VOL 4 (2020) NO 4

e-ISSN : 2549-9904
ISSN : 2549-9610

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

Laying Chicken Algorithm (LCA) Based for Clustering

Iwan Tri Riyadi Yanto[#], Ririn Setiyowati[*],Nursyiva Irsalinda[**], Rasyidah[***], Tri Lestari[***]

[#] *Information System Department, Universitas Ahmad Dahlan, Yogyakarta, Indonesia*
E-mail: yanto.itr@is.uad.ac.id

[*]*Mathematics Department, Universitas Sebelas Maret, Indonesia*
E-mail: ririnsetiyowati@staff.uns.ac.id

[**]*Mathematics Department, Universitas Ahmad Dahlan, Yogyakarta, Indonesia*
E-mail: nursyiva.irsalinda@math.uad.ac.id

[***]*Information Technology Department, Politeknik Negeri Padang, Indonesia*
E-mail:rasyidah@pnp.ac.id, lestari@pnp.ac.id

***Abstract*— Many researches and applications related to fuzzy clustering are still very important and interesting. One of them is focusing on sensitive fuzzy based for clustering regarding to the initial centroid. If the initial centroid is bad, it will not converge to a good clustering results. This is due to iteration process easily stuck in local optimum. A stochastic global optimization is used to handle this optimization problem. Laying Chicken Algorithm (LCA) is one current of the stochastic global optimization as a multi swarm optimization. It adapts individual laying hens in the process of incubating their eggs to improve the Chicken Swarm Optimization (CSO). In this article, Fuzzy C-Means (FCM) and LCA were modified to repair local optimum of Fuzzy Clustering proposed. The LCA is used to find the global optimum of FCM. Data were redefined to be a matrix of identity as initial population chicken of LCA. Then, the improvement of population and solution were updated to get the optimal solution. The experiment was run by using UCI dataset. The comparison was conducted by evaluating the term of Davies Boulding index, rank index and accuracy of overall average. The results indicate that FCMLCA method has better performance than the compared methods.**

***Keywords*— fuzzy clustering; FCM; multi swarm optimization; LCA.**

I. INTRODUCTION

Clustering is a process of placing the data object to the mutually related set group called clusters. Clustering technique is applied in many application areas such as pattern recognition [1], segmentation market [2], and image processing [3], etc.

Kapoor et al. [4] classifies clustering algorithms into three types, namely hard, fuzzy, and probabilistic. The most famous hard clustering algorithm is K-mean clustering. This algorithm creates partition of the data object to the k clusters [5], [6]. This model does not correspond to the real data set which does not contain clear restriction among the clusters. To overcome this, the researcher used the theory of fuzzy for the clustering process. Fuzzy algorithm can map data objects partially into various clusters. Degree of membership in fuzzy class depends on the proximity of the data objects to the center of the class. In 1974, Bezdek introduced Fuzzy C Mean (FCM) as the most famous fuzzy algorithm[7].

The FCM algorithm is very effective. However, the iteration process of this algorithm easily crashes at the local optimum due to the random selection of data centers (centroids)[8]. To solve this problem, the evolutionary algorithms that have been successfully applied to clustering are genetic algorithm (GA) [9], simulated annealing (SA) [10], ant colony algorithm [11] and particle swarm optimization (PSO) [12]. PSO is an optimization algorithm based on population that can be implemented easily to resolve the optimization problem or problem that can be transformed into optimization function [13].

Li et al introduce combination of PSO with Fuzzy which is known as "*Fuzzy Particle Swarm Optimization*" (FPSO) [14]. FPSO can anticipate the problem of being stuck at the local optimum, but this algorithm requires a longer response time than FCM. In addition, PSO is a single swarm optimization method, namely a global optimization method that builds a calculation formula from the results of observing an individual's social behavior which is influenced by the behavior of other individuals in a social group [12], [13].

