Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 04, April -2018

# AN ENHANCEMENT OF WOA APPROACH FROM LARGE DATA FEATURE SELECTION

PRIYA RAGHUWANSHI [1], PROF.DEEPAK KUMAR GOUR[2]

[1]*Department of Computer Science Engineering, RGPV, Bhopal*

[2]*TIT, Bhopal, Madhya Pradesh India*

**ABSTRACT:** *In the given work we have explained about the Whale Optimization Algorithm (WOA) that provides the energy to homes, organisations or mechanical offices in circulation frameworks. The whale optimization algorithm motivated by humpback whales. Its calculation is benchmarked on 29 understood capacities. Also we have compared the proposed algorithm with the existing algorithm so to get the accurate results. Hence, we find accurate results by the help of the proposed algorithm.*

***Keywords:*** *WOA, Data mining, Clustering, WGC.*

## I INTRODUCTION

Data mining is methodology that takes information and converts that into useful contents which are helpful in future [1]. Such information objects, which are overwhelming or conflicting with the staying set of information, are called exceptions [2]. An anomaly is an informational index which is not quite the same as the rest of the information. Exception is additionally alluded to as distortion, degenerates and measurements writing. In many applications the information is made by at least one producing forms, which could either reflect action in the framework or perceptions gathered about substances. At the point when the creating methodology acts calmly, it brings about the making of exceptions. In this manner, an exception frequently contains valuable data about inconsistency attributes of the frameworks and substances, which affect the information age process. In Data Mining classification clustering [5]and regression are the three key approaches. Classification is a supervised learning approach in which students are grouped into identified classes [3]. Classification rules may be identified from a part of data known as training data and further it may be tested for rest of the data [4].

## II LITERATURE REVIEW

In this section different literature survey approach which works towards mining large data input such as KDD and other made available data from UCI and real time resources.

In this paper [11] Author presented an optimized algorithm which is WGC algorithm used for the better clustering performance over the dataset. WGC algorithm mainly focused on centroid value and its optimization. Proper value centroid estimation is performed through the WGC clustering technique. There are three dataset is taken on which their approach is performed; they are Banknote authentication dataset, wine dataset and Iris dataset which is taken from the UCI.

The algorithm is computed comparison with PSC, MPSC, GWO, EGWO, KEGWO and WOA. Proposed solution WGC find suitable with the parameter such as MSE and F-measure parameter. This algorithm involves data input process, population generation phase. Further a random agent search and fitness function to compute fitness value is given. Returning best fitness value and thus finding an appropriate cluster is performed by the proposed approach WGC. This paper proposes a novel nature-inspired meta-heuristic optimization algorithm, called Whale Optimization Algorithm (WOA), which mimics the social behaviour of humpback whales. The algorithm is inspired by the bubble-net hunting strategy. WOA is tested with 29 mathematical optimization problems and 6 structural design problems. Optimization results prove that the WOA algorithm is very competitive compared to the state-of-art meta-heuristic algorithms as well as conventional methods.

In this paper [12] author proposed an algorithm which is PSO, particle swarm optimization for the efficient clustering. [6] Particle swarm optimization algorithm works on the basis of complex computations. This algorithm generates number of clusters and fine solution with its internal architecture. This technique use entropy parameter and weak cluster computation. Weight based minimization approach is computed by the author which focused on dataset made available from UCI. They have compared their result with a detection rate, F-Measure and cluster generation from the algorithm. Further there are variants and improvement on weight and data handling complexity is worked with MCSS-PSO Meta heuristic approach, Fuzzy based PSO which is PSOFKM for efficient cluster optimization.

 

In this paper [13] an approach which is LOR line overlap rate based approach. Also hierarchical based overlap rate technique used for the data parsing and optimization. They have used Gaussian mixture model for data production. This algorithm help in creating one cluster and verify the match over the data overlapping. Data overlap matching and combining them into the single cluster is performed by LOR based approach [7]. Finding cluster using the given approach is efficient while large dataset with low density is available. A large data input with diverse and more noise level data need extra process to pre-filter, pre-process data. This technique is suitable where a data clarity input is given.

In this paper [14] author given the PSO approach and genetic algorithm for the data dynamic clustering. Genetic algorithm makes use of various intermediate techniques which includes the initial population generation from the input dataset [8]. Further on chromosome finding from it. Performing mutation over the finding chromosome and new population generation. Further a rule applying and fitness function applies for the cluster generation is given in this approach. Genetic algorithm help in dynamic feature selection and finding further cluster from that. It help in data pre-fetching, analysing and optimizing value of cluster usage accordingly.

In this paper [15] author worked towards the cluster generation and method is K-Mean, which is very popular among the clustering approaches. This technique finds many applications including outlier detection from the input dataset taken from UCI directory. These approaches first of all perform input data partitioning. First of all possible clusters are found and then minimization is drawn through the K-Mean approach [9]. Centroid value calculation is performed and all possible clusters finding from the data is gathered. Variance computation, minimum length computation and clusters are return by the returning process of algorithm. A variance computation and means of all computed data is important computation part of the approach.

In this paper [16] author presented an ANN (Artificial neural network) based approach for cluster detection. Author presented a fuzzy clustering-Artificial neural network (FC-ANN) to enhance the precision and accuracy rate. Author added and further divided the training data into different subset using fuzzy clustering. For each subset of training data, [10] ANN was applied to learn the system precisely. After learning of each subset, he used fuzzy aggregation module to learn again and combine the different ANN's results. A demonstration and work comparison with UCI dataset and compared with existing technique such as Naïve Bayes and other competitive algorithm. An ANN based approach help in rule definition and rule detection to optimize input data in more efficient manner.

## III RELATED WORK

This section intends to give the readers the necessary background to understand the key differences between the compared algorithms.

**F-Measure:** It can be defined as an information retrieval (IR) system has recall R and precision P on a test document collection and an information need.

**Mae:** It is the abbreviation of the mean absolute error which is a measure of difference between two continuous variables.

**Accuracy:** It can be defined as the sum of all the values that is divided by the given set of numbers.

## IV PROPOSED ALGORITHM

Disseminated generator (DG) assets are little scale electric power creating plants that can give energy to homes, organizations or mechanical offices in circulation frameworks. Power misfortune diminishments, voltage profile change and expanding unwavering quality are a few focal points of DG units. The above advantages can be accomplished by ideal arrangement of DGs. Whale streamlining calculation (WOA), a novel met heuristic calculation, is utilized to decide the ideal DG estimate. WOA is displayed in view of the remarkable chasing conduct of humpback whales.

- The Whale Optimization Algorithm motivated by humpback whales is proposed.
- The WOA calculation is benchmarked on 29 understood test capacities.
- The outcomes on the unimodal capacities demonstrate the unrivalled abuse of WOA.
- The investigation capacity of WOA is affirmed by the outcomes on multimodal capacities.
- The outcomes on auxiliary plan issues affirm the execution of WOA.

## V PROPOSED METHODOLOGY

The above algorithm explains about the whale optimization algorithm work flow which is explained in the form of the algorithm where the instances has taken as an input values also the check has been done at each step and then it get exit. First of all, communication node setup, data entity, node scenario is generated along with the mobility, data packet size, packet rate etc.

1. First of all we have taken input of given datasets attributes, for the further calculation.
2. Then the steps begin for each instance that has been calculated by the particular value.
3. Then the data values has been read by the function named ReadDataValues (), then, noise removal has done only if the value of instance is equal to zero. Then, the function RemoveupdateDS () will work or else it will move to the ExactProcess ().
4. The ExactProcess () will initials the population which is equal to the total whale population.
5. Apply FNN () has done in this step.
6. At the end the Compute Usage () and Compute Efficiency Param () so to get the usage and the efficiency parameters in a proper set of values.
7. Exit.

## VI RESULT ANALYSIS

**Table 1. Statically Analysis – Contact lenses Dataset**

| **Algorithm name/Parameter** | RF | RT | WOA +SA | WOA-FFANN |
|---|---|---|---|---|
| **F- Measure** | 90 | 91 | 93 | 94.76 |
| **MAE** | 78 | 81 | 85.6 | 91.90 |
| **ACCURACY** | 91 | 93.8 | 90.04 | 94.8 |

The above table explains about the algorithms and the calculations that are performed on the basis of parameter.

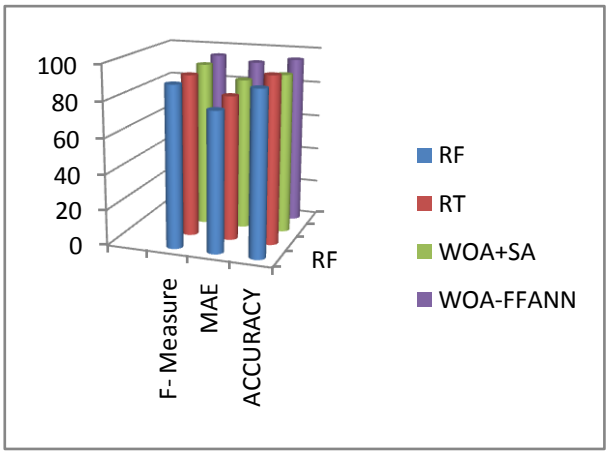


**Figure 1. Graphical Analysis – Contact lenses Dataset**

The above graph explains about the several algorithms F-measure, MAE, Accuracy on the basis of different calculations parameters.

## VII CONCLUSION

Data mining and extracting cluster from the available technique is an important aspect task. Mining proper and accuracy level data is essential from the exponential growing data from large resources. Data clustering give a separation and analysis of data in different phases. Data clustering help in detective multiple levels of data availability and their impact. Clustering can help in finding undetermined or un-detected data from the human techniques. Machine learning help in creating data cluster from the input resources and analysis. In this paper a survey of the available technique for data mining clustering is performed. Data usage approach over the large input data processing, finding centroid for cluster

 

detection is performed. Our further work is going to find an optimized solution for the mention limitation and overcome by our proposed enhance solution for data cluster finding approach and their application

## REFERENCES

[1] Sonali Agarwal, Data Mining in Education: Data Classification and Decision Tree Approach, International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, April 2012

[2] G. N. Pandey Data Classification and Decision Tree Approach, International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 2, No. 2, April 2012

[3] A. Merceron and K. Yacef, Educational Data Mining: a Case Study, In C. Looi; G. McCalla; B. Bredeweg; J. Breuker, editor, Proceedings of the 12th international Conference on Artificial Intelligence in Education AIED, pp. 467–474. Amsterdam, IOS Press, 2005

[4] I. H. Witten and E. Frank, Data mining: Practical Machine Learning Tools and Techniques, 2nd ed, Morgan-Kaufman Series of Data Management Systems San Francisco Elsevier, 2005.

[5] Neha Sharma, Routing and clustering optimization techniques in WSN: A Review, IEEE – 40222.

[6] Zhenfeng Hu, an Intelligent Test Paper Generation Method Based on Genetic Particle Swarm Optimization, 2016 13th Web Information Systems and Applications Conference.

[7] Jun Ahn, Kye Young Jeong, Fast cone-bearn-based LOR reconstruction for 3-D PET, 2012 IEEE Nuclear Science Symposiwn and Medical Imaging Conference Record (NSS/MIC).

[8] Omid E. David, H. Jaap van den Herik, Genetic Algorithms for Evolving Computer Chess Programs, IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION.

[9] Vaibhav Kumar, K-mean Clustering based Cooperative Spectrum Sensing in Generalized κ-μ Fading Channels.

[10] Ms. Sonali. B. Maind, Artificial Neural Network, and International Journal on Recent and Innovation Trends in Computing and Communication Volume: 2 Issue: 1.

[11] Amolkumar NarayanJadhav, WGC: Hybridization of exponential grey wolf optimizer with whale optimization for data clustering, Alexandria Engineering Journal (2017).

[12] Hemlata S. Urade, Study and Analysis of Particle Swarm Optimization, 2nd National Conference on Information and Communication Technology (NCICT) 2011.

[13] Ch. Raju, Analysis of Parsing Techniques & Survey on Compiler Applications, IJCSMC, Vol. 2, Issue. 10, October 2013, pg.115 – 125.

[14] Thomas R. Schneider, A genetic algorithm, ISSN 0907-4449 Accepted 12 November 2001.

[15] Jyoti Yadav, A Review of K-mean Algorithm, International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Issue 7- July 2013.

[16] Ms. Sonali. B. Maind, Artificial Neural Network, ISSN: 2321-8169 Volume: 2 Issue: 1 96 – 100.

 