.1.2.Hunting:

This method can be used to find and encircle prey. In most cases, the alpha is in command of the hunt. Hunting is something the beta and delta might undertake from time to time. We assume that the delta, beta, and alpha have a better grasp of possible prey sites in order to numerically imitate grey wolf hunting behaviour. As a consequence, we'll maintain the top three results we've identified so far and ask additional search agents to update their places in line with the best search agent's position. This is seen in the formulas that follow.

$$\vec{D}_\alpha = \left|\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}\right|, \vec{D}_\beta = \left|\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}\right|, \vec{D}_\omega = \left|\vec{C}_3 \cdot \vec{X}_\omega - \vec{X}\right|$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \left(\vec{D}_\alpha\right) \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \left(\vec{D}_\beta\right), \vec{X}_3 = \vec{X}_\omega - \vec{A}_3 \cdot \left(\vec{D}_\omega\right)$$

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

These equations, which are dependent on alpha, beta, and delta, are used by a search agent to modify its position in an n-dimensional search space. Furthermore, the ultimate position would be in a random point within a circle created by the delta, beta, and alpha values of the search space. To put it another way, delta, beta, and alpha make educated guesses about the location of the prey, while other wolves move about it at random.

4. RESULTS AND DISCUSSION

The MLP dataset was first initialized in this study, and then the grey wolf optimizer was used. Following that, a number of variables such as MSE and classification rate are utilized to assess whether or not this method is superior. In previous studies, authors/researchers have utilized a variety of methods for data mining, however just one approach is used in this study. Data mining is the process of identifying data trends via the use of specialized algorithms. The training of MLP using grey wolf optimization approaches for data mining is presented in this research article. The goal of this study is to use GWO to do data mining activities in order to improve MSE and classification rate. This technique proves to be an extremely efficient way

Journal of The Gujarat Research Society

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

for training MLPs. As can be seen above, it maintained a respectable classification rate and MSE for the most part.

It also had short runtimes in most cases. The convergence rate of the grey wolf algorithm is also rather excellent. It has a very high convergence rate. GWO is expected to achieve such high results because it has a high level of local-optimal avoidance and adaptive parameters that allow it to balance exploitation and exploration. In the great majority of cases, GWO converges before any other approach. This optimization approach is highly recommended for data mining projects. MSE calculates the sum of the squares of the errors, or the mean squared difference between the actual and estimated values. MSE is one of the factors used to assess whether or not the grey wolf optimizer is good. Table 1 displays the results of classification rate and MSE for various run times, while Figure 2 shows a graph of MSE vs. runtime.

Table 1: Illustrates the results of MSE and classification rate on different run time

S. No.	Run time (sec)	Classification rate	MSE
1	900	82.667%	0.0215
2	8482	98.6667%	0.0256
3	562	90.6667%	0.0246
4	2614	100%	0.001324
5	1814	99%	0.001125

