Fuzzy c-Means with Improved Particle Swarm Optimization

Jie Li
Department of Risk Engineering
University of Tsukuba
Ibaraki, Japan
jelly.jelly@hotmail.com

Yasunori Endo
Faculty of Engineering, Information and Systems
University of Tsukuba
Ibaraki, Japan
endo@risk.tsukuba.ac.jp

***Abstract*—Fuzzy clustering algorithm has become a relevant research field of unsupervised learning due to that the uncertainties between patterns can be described more accurately. Based on objective function, fuzzy clustering algorithm uses a constrained optimization mathematical problem to represent the clustering problem. It then determines the division of data sets and fuzzy clustering results by solving the optimization problem. Fuzzy c-means (FCM) is one of the best known for its simplicity and efficiency. However, it shows some weaknesses, particularly its tendency to fall into local optima and dependence on initial values. Particle Swarm Optimization (PSO) is one of the heuristic methods that usually implemented on function optimization problems since it has a robust global search capability. In this paper, a new concept of worst position is introduced to PSO that gives a chance for particles to change flying directions. Moreover, new hybrid algorithms based on FCM and improved PSO with worst position (PSOWP) both in L_1 norm and L_2 norm are proposed, which avoid falling into local optimum with faster convergence speed.**

***Index Terms*—Fuzzy Clustering, Fuzzy c-Means(FCM), Particle Swarm Optimization(PSO)**

I. Introduction

Fuzzy c-Means (FCM) is the most common method in fuzzy clustering. In 1981, FCM was first proposed by Bezdek [4] added fuzzy factor to hard means clustering. FCM is an effective clustering method. However, because of its random initial value selection, easily being trapped in local optima and low robustness, there is plenty of researches to solve these shortcomings. The fuzzy clustering problem is an optimization problem. To obtain optimal solutions to such large problems could be complex, therefore, approximate methods are required. Recently, evolutionary algorithms such as Genetic Algorithm (GA) [3], Colony Ant Optimization (ACO) [7], and Particle Swarm Optimization (PSO) [6] have been successfully applied. Runkler and Katz [8] applied PSO to cluster data using fuzzy clustering. They introduced two new methods to minimize the two reformulated versions of FCM objective function by PSO. On the other hand, to make FCM more robust, Bobrowski and Bezdek proposed clustering based on L_1 and L_∞ norm [9]. On this basis, Jajuga [10] and Miyamoto [11] proposed efficient algorithms to simplify the calculation complexity on L_1 norm that applies to large datasets.

Particle Swarm Optimization (PSO) was proposed based on the social behaviors of birds flocking and fish schooling. PSO is usually implemented on function optimization problems since it has a strong global search capability and fast converge speed. Many changes in PSO have been proposed to improve its results and convergence time. Shi and Eberhart [13] introduced inertia weight to velocity updating equation, in order to achieve balance in global and local searching with dynamically adjusted inertia weight, which is called standard PSO. Clere [18] introduced constriction factor to velocity updating equation that ensures algorithm converges and relaxes the constraints of velocity. Liu and Xu proposed a version of the PSO based on human behavior algorithm [19] which gives a chance to adjust the direction of the particle's velocity and achieve better performance on convergence time. In hierarchical PSO [20] a particle is influenced by its own so far best position and by the best position of the particle that is directly above it in the hierarchy. New Particle Swarm Optimization(NPSO) [21] is based on experience that an individual learns both from previous best and individuals' mistakes.

On the other hand, FCM can also be attributed to optimization problems under the constraint of membership. Therefore, many PSO-based fuzzy clustering algorithms were proposed [15]–[17]. But in most of these algorithms, the particle is encoded by cluster centers, less of them concerned the method of encoding particle as membership. Izakian and Abraham proposed a hybrid fuzzy clustering algorithm based on FCM and PSO and encoded the particle by membership, called FCM-PSO [25]. Their experiments showed the better potential of the PSO-based method on clustering accuracy than FCM.

The main contributions of this paper are the introduction of new conception to standard PSO and proposals of new methods for fuzzy clustering problems based on objective function optimization by hybridizing FCM and PSOWP. The expectations of proposed methods are (i) better and more stable solutions, that by expanding exploration ability of PSO with a worst position in solution space to jump out of initial value sensitivity and local optima of FCM both in L_1 and L_2 norm; (ii) faster convergence speed because of chances to change directions of particles provided by worst positions. This paper is organized as follows. In Section II, the origi-

978-1-7281-6932-3/20/$31.00 ©2020 IEEE

nal FCM and FCM-PSO for fuzzy clustering algorithms are reviewed. The proposed PSO with worst position (PSOWP), a hybrid between FCM and PSO with the worst position (FCM-PSOWP) and L_1 FCM-PSOWP are given in Section III. Experimental results are provided in Section IV. And some conclusions are given in Section V.

II. Related Works

A. Fuzzy c-means (FCM)

Ruspini [5] introduced fuzzy mathematical theory into clustering and proposed concept of fuzzy division, that membership $u_{ki} \in \{0, 1\}$ was improved as $u_{ki} \in [0, 1]$; Bezdek [4] introduced fuzzy parameter m $(m > 1)$ and made membership to be fuzzificated by m in 1973, which is the well-known and widely used Fuzzy c-Means (FCM).

FCM treats clustering as soft clusters to which each object has a membership represents a higher similarity between the object and cluster.

Let objects considered as a set of n vectors $X = \{x_1, x_2, ..., x_n\}$ for clustering into c clusters. The cluster center is denoted by $Y = \{y_1, y_2, ..., y_c\}$. A membership matrix $U = (u_{ki})$ is known as fuzzy partition matrix to describe the degree of objects belong to clusters, and the membership element $u_{ki} \in [0, 1]$. Euclidean distance in L_2 norm $d_{ki} = ||x_k - y_i||^2$ is defined as the distance between object x_k and cluster center y_i. According to above definitions, the objective function of FCM can be computed with membership and Euclidean distance as follow equation:

$$\text{minimize } J_{FCM}(U, Y) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ki}^m ||x_k - y_i||^2 \quad (1)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{i=1}^{c} u_{ki} = 1 \quad (k = 1, 2, ..., n) \\ & 0 < \sum_{k=1}^{n} u_{ki} < n \quad (i = 1, 2, ..., c) \end{aligned} \quad (2)$$

Here, m is the fuzzy parameter which satisfies $m > 1$. Eq.(1) is the objective function of FCM, Eq.(2) is the constraint of membership. Cluster centers Y and membership U are alternately updated with equations (3) and (4) obtained by the Lagrange method.

$$y_i = \frac{\sum_{k=1}^{n} (u_{ki})^m x_k}{\sum_{k=1}^{n} (u_{ki})^m} \quad (3)$$

$$u_{ki} = \frac{(\frac{1}{d_{ki}})^{\frac{1}{m-1}}}{\sum_{j=1}^{c} (\frac{1}{d_{ki}})^{\frac{1}{m-1}}} \quad (4)$$

The iteration of FCM will stop when $|y_i^{(t+1)} - y_i^{(t)}| < \epsilon$, which means the change of cluster centers on current iteration and next iteration is smaller than a threshold ϵ, here ϵ is a termination criterion between 0 and 1, t is the iteration steps. This procedure converges to a local minimum or a saddle point of $J_{\text{FCM}}(U, Y)$. Therefore, FCM algorithm is composed of following steps.

Algorithm 1 FCM

1: Give the number of cluster c.
2: Initialize membership u_{ki} randomly.
3: Update cluster center Y by Eq.(3) with fixing U.
4: Update membership U by Eq.(4) with fixing Y
5: If Y changes from previous Y, go back to Step 3. Otherwise, stop.

Due to non-convex objective function of FCM, and the algorithm is implemented by the iterative hill-climbing method, it is particularly sensitive to initial values, in addition, it is also easily to fall into local optimum that is hard to obtain the global optimal solution.

B. Particle Swarm Optimization (PSO)

PSO is a inspired the behavior of bird flocking or fish schooling, that is first proposed by Kennedy and Eberhart [6] in 1995. Each bird is a particle in the algorithm, which is a feasible solution to the problem. The particle constantly changes its position and flight speed until finally finds optimal solution.

PSO first randomly generates a population of particle with random velocities in solution space, positions of particles are potential solutions with a fitness value determined by objective function. Each particle flies in the solution space, and its direction and distance are determined by velocity. The positions of the next generation particles are determined by combining position and velocity. In each iteration, particles have to store two information, one is the optimal solution found in the particle's historical updating $Pbest$, and the optimal solution found by the entire population so far $Gbest$. Each particle will modify velocities according to personal best position $Pbest$ and globa best position $Gbest$.

Considering the l-th particle among total N particles, its position and velocity at iteration t are denoted by $X_l(t)$ and $V_l(t)$, and its personal best position in t-th iteration is $Pbest_l(t)$. Thus the new velocity and position of l-th particle in iteration $t+1$ will be calculated by using following:

$$\begin{aligned} V_l(t+1) = \omega V_l(t) + c_1 r_1 (Pbest_l(t) - X_l(t)) \\ + c_2 r_2 (Gbest - X_l(t)) \end{aligned} \quad (5)$$

$$X_l(t+1) = X_l(t) + V_l(t+1) \quad (6)$$

Here $1 \leq l \leq N$, ω is inertia weight, c_1, c_2 are positive constants called acceleration coefficients that control the influence of $Pbest$ and $Gbest$ while searching solutions, r_1 and r_2 are random values in range $[0, 1]$.

C. Hybrid FCM-PSO for Fuzzy Clustering (FCM-PSO)

Pang [22] proposed a modified PSO for TSP [23] called fuzzy particle swarm optimization (FPSO). According to Peng, position and velocity of particles redefