Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 8 July, 2021: 6157 – 6168

Research Article

A Chaotic Elephant Herding Optimization-Established Classification (EHOC) for Large Dataset to expand the appearance of Heterogeneous Distributed Situation

Rajeev Pandey

Ph.D. Scholar, Computer Science & Engineering Department, UIT, RGPV, Bhopal (M.P.), India, rajeevpandey@rgtu.net

Sanjay Silakari

Professor, Computer Science & Engineering Department, UIT, RGPV, Bhopal (M.P.), India, ssilakari@yahoo.com

Abstract

Data clustering (or classification) is broadly exploited for an enormous range of applications in numerous regions such as education, remedial, invention, and organizations in heterogeneous distributed environments. The large heterogeneous data are processed and examined by using various classification methods to enhance the quality of important information transferring over distributed environment. Here, a Chaotic Elephant Herding Optimization based Classification (CEHOC) is implemented to classify the large amount of heterogeneous datasets to improve the performance of distributed computing. The key motive of CEHOC is to obtain a well arranged distribution of data elements over heterogeneous resources for distributed environment by utilizing some circumstances of chaos theory for population selection. The chaos function is introduced for improving the exploration and exploitation power of search agents in optimization algorithm and for performing the selection of centroids and members of classes (or clusters) optimally and precisely. The MATLAB 2021a tool is used to implement the CEHOC algorithm for four large datasets and the outcomes describe the superior effectiveness of CEHOC algorithm according to parameters like purity index, F-measure, intra-cluster distance, time complexity and standard deviation against previous algorithms such as K-Means, PSO, ACO and EHO.

Keywords: Chaos Function, Centroids, Classification, Clustering, Elephant Herding Optimization, F-Measure, Purity Index.

1. Introduction

The big data [1, 2] is processed and examined in numerous applications by exploiting huge features such as memory, execution, and management of information. The massive quantity of data is examined under pre-processing ahead of scrutiny to diminish the data dismissal with improving the data precision and competence. A powerful distributed architecture was well comprised the various data processing techniques at server and client end. The architecture generated high speed and minimum cost data analysis for serial as well as parallel transmission [3, 4]. The big data is examined by concerning several meta-heuristic optimization methods such as Ant Colony Optimization (ACO), Fruit Fly Optimization (FFO), Ant Lion Optimization (ALO), Grey Wolf Optimization Algorithm (GWOA) [5] and Particle

Swarm Optimization (PSO) to analyze the huge amount of data optimally and precisely in distributed computing environment [6, 7, 8]. The analogical reasoning is also performed in big data analysis for deep learning utilizing the neural network and feed-forward network. The case based reasoning is developed for query processing to reduce the query accessing cost and to enhance the accessing time of queries. The medical data are analyzed to visualize the diseases with qualitative and quantitative explanation for assuring the best treatment over diseases [9, 10, 11]. The convex optimization is the one of the best optimization methods for big data analysis in parallel and distributed computing for heterogeneous sources. The signal processing based convexity is used the least square method for sampling and computation of terabytes of data. The proximal mapping, randomization, parallel and distributed computing are key concerns in data optimal analysis [12, 13].

The distributed clustering was implemented over heterogeneous data by utilizing the deep learning based auto-encoder. The auto-encoder was automatically selected the number of cluster heads and after that; members of clusters were allocated by using deep learning algorithm. The results were compared with well known approach of clustering like K-Means [14, 15]. The convolution neural network is combined with machine learning techniques to improve the clarity and precision for selection of disease type as well as their treatment. The pipeline with auto-encoder is used to select the superior classification algorithm for clinical management and concept based learning about disease [16]. The decision making over diseases is performed by using artificial neural network and support vector machine with training and testing health datasets. The different types of cancers are analyzed with the help of artificial intelligence for proper and accurate treatment of disease [17].

The machine learning is also operated with tensor processing units and central processing unit to manage and train the datasets optimally. The training of data is processed by deep learning used in various platforms like Google and real time applications [18, 19, 20]. The classification is similar process as clustering, in which the several datasets are divided into groups for multiple uses. The decision support system of clinical data is used for classification of medical data to predict the disease [21, 22]. The classification of medical data is also performed by using convolution neural network with the help of a novel loss function. The multi pool procedure is applied to capture the location data of initial characteristics of clinical data, which helps to evaluate the seriousness of disease and treatment [23].

In above literature, numerous classification techniques are implemented on multiple large datasets to increase the data removal and processing. More than a few optimization methods such as ACO, ALO, PSO and FFO are also explained for classification. Yet, entire problems are not evaluated by using one method and convergence speed is a key concern of optimization methods. Here a Chaotic Elephant Herding Optimization based Classification (CEHOC) is implemented to classify the large amount of heterogeneous datasets, which is advanced form of Elephant Herding Optimization (EHO) algorithm. The key purpose of CEHOC is to acquire a well orderly distribution of data instances over heterogeneous resources for distributed environment by using chaos theory for population selection.

2. The Elephant Herding Optimization (EHO) Algorithm

An EHO is a nature motivated algorithm imitative from behaviour of elephant herding to extricate universal optimization problems. The EHO describes three basic concepts: (i) several predestined elephants are combined to form clans and a small number of clans are combined to obtain elephant`s population; (ii) various predestined male elephants force to depart from their clans and endure separate out of the way from most important elephant swarm at each formation; (iii) the elephants equally endure in clan beside the assistance of a matriarch.

2.1. Modifying Process of Clan

The subsequent elephant's position in clan A_c is impacted by matriarch A_c. The e^{th} elephant's position is altered in clan A_c by utilizing eq. (1).

$$X_{new,A_c,e} = X_{A_c,e} + \mu \times \left(X_{best,A_c} - X_{A_c,e}\right) \times random \tag{1}$$

Here,

$X_{new,A_c,e}$ & $X_{A_c,e}$ = e^{th} elephant next and present position in clan A_c respectively.

μ = An extent term obtaining the impact of matriarch A_c on $X_{Ac,e}$. $\{\mu \in [0,1]\}$

X_{best,A_c} = The elephant in clan A_c with maximum fitness value known as matriarch.

$random$= Arbitrary number. $\{random \in [0,1]\}$

The position of elephant with the best fitness value in each clan is altered by utilizing eq. (2), where eq. (1) is not utilized for alteration, i.e., $X_{A_c,e} = X_{best,A_c}$.

$$X_{new,A_c,e} = \gamma \times X_{center,A_c} \tag{2}$$

Where,

γ = A variable obtaining the rein of X_{center,A_c} on $X_{new,A_c,e}$. $\{\gamma \in [0,1]\}$

X_{center,A_c}= Clan A_c mid-point (centre).

The m^{th} dimension $\{1 \le m \le M, \quad M = Complete\ Dimensions\}$ is introduced in eq. (2) to generate the e^{th} elephant position (eq. (3)).

$$X_{center,A_c,m} = \frac{1}{N_{A_c}} \times \sum_{e=1}^{N_{A_c}} X_{A_c,e,m} \tag{3}$$

Where,

N_{A_c}= Total A_c clan elephants. (N_{cln} = Total clans in population of elephants)

$X_{center,A_c,m}$ = Elephant position mid-point (centre) in clan A_c (m^{th} dimension).

$X_{A_c,e,m}$= e^{th} elephant position in clan A_c (m^{th} dimension).

2.2. Separating Process

Afore, increasing the exploration potency of EHO algorithm, accede that the separating process will recognize the elephants having worst fitness value at every formation as formulated in eq. (4).

$$X_{wst,A_c} = X_{mnm} + (X_{mxm} - X_{mnm} + 1) \times random \tag{4}$$

Where,

X_{mxm} & X_{mnm}= Maximum and minimum bound of elephant position.

$random$ = Speculative and consistent distribution. $\{random \in [0,1]\}$

X_{wst,A_c}= Elephant in the clan A_c with worst fitness value.

3. A Chaotic Elephant Herding Optimization based Classification (CEHOC) Algorithm

An improved EHO algorithm is implemented utilizing the chaotic function. The EHO algorithm is well appropriate for searching and utilization in enormous exploring area and obtains the proficient outcomes for numeric optimization. Yet, the numeric optimization is realistically disparate from data classification (or clustering) strategy. Now, various inherent features are calibrated and a chaotic EHO is applied using chaotic function for data classification.

3.1. Population Selection of EHO Algorithm through Chaos Function

The chaos function is tremendously interconnected to preliminary situation and fruitfully exploited for capricious number procreation utilizing logistic map. The chaotic function is formulized by eq. (4).

$$\beta_{k+1} = \delta \times \beta_k \times (1 - \beta_k) \tag{4}$$

Here,

δ = Constant coefficient$\{\delta \in [1, 4]\}$.

β = Random variables ($\beta \in [0,1]$, $\quad k = 0,1,2, \dots \dots \dots$)

The EHO Population (P^E) is assigned the primary values by evaluating chaotic function (eq. (4)) for increasing the proficiency of EHO with appropriate exploitation of enormous solution region.

3.2. The Entire CEHOC Algorithm for Data Classification

The population selection through chaos function is performed in Chaotic Elephant Herding Optimization based Classification (CEHOC) algorithm to classify the large datasets in optimal manner. Primarily, the initial value of EHO Population (P^E) is calculated by chaotic function (eq. (4)) and every individual is symbolized as a matrix with length $I_x = I \times C$, (I = Total dataset instances & C = Total cluster centroids & I_x = Individual dimension). The C centroids places are set into matrix, where 1st centroid represents through 1st I components, 2nd centroid represents through 2nd I components, and so on. The first attribute matrix values are obtained capricious and regularly between least and highest attribute's value in the active dataset contained by the Utmost number of repetitions (Ur). Subsequently, the CEHOC is used to obtain fitness values of total individuals by utilizing eq. (1) to eq. (4).

Algorithm 1: The Entire CEHOC Algorithm	
Algorithm	**Number of Executions**
START	
Assign creation counter j =1 and the utmost repetitions (Ur)	1
Set the initial value of population P^E of EHO by utilizing chaotic function (eq. (4))	1
Calculate fitness for each elephant	P^E

WHILE (j < Ur) **DO**	Ur+1
Perform Sorting to total elephants based on their fitness	Ur
FOR A_c = 1 to N_{cln} **DO**	Ur*(N_{cln}+1)
FOR e = 1 to N_{A_c} **DO**	Ur*N_{cln}*(N_{Ac}+1)
Modify $X_{A_c,e}$ and obtain $X_{new,A_c,e}$ (eq. (1))	Ur*N_{cln}*N_{Ac}
IF $X_{A_c,e} = X_{best,A_c}$ **THEN**	Ur*N_{cln}*N_{Ac}
Modify $X_{A_c,e}$ and obtain $X_{new,A_c,e}$ (eq. (2))	Ur*N_{cln}*N_{Ac}
END IF	
END FOR	
END FOR	
FOR A_c = 1 to N_{cln} **DO**	Ur*(N_{cln}+1)
Change the A_c clan elephant with worst fitness (eq. (4))	Ur*N_{cln}
END FOR	
Calculate population by recently modified locations	Ur
j = j+1;	Ur
END WHILE	
Return optimal fitness elephant	1
STOP	

4. Results and Analysis

Here, the complete six large datasets and performance parameters for CEHOC algorithm are described in brief. The whole algorithms are developed by utilizing MATLAB 2021a environment with windows 8 operating system and examined over six large datasets (Table 1). The CEHOC algorithm is obtained the outcomes for 600 repetitions according to parameters like purity index, F-measure, intra-cluster distance, time complexity and standard deviation against previous algorithms such as K-Means, PSO, ACO and EHO over 30 different executions.

4.1. Large Datasets

The CEHOC algorithm is processed on four large independent datasets receiving form UCI repository. The datasets are dry bean, waveform, online_shoppers_intention and shill bidding (Table 1).

Table1: Datasets

Sr. No.	Dataset	No. of Instances	No. of Attributes	No. of Classes/Clusters

1	Dry Bean	13611	17	7
2	Waveform	5000	40	3
3	Online_shoppers_intention	12330	18	2
4	Shill Bidding	6321	13	2

4.2. Performance Parameters

The CEHOC algorithm is examined according to parameters like purity index, F-measure, intra-cluster distance, time complexity and standard deviation.

4.2.1. Intra-cluster Distance

First of all, the distances among data instances are obtained inside a cluster. Later then, the intra-cluster distance is generated by calculating the mean of distances. The best classification is established by means of lowest intra-cluster distance. The distance average is calculated by the distances of centroid from complete data instances in a cluster (or class) and this strategy is applied on each class. Finally, the intra-cluster distance average value is generated by combining all mean distances of all classes.

Table 2: Intra-cluster distance based Average Ranking

Dataset	K-Means	PSO	ACO	EHO	CEHOC
Dry Bean	2547.34 (5)	2451.72 (3)	2513.67 (4)	2345.12 (2)	2234.35 (1)
Waveform	3.8734 (4)	4.0367 (5)	3.8571 (3)	3.4128 (2)	3.1025 (1)
Online_shoppers_intention	1752.69 (5)	1617.25 (4)	1547.69 (3)	1378.67 (2)	1287.25 (1)
Shill Bidding	84.3645 (5)	81.3624 (4)	80.7581 (3)	79.3614 (2)	77.2654 (1)
Average Ranking	4.75	4	3.25	2	1

Table 2 explains that the CEHOC calculates lowest intra-cluster distance for complete four large datasets. The CEHOC provides 10% improved outcomes than EHO, 24% improved outcomes than ACO, 21% improved outcomes than PSO and 31% improved outcomes than K-Means according to intra-cluster distance for complete four datasets. The average ranking is obtained for entire algorithms according to least to highest intra-cluster distance (from 1 to 4.75).

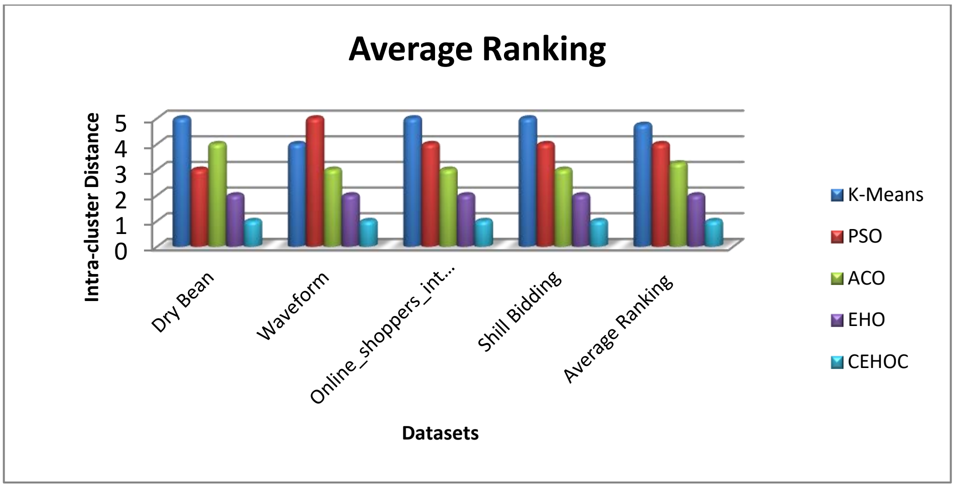

Figure 1: Intra-cluster distance based Average Ranking

Figure 1 shows the superior rank of ECHOC having lowest intra-cluster distance in opposition to K-Means, PSO, ACO and EHO algorithms for complete datasets.

4.2.2. Standard Deviation

The rigid data classification in the section of mean value is obtained through a geometric feature notated as Standard Deviation (StDe). The best classification is established by means of lowest standard deviation (eq. (5)).

$$StDe = \sqrt{\frac{\sum(U - \bar{U})}{|I|}} \tag{5}$$

Here,

$|I|$= Dataset size

V = Dataset points

$\bar{V}$= Mean dataset point value

Table 3: Standard Deviation for four Datasets

Dataset	K-Means	PSO	ACO	EHO	CEHOC
Dry Bean	18.5784	16.5847	14.3625	8.6547	7.3684
Waveform	0.3762	0.3964	0.3415	0.2547	0.1934
Online_shoppers_intention	37.2684	25.4287	26.8754	24.3256	21.8759
Shill Bidding	0.4758	0.4137	0.4367	0.3839	0.3571

Table 3 explains that the CEHOC calculated lowest standard deviation for complete four large datasets. The CEHOC provides 15% advanced results than EHO, 21% advanced results than ACO, 33% advanced results than PSO and 46% advanced results than K-Means according to standard deviation for complete four datasets.

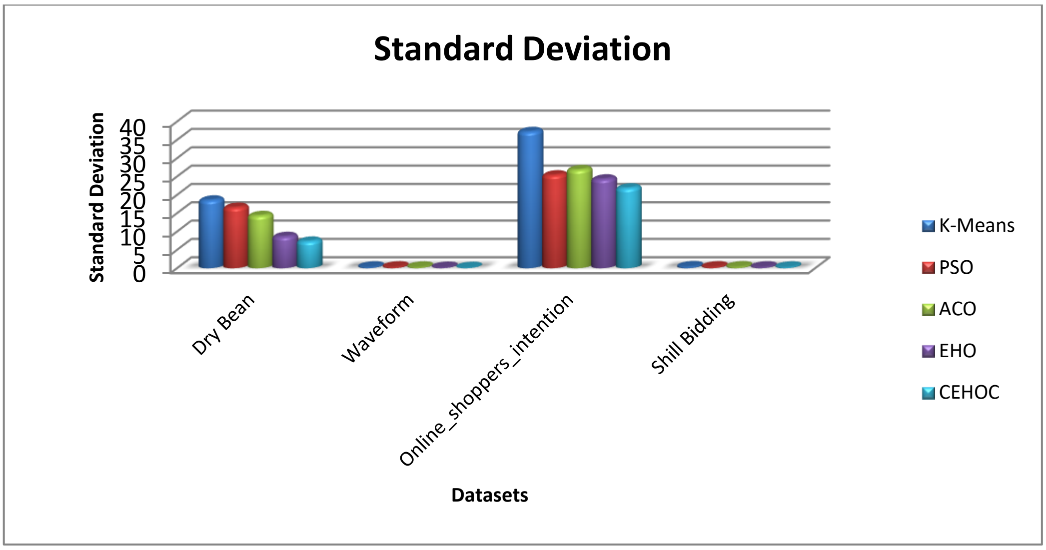

Figure 2: Standard Deviation for four Datasets

Figure 2 represents the lowest standard deviation of CEHOC in opposition to K-Means, PSO, ACO and EHO for complete datasets.

4.2.3. Purity Index

The appropriateness of classification (or clustering) technique is notated as Purity (Prt), where precise classification is generated for data instances. For this reason, complete instances of a distinct class can be correctly assigned to a distinct cluster. Eq. (6) and eq. (7) are calculated for generating Purity Index (Pu_In) by utilizing purity. The high purity is developed by means of utmost Pu_In value nearby to 1.

$$Prt(L_y) = \frac{maximum(|L_{yd}|)}{|L_y|} \tag{6}$$

$$Pu\_In = \sum_{y=1}^{C} \frac{(|L_y| Prt(L_y))}{|I|} \tag{7}$$

Here,

$Prt(L_y)$ = y^th cluster purity.

$|L_y|$= y^th cluster size.

$|L_{yd}|$= Total data instances of d^th class assigned to y^th cluster.

Table 4: Purity Index (Pu_In) for four Datasets

Dataset	K-Means	PSO	ACO	EHO	CEHOC
Dry Bean	0.71	0.80	0.79	0.83	0.85
Waveform	0.76	0.83	0.85	0.89	0.92
Online_shoppers_intention	0.73	0.77	0.79	0.81	0.83
Shill Bidding	0.78	0.83	0.82	0.86	0.90

Table 4 describes that the CEHOC evaluated utmost purity index for complete four datasets. The CBO-IE provides 4% improved results than EHO, 9% improved results than ACO, 14% improved results than PSO and 18% improved results than K-Means according to purity index for complete four datasets.

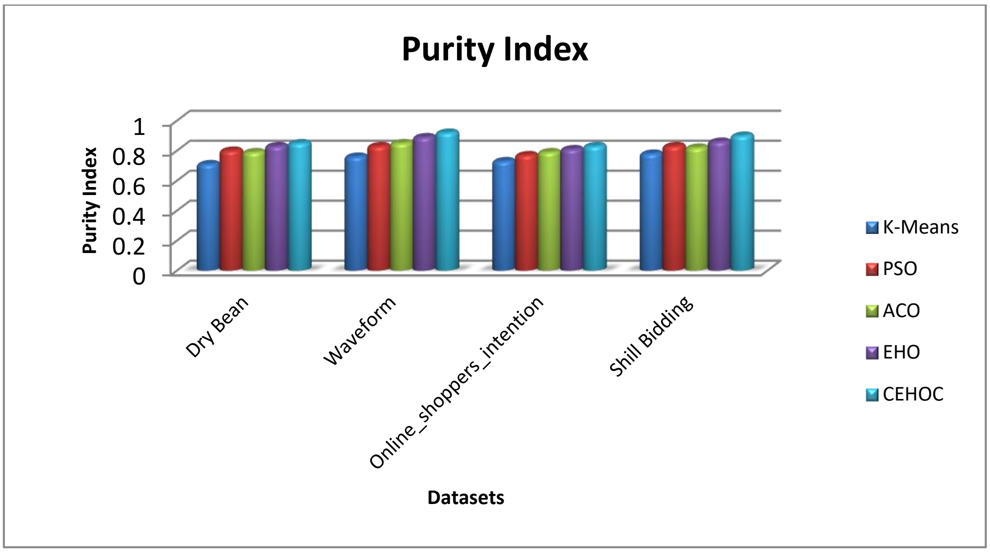

Figure 3: Purity index for four Datasets

Figure 3 illustrates the utmost purity index of CEHOC opposition to K-Means, PSO, ACO and EHO for complete datasets.

4.2.4. F-Measure

First of all, Precision (Pn) and Recall (Rl) are obtaining to recuperate the data (eq. (8) and eq. (9)). Later then, Pn and Rl are merged to evaluate F-Measure (F_M) (eq. (10) and eq. (11)).

$$Pn(d,y) = \frac{|L_{yd}|}{|L_y|} \qquad (8)$$

$$Rl(d,y) = \frac{|L_{yd}|}{|L_d|} \qquad (9)$$

$$F\_M(d,y) = \frac{2 \times Pn(d,y) \times Rl(d,y)}{Pn(d,y) + Rl(d,y)} \qquad (10)$$

$$F\_M = \sum_{d=1}^{C} \frac{|L_d|}{|I|} maximum\{F_{M(d,y)}\} \qquad (11)$$

Here,

$|L_d|$= dth class size.

Table 5: F-Measure for four Datasets

Dataset	K-Means	PSO	ACO	EHO	CEHOC
Dry Bean	0.68	0.77	0.76	0.80	0.82
Waveform	0.73	0.80	0.82	0.86	0.89

Online_shoppers_intention	0.70	0.74	0.76	0.78	0.80
Shill Bidding	0.75	0.80	0.79	0.83	0.87

Table 5 illustrates that the CEHOC calculates utmost F-Measure for complete four datasets. The CEHOC provides 6% improved outputs than EHO, 10% improved outputs than ACO, 16% improved outputs than PSO and 20% improved outputs than K-Means according to F-Measure for complete four datasets.

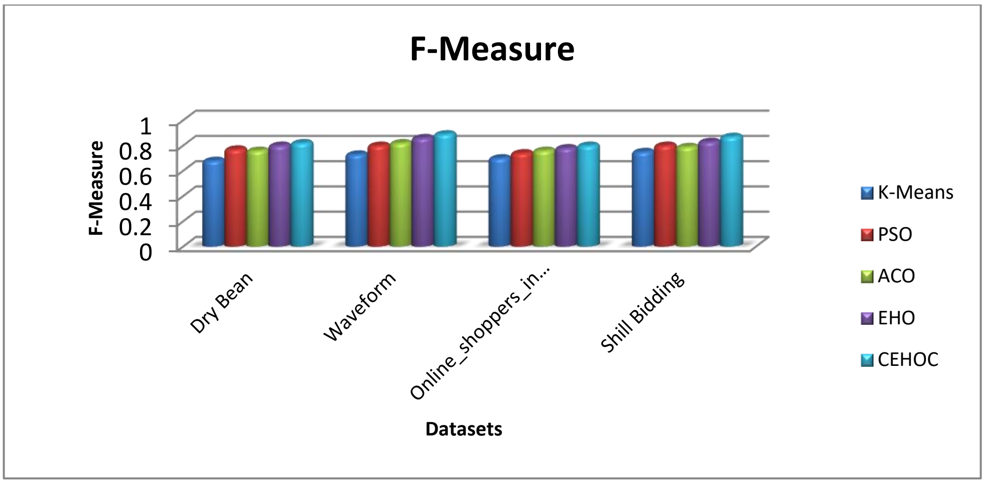

Figure 4: F-Measure for four Datasets

Figure 4 illustrates the utmost F-Measure of CEHOC opposition to K-Means, PSO, ACO and EHO for complete datasets.

The Chaos function is introduced for initial population generation of EHO to enhance the exploration and exploitation power of elephants in clan, which utilizes for cluster member allocation strategy more specifically to provide superior data instances distribution in dataset. Hence, superior quality of cluster centroids and members are generated optimally by using CEHOC algorithm. Therefore, CEHOC provides improved outputs than K-Means, PSO, ACO and EHO algorithms.

4.3. Time Complexity of CEHOC Algorithm

The time complexity is calculated by executing the classification process and count the number of executions. The step cost is set to be 1 unit for every execution. The Entire Execution Cost (EEC) is calculated with the help of algorithm 1 through eq. (12) and eq. (13).

$$EEC = 1 + 1 + P^E + Ur + 1 + Ur + Ur * (N_{cln} + 1) + Ur * N_{cln} * \left(N_{A_c} + 1\right) + Ur * N_{cln} * N_{A_c} + Ur * N_{cln} * N_{A_c} + Ur * N_{cln} * N_{A_c} + Ur * (N_{clm} + 1) + Ur * N_{cln} + Ur + Ur + 1 \tag{12}$$

$$EEC = 4 \times Ur \times N_{cln} \times N_{A_c} + 4 \times Ur \times N_{cln} + 6 \times Ur + P^E + 4 \tag{13}$$

Infer to be entire parameters are equivalent in eq. 13 in worst case; hence eq. 14 is obtained as follows:

$$EEC = 4n^3 + 4n^2 + 7n + 4 \tag{14}$$

The time complexities of classification algorithms are $O(n^3)$ for CEHOC, $O(n^3)$ for EHO, $O(n^3)$ for PSO, $O(n^3)$ for ACO and $O(n^2)$ for K-Means in worst case. For this reason, all algorithms are operated in polynomial time.

5. Conclusion

The several application regions like education, invention, remedial and organization are widely utilized the data clustering (or classification) in heterogeneous distributed environments. A lot of classification strategies are used for processing and examining of large heterogeneous data to increase the transmission quality of information over distributed environment. In this paper, a Chaotic Elephant Herding Optimization based Classification (CEHOC) is implemented to provide a classification of large heterogeneous datasets to get better efficiency of distributed computing. The appropriate data element distribution is a main motive of the CEHOC algorithm over heterogeneous resources of distributed environment by using chaos theory for selecting the population. The exploitation and exploration power of searching of optimization is enhanced by using chaos function, which is further utilized for optimal and precise cluster (or class) centroids and members selection. The CEHOC algorithm is implemented in MATLAB 2021a tool for four large datasets and the outputs explains the better-quality effectiveness of CEHOC algorithm according to parameters like purity index, F-measure, intra-cluster distance, time complexity and standard deviation against previous algorithms such as K-Means, PSO, ACO and EHO.

References

1. Rajeshkumar J and Kousalya K., “Diabetes data classification using whale optimization algorithm and back propagation neural network”, Int. Res. J. Pharm., Vol. 8(11), pp. 219-222, 2017. http://dx.doi.org/10.7897/2230-8407.0811242
2. Zhi-Hua Zhou, Nitesh V. Chawla, Yaochu Jin, and Graham J. Williams, “Big Data Opportunities and Challenges: Discussions from Data Analytics Perspectives”, IEEE Computational Intelligence Magazine, pp-1-20, 2015.
3. S. Kiruthika, EVRM. Kalaimani, R. Sudha and K. Suganthi, “ACO Feature selection and Novel Black Widow meta-heuristic Learning rate optimized CNN for Early diagnosis of Parkinson`s disease”, Turkish Journal of Computer and Mathematics Education, Vol. 12 (7), pp. 809-817, 2021.
4. Mohammed S. Al-kahtani, and Lutful Karim, “An Efficient Distributed Algorithm for Big Data Processing”, Arab J Sci Eng, Springer, Vol. 42, pp-3149–3157, 2017. (DOI 10.1007/s13369-016-2405-y)
5. Ankur Khare, Rajendra Gupta and Piyush Kumar Shukla, “A Grey Wolf Optimization Algorithm (GWOA) for Node Capture Attack to Enhance the Security of Wireless Sensor Network”, International Journal of Scientific & Technology Research, Vol. 9 (3), pp. 206-209, 2020.
6. Daniel Sonntag, and Hans-Jürgen Profitlich, “An architecture of open-source tools to combine textual information extraction, faceted search and information visualisation”, Artificial Intelligence In Medicine, Elsevier, Vol. 93, pp-13-28, 2019. (https://doi.org/10.1016/j.artmed.2018.08.003)
7. Rohyoung Myung, Heonchang Yu, and Daewon Lee, “Optimizing Parallelism of Big Data Analytics at Distributed Computing System”, International Journal on Advanced Science Engineering Technology, Vol. 7(5), pp-1716-1721, 2017.
8. F. Leena Vinmalar and A. Kumar Kombaiya, “An Improved Dragonfly Optimization Algorithm based Feature Selection in High Dimensional Gene Expression Analysis for Lung

Cancer Recognition", International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol. 9 (8), pp. 896-908, 2020.

9. Jean-Baptiste Lamy, Boomadevi Sekar, Gilles Guezenneca, Jacques Bouauda, and Brigitte Séroussi, "Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach", Artificial Intelligence in Medicine, Elsevier, Vol. 94, pp-42-53, 2019. (https://doi.org/10.1016/j.artmed.2019.01.001)
10. Maisa Daouda, and Michael Mayo, "A survey of neural network-based cancer prediction models from microarray data", Artificial Intelligence in Medicine, Elsevier, Vol. 97, pp-204-214, 2019. (https://doi.org/10.1016/j.artmed.2019.01.006)
11. Xueting Cui, Ying Li, Jiahao Fan, Tan Wang and Yuefeng Zheng, "A Hybrid Improved Dragonfly Algorithm for Feature Selection", IEEE Access, Vol. 8, pp. 155619-155629, 2020.
12. Farrukh Nadeem , Daniyal Alghazzawi, Abdulfattah Mashat, Khalid Faqeeh, and Abdullah Almalaise, "Using Machine Learning Ensemble Methods to Predict Execution Time of e-Science Workflows in Heterogeneous Distributed Systems", IEEE Access, Vol. 7, pp-25138-25149, 2019. (Digital Object Identifier 10.1109/ACCESS.2019.2899985)
13. Hamouda Chantar, Mohammad Tubishat, Mansour Essgaer and Seyedali Mirjalili, "Hybrid Binary Dragonfly Algorithm with Simulated Annealing for Feature Selection", SN Computer Science, Vol. 2(295), pp. 1-11, 2021.
14. Chin-Yi Chen and Jih-Jeng Huang, "Double Deep Autoencoder for Heterogeneous Distributed Clustering", Information, MDPI, Vol. 10 (144), pp-1-15, 2019. (doi:10.3390/info10040144)
15. Victor M. Herrera, Taghi M. Khoshgoftaar, Flavio Villanustre and Borko Furht, "Random forest implementation and optimization for Big Data analytics on LexisNexis's high performance computing cluster platform", Journal of Big Data, pp-1-36, 2019.
16. Yihan Denga, Andre Sanderb, Lukas Faulstichb, and Kerstin Deneckea, "Towards automatic encoding of medical procedures using convolutional neural networks and autoencoders", Artificial Intelligence in Medicine, Vol. 93, pp-29-42, 2019. (https://doi.org/10.1016/j.artmed.2018.10.001)
17. Nida Shahid, Tim Rappon, and Whitney Berta, "Applications of artificial neural networks in health care organizational decision-making: A scoping review", PLOS ONE, pp-1-22, 2019. (https://doi.org/10.1371/journal.pone.0212356)
18. Parvaneh Shabanzadeh and Rubiyah Yusof, "An Efficient Optimization Method for Solving Unsupervised Data Classification Problems", Computational and Mathematical Methods in Medicine, Hindawi, Vol. 2015, pp. 1-9, 2015.
19. Samir Mitta, "Cognitive Computing Architectures for Machine (Deep) Learning at Scale", Proceedings, MDPI, Vol. 1 (186), 2019.
20. Tal Ben-Nun and Torsten Hoefler, "Demystifying Parallel and Distributed Deep Learning: An In-Depth Concurrency Analysis", cs.LG, pp-1-47, 2019.
21. Bayu Adhi Tama and Sunghoon Lim, "A Comparative Performance Evaluation of Classification Algorithms for Clinical Decision Support Systems", Mathematics, MDPI, Vol. 8 (1814), pp. 1-25, 2020. (doi:10.3390/math8101814)
22. Martin Sarnovsky and Marek Olejnik, "Improvement in the Efficiency of a Distributed Multi-Label Text Classification Algorithm Using Infrastructure and Task-Related Data", Informatics, MDPI, Vol. 6 (12), pp-1-15, 2019. (doi:10.3390/informatics6010012)
23. Bin Hea, Yi Guana, and Rui Dai, "Classifying medical relations in clinical text via convolutional neural networks", Artificial Intelligence In Medicine, Elsevier Vol. 93, pp-43-49, 2019. (https://doi.org/10.1016/j.artmed.2018.05.001)