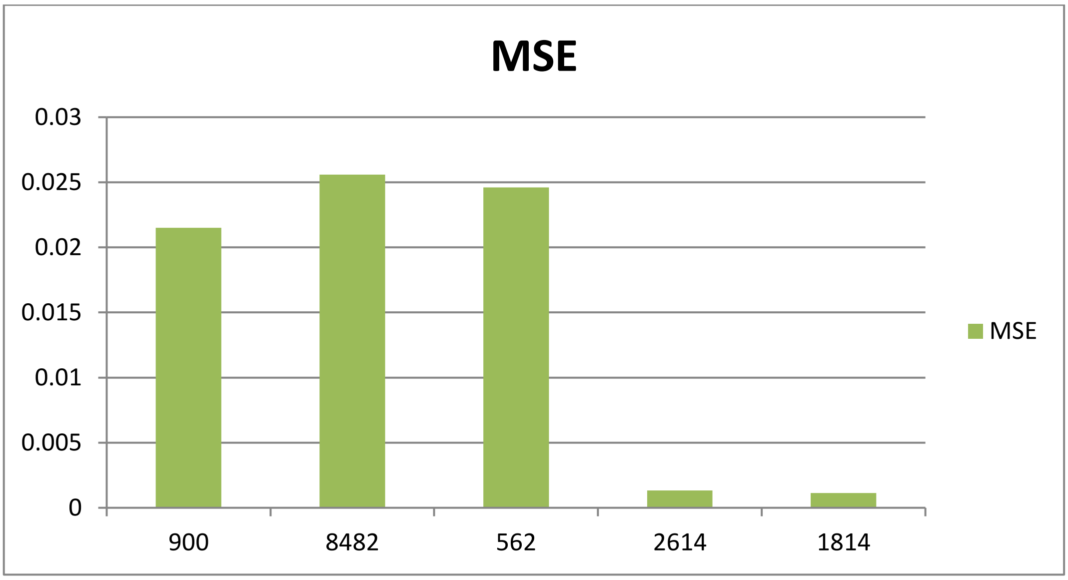

Figure 2: The above diagram shows the graph of MSE with run time

5. CONCLUSION

ISSN: 0374-8588
Volume 22 Issue 1, January 2020

Nature-inspired optimization algorithms are useful in a variety of fields, including engineering, image processing, and data collections. It is divided into two categories: Swarm Intelligence and Evolutionary Algorithms. Many studies have already shown a variety of data mining approaches, but none of them have stated which optimization strategy produces superior data mining outcomes. The researcher proposes a grey wolf optimization approach for data mining that uses ANN in this research article. The training MLP neural network is utilized in this, and MSE, classification rate, and runtime are among the parameters that demonstrate GWO is the best for data mining. In addition, simulation results demonstrate that grey wolf optimizer (GWO) based neural networks produce better results than other approaches. Hybrid algorithms, which integrate two distinct optimization techniques, can help Data mining jobs in the future.

REFERENCES

[1] M. J. Hamid Mughal, "Data mining: Web data mining techniques, tools and algorithms: An overview," *Int. J. Adv. Comput. Sci. Appl.*, 2018, doi: 10.14569/IJACSA.2018.090630.

[2] W. L. Chang, J. Kanesan, A. J. Kulkarni, and H. Ramiah, "Data clustering using seed disperser ant algorithm," *Turkish J. Electr. Eng. Comput. Sci.*, 2017, doi: 10.3906/elk-1512-231.

[3] X. S. Yang, *Nature-Inspired Optimization Algorithms*. 2014.

[4] "Nature-Inspired Optimization Algorithms," in *Nature-Inspired Optimization Algorithms*, 2014.

[5] R. E. Precup, P. Angelov, B. S. J. Costa, and M. Sayed-Mouchaweh, "An overview on fault diagnosis and nature-inspired optimal control of industrial process applications," *Comput. Ind.*, 2015, doi: 10.1016/j.compind.2015.03.001.

[6] S. Zhang and Y. Zhou, "Grey Wolf Optimizer Based on Powell Local Optimization Method for Clustering Analysis," *Discret. Dyn. Nat. Soc.*, 2015, doi: 10.1155/2015/481360.

[7] A. Safari, H. G. Lemu, S. Jafari, and M. Assadi, "A comparative analysis of nature-inspired optimization approaches to 2d geometric modelling for turbomachinery applications," *Math. Probl. Eng.*, vol. 2013, 2013, doi: 10.1155/2013/716237.

[8] A. J. Mohammed, Y. Yusof, and H. Husni, "Nature inspired data mining algorithm for document clustering in information retrieval," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2014, doi: 10.1007/978-3-319-12844-3_33.

[9] J. M. Johnson and Y. Rahmat-samii, "Genetic algorithm optimization and its application to antenna design," *IEEE Antennas Propag. Soc. AP-S Int. Symp.*, 1994, doi: 10.1109/aps.1994.407746.

[10] N. Uniyal, S. Pant, and A. Kumar, "An overview of few nature inspired optimization techniques and its reliability applications," *Int. J. Math. Eng. Manag. Sci.*, vol. 5, no. 4, pp. 732–743, 2018, doi: 10.33889/IJMEMS.2020.5.4.058.

[11] O. Häggström, "Intelligent design and the NFL theorems," *Biol. Philos.*, 2007, doi: 10.1007/s10539-006-9040-z.