Method development is still being done to get better performance. Xianbing Meng et al developed a multi-swarm optimization method, namely Chicken Swarm Optimization (CSO). Compared to the single swarm optimization method, CSO can achieve better optimization results in robustness and accuracy [15], [16]. The continuation of CSO has been developed in 2017. Hosseini E developed Laying Chicken Algorithm (LCA). LCA is not a single swarm optimization. This algorithm adapts individual laying hens in the process of incubating their eggs. LCA is more simple because it is not influenced by other individuals such as in PSO so that the optimization problem solving time is shorter than other metaheuristic algorithms [17]. Therefore, LCA will be combined with FCM in solving clustering problem in this study.

The remainder of this paper is the FCM described in Section II. PSO is described in Section III. The FCMPSO is presented in Section IV briefly. In Section V, the principles of LCA are presented. Sections VI and VII present the proposed method and experimental results, respectively. The last section contains conclusions and further work

II. Principle Of Fuzzy C-Means

Fuzzy C Mean (FCM) is a clustering technique with a degree of membership to determine the presence of each data point in a group [13], [18]. Suppose there is n objects $o = \{o_1, o_2, \ldots, o_n\}$ in $\mathbb{R}^d$ dimensional space, FCM partition the objects in $c\ (1 < c < n)$ cluster with $z = \{z_1, z_2, \ldots, z_n\}$ cluster center or centroid. Fuzzy cluster objects are described as n rows and c column fuzzy matrix μ. The elements of the matrix μ in row i and column j are denoted by μ_{ij} which shows the membership function of the i-th object with the j-th cluster.

$$J_m = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^{m} d_{ij} \tag{1}$$

where m is a scalar which states the exponential weight with $m > 1$ and fuzziness control of the cluster results.

$$d_{ij} = \left\| o_i - z_j \right\| \tag{2}$$

d_{ij} is an Euclidean distance from object i to cluster j and subject to

$$\begin{gathered} \mu_{ij} \in [0,1] \forall i = 1,2, \ldots, n; \forall = 1,2, \ldots c. \\ \sum_{j=1}^{c} \mu_{ij} = 1;\ \forall i = 1,2, \ldots, n \\ 0 < \sum_{i=1}^{n} \mu_{ij} < n;\ \forall j - 1,2, \ldots, c \end{gathered} \tag{3}$$

III. Principle Of Particle Swarm Optimization

Particle Swarm Optimization (PSO) is an optimization method based on the social behaviour of organism that consists of individual actions and other influences in a social group [19]. PSO finds the optimal solutions through collaboration among individual in a swarm population. Each individual is called a particle or agent with two components (variables) namely position and velocity [20]. For a problem in the swarm with n dimensions and m particles then the position and velocity of i-th particle are denoted as in (4).

$$\begin{gathered} x_i - [x_{i1}, x_{i2}, \ldots, x_{in}]^T \\ V_i = [V_{i1}, V_{i2}, \ldots, V_{in}]^T, for\ i = 1,2, \ldots, m, \end{gathered} \tag{4}$$

PSO Algorithm is expressed as in (5):

$$X(t+1) = X(t) + V(t+1) \tag{5}$$

where,

$$V_i(t+1) = x\left(\omega V_i(t) + \phi_1(p(t) - x_i(t)) + \phi_2(g(t) - x_i(t))\right) \tag{6}$$

Vector parameter $x_i(t)$ is notated as an estimator from optimum value in t-th iteration $V_i(t)$ called the velocity vector which is the vector modification of the vector parameter (momentum for the next iteration). $p(t)$ is the best estimator achieved by the particle. $g(t)$ is the best outcome of the results ever achieved by all particles until t-th iteration. $p(t)$ and $g(t)$ are personal best and global best respectively. Coefficient Φ_1 and Φ_2 are two positive numbers in a certain range using uniform distribution with upper bound to determine the balance between the best individual estimator and the best swarm estimator. Coefficient ω is denoted as inertia adjustment effect and X as renewal coefficient.

IV. Principle of FCM-PSO

In Fuzzy Particle swarm optimization (FPSO), X is the particle position that shows the fuzzy relation of the data object $o = \{o_1, o_2, \ldots, o_n\}$ to the center of the cluster or centroid $z - \{z_1, z_2, \ldots, z_n\}$ X can be expressed in a matrix as in (7):

$$X = \begin{bmatrix} \mu_{ij} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nc} \end{bmatrix} \tag{7}$$

where μ_{ij} is membership function of i-th object with j-th cluster. Matrix position on each particle is the same as the fuzzy matrix μ in FCM, likewise the velocity of each particle is represented by a matrix of size n rows and c column with elements in the interval $[-1,1]$. The change in position and velocity of the particles are based on the matrix operations as in (8):

$$V(t+1) = \omega V(t) + c_1 r_1 (pbest(t) - X(t)) c_2 r_2 (gbest(t) - X(t)) \tag{8}$$

where, c_1 is positive constants of *pbest*, c_2 is positive constant of *gbest*, r_1 and r_2 are randomly valued in range [0, 1], w is weight

$$X(t+1) = X(t) + V(t+1) \tag{9}$$

with X is velocity particle and V is position particles then transformed into the formula (10):

$$X_{normal} = \begin{bmatrix} \mu_{ij} / \sum_{j=1}^{c} \mu_{(1j)} & \cdots & \mu_{ic} / \sum_{j=1}^{c} \mu_{(1j)} \\ \vdots & \ddots & \vdots \\ \mu_{n1} / \sum_{i=1}^{c} \mu_{(1j)} & \cdots & \mu_{nc} / \sum_{j=1}^{c} \mu_{(nj)} \end{bmatrix} \tag{10}$$

V. PRINCIPLE OF LAYING CHICKEN ALGORITHM

This algorithm is an algorithm that adapts the behavior of laying hens in turning eggs into chickens [17]. In continuous programming problems, eggs are represented as feasible solution while the chicken is the optimal solution. In farms, breeders use fake eggs to keep laying hens in their nests because laying hens prefer the same location or nest, not an empty nest without eggs. That behavior of laying hens is an idea to make a feasible initial solution which makes the initial population in an algorithm.

A. Laying Chicken Algorithm (LCA) Concept

1) Initial Population
Initial population of the settlement has been made near a feasible solution so that the next simulation is to define the initial which is a neighborhood with n dimension from R^n which is defined as in (11):

$$\sqrt{(x_1 - y_1)^2} - (x_2 - y_2)^2 + \dots + (x_n - y_n)^2 \le k \quad (11)$$

with X is initial solution, n-dimension vector denoted by Y and positive integer k. This algorithm will be more efficient when k gets smaller because not many solutions will be missed around the initial solution.

2) Population Improvement
Each solution in the population must be converted into X if the objective function is no better than the objective function of X. Actually, the value of the particles has been changed in vector direction that connects with X. This solution is modified as in (12):

$$x_{j+1} = x_j + ad_{j0} \quad (12)$$

with d_{j0} is a vector of x_j to x_0 and $f(x_j) < f(x_0)$, $0 < \alpha < 2k$ in the maximization problem. For $\alpha \to 0$ will not change the solution well, thus the interval $0 \le \alpha \le \frac{k}{4}$ is eliminated and we use the best α in interval $\frac{k}{4} \le \alpha \le 2k$.

3) Solution Improvement
The rotation of the egg by the hen where it turns the egg three or four times each day has inspired the final property of this simulation method. In this step except the best solution, all members of the population change slightly in (13).

$$(x_{i+1}, y_{i+1}) = (x_i \pm \varepsilon, y_i \pm \varepsilon) \quad (13)$$

with ε is a given positive integer. Each solution j which $x_j < x_{best}$ can be changed using formula (14):

$$(x_{i+1}, y_{i+1}) = (x_i + \varepsilon, y_i) \quad (14)$$

Meanwhile, for each solution k which $x_k > x_{best}$ can be changed in (15).

$$(x_{k+1}, y_{k+1}) = (x_k - \varepsilon, y_k) \quad (15)$$

In each iteration, the best solution can be stored and other selected solutions close to the best solution becomes the initial population in the next iteration.

B. Laying Chicken Algorithm (LCA)

1) Determine the initially feasible solution (x_0, y_0), the number of iteration N and any positive integers ε_i
2) Build the initial population around (x_0, y_0).
3) Each solution in step 2 which has an objective function no better than (x_0, y_0), must change direction toward (x_0, y_0) and will obtain the (x_{best}, y_{best}).
4) Change the whole solution except the best solution.
5) Objective function of the all solution updated supposes to be $(x_0, y_0) = (x_{best}, y_{best})$.
6) If $|f(x_{i\,best}) - f(x_{i+1\,best})| < \varepsilon_i$ or the number of iteration is more than N then the algorithm is terminated. Otherwise then return to step 2.

VI. PROPOSED METHOD

This research is a literature study by reading relevant books and journals. In this case, the theory of FCM, PSO, FPSO and CSO for clustering has been studied. Then, a new method is established by applying LCA to solve the FCM problem. The method to be developed is based on the LCA algorithm to modify iterations in determining the optimum point in the FCM to increase its accuracy value.

Here, LCA will be applied to clustering technique. We redefine the proposed method into chicken-domination identity and relationships. Let, X is chicken position shows *th*-fuzzy relation from datasets $o = \{o_1, o_2, \dots, o_n\}$, $z = \{z_1, z_2, \dots, z_n\}$, is a set of cluster centers. X is expressed by:

$$X = \begin{bmatrix} \mu_{ij} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nc} \end{bmatrix} \quad (16)$$

where, μ_{ij} is datasets of the i-th in j-th cluster constrains by $\mu_{ij} \subset [0,1], \forall i = 1,2,\dots,n;\ \forall j = 1,2,\dots,c$

Afterwards, we obtained the μ position matrix of each chicken, we performed an update of each particle using the $n \times c$ matrix in the range [0.1]. Furthermore, the position of the rooster, chicken and chicken is updated respectively based on the matrix operation using

$$\begin{gathered} \sqrt{(x_1 - y_1)^2} = (x_2 - y_2)^2 + \dots + (x_n - y_n)^2 \le k \\ x_{j+1} = x_j + ad_{j0} \\ (x_{i+1}, y_{i+1}) = (x_i \pm \varepsilon, y_i \pm \varepsilon) \end{gathered} \quad (17)$$

It is possible to violate constraints during the process of updating the position of the matrix. Then, evaluate the fitness function for a general solution is to use

$$f(X) = \frac{K}{J_m}$$

The notation J_m represents the FCM destination function and K is a constant.

The best model of LCA based clustering is indicated by the smaller J_m. It means that similarity of clustering and also higher individual fitness $f(X)$ have good performance. Furthermore, the steps of proposed technique are as follows:

1. Initializing parameters *P*,

2. Creating a chicken swarm with P chicken groups (roosters, hens and chicks)
3. Initializing X for the swarm of each chicken group,
4. Calculating the cluster centers,
5. Calculating the fitness values,
6. Updating the matrix position by improvement of LCA and solution improvement of LCA
7. Returning to step 4,
8. Until terminating condition is met.

The maximum iteration of fitness value in the LCA method is proposed as a termination condition.

VII. EXPERIMENTAL RESULT

This section presents the empirical work, performed by six datasets that documented from the UCI website. Then, datasets analysis on PC with Intel i5-8400 six core CPU 2.8 GHz and 8 GB RAM using MATLAB programming language with Windows 10, 64-bit as an operating system. The datasets were captured from the UCI that can be viewed in Table I

TABLE I
DATA SET

Data	The Number of Attribute	Data Size
Cancer	9	6147
E-coli	7	2352
Glass	10	214
Ionosphere	34	11934
Iris	4	600
Spam base	57	262257

All size of selected data is horizontally or vertically different. Data Selection process was done by using the proposed technique performance. Then, some datasets were modified by deleting the sample which had incomplete data. The experiment runs with the specified number of 50 with a population of 100 with a maximum number of iteration fuzziness index variations is $m \in [1.1,2.0]$, and then the average Davies Bouldin index, rank index and accuracy are calculated.

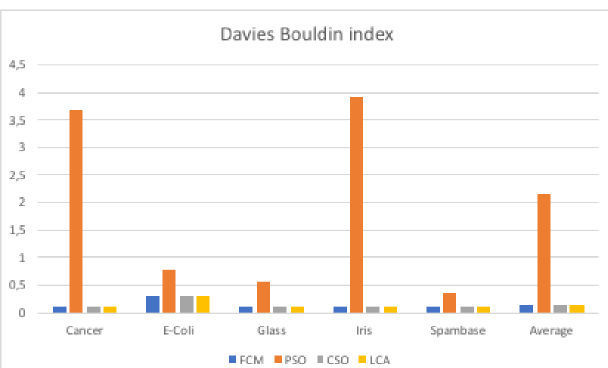

Fig 1. Illustrates that the LCA based clustering technique

Fig 1. Illustrates that the LCA based clustering technique has the lower Davied Bouldin index compared to CSO based on six datasets. It shows that the proposed approach is better than the baseline technique.

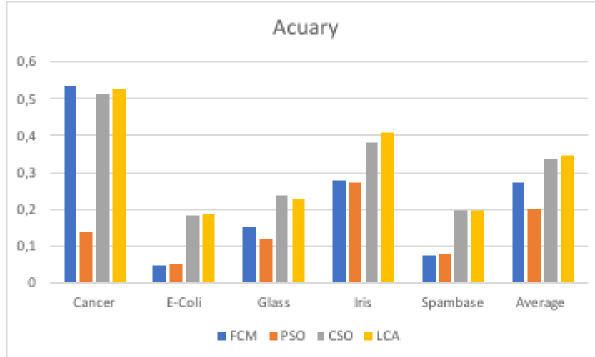

Fig. 2 Accuracy Comparison

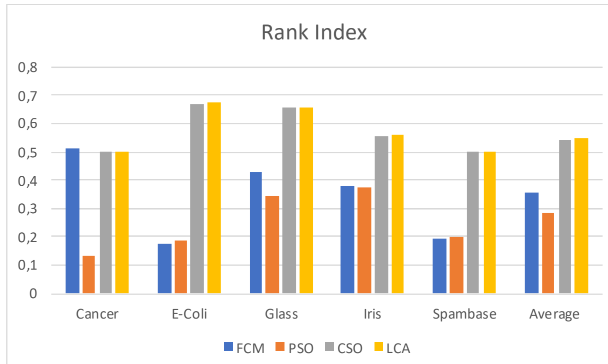

Fig. 3 Rank Index Comparison

The overall average of accuracy is showed in figure 2. It can be viewed that the proposed approach is better than CSO based clustering. Meanwhile based on rank index which is illustrated in figure 3, it can be seen that both techniques have almost similar result in average.

VIII. CONCLUSIONS

The clustering with emphasis on LCA technique has been implemented as an alternative approach in fuzzy clustering problem. Furthermore, the LCA based technique has been successfully utilized by using UCI datasets. It is indicated by the experiment results that the proposed approach has better performance in term of Davies Bouldin index, rank index and accuracy in overall average, respectively

REFERENCES

[1] Y. Qi, B. Luo, X. Wang, and L. Wu, "Load pattern recognition method based on fuzzy clustering and decision tree," in *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*, 2017, pp. 1–5.

[2] A. S. M. S. Hossain, "Customer segmentation using centroid based and density based clustering algorithms," in *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, 2017, pp. 1–6.

[3] M. A. Almahfud, R. Setyawan, C. A. Sari, D. R. I. M. Setiadi, and E. H. Rachmawanto, "An Effective MRI Brain Image Segmentation using Joint Clustering (K-Means and Fuzzy C-Means)," in *2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2018, pp. 11–16.

[4] A. Kapoor and A. Singhal, "A comparative study of K-Means, K-Means++ and Fuzzy C-Means clustering algorithms," in *2017 3rd International Conference on Computational Intelligence Communication Technology (CICT)*, 2017, pp. 1–6.

[5] K. P. Sinaga and M. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020.

[6] H. Zhou, “K-Means Clustering BT - Learn Data Mining Through Excel: A Step-by-Step Approach for Understanding Machine Learning Methods,” H. Zhou, Ed. Berkeley, CA: Apress, 2020, pp. 35–47.

[7] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers, 1981.

[8] H. Izakian and A. Abraham, “Fuzzy C-means and fuzzy swarm for fuzzy clustering problem,” *Expert Syst. Appl.*, vol. 38, no. 3, pp. 1835–1838, Mar. 2011.

[9] S. Irfan, G. Dwivedi, and S. Ghosh, “Optimization of K-means clustering using genetic algorithm,” in *2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN)*, 2017, pp. 156–161.

[10] S. Gavrilov, D. Zheleznikov, V. Khvatov, and R. Chochaev, “Clustering optimization based on simulated annealing algorithm for reconfigurable systems-on-chip,” in *2018 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, 2018, pp. 1492–1495.

[11] T. N. Reddy and K. P. Supreethi, “Optimization of K-means algorithm: Ant colony optimization,” in *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, 2017, pp. 530–535.

[12] C. Hua and W. Wei, “A Particle Swarm Optimization K-Means Algorithm for Mongolian Elements Clustering,” in *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2019, pp. 1559–1564.

[13] Y. Shen *et al.*, “Research on Swarm Size of Multi-swarm Particle Swarm Optimization Algorithm,” in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 2018, pp. 2243–2247.

[14] L. Li, X. Liu, and M. Xu, “A Novel Fuzzy Clustering Based on Particle Swarm Optimization,” in *2007 First IEEE International Symposium on Information Technologies and Applications in Education*, 2007, pp. 88–90.

[15] X. Meng, Y. Liu, X. Gao, and H. Zhang, “A New Bio-inspired Algorithm: Chicken Swarm Optimization,” in *Advances in Swarm Intelligence SE - 10*, vol. 8794, Y. Tan, Y. Shi, and C. C. Coello, Eds. Springer International Publishing, 2014, pp. 86–94.

[16] Suprihatin, I. T. R. Yanto, N. Irsalinda, T. Purwaningsih, Haviluddin, and A. P. Wibawa, “A performance of modified fuzzy C-means (FCM) and chicken swarm optimization (CSO),” in *2017 3rd International Conference on Science in Information Technology (ICSITech)*, 2017, pp. 171–175.

[17] E. Hosseini, “Laying Chicken Algorithm: A New Meta-Heuristic Approach to Solve Continuous Programming Problems,” *J. Appl. Comput. Math.*, vol. 06, no. 01, pp. 1–8, 2017.

[18] Y. Xiaojing, J. Qingju, and L. Xinke, “Center Particle Swarm Optimization Algorithm,” in *2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 2019, pp. 2084–2087.

[19] D. Wang, D. Tan, and L. Liu, “Particle swarm optimization algorithm: an overview,” *Soft Comput.*, vol. 22, no. 2, pp. 387–408, 2018.

[20] F.-S. Wang and L.-H. Chen, “Particle Swarm Optimization (PSO) BT - Encyclopedia of Systems Biology,” W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, Eds. New York, NY: Springer New York, 2013, pp. 1649–1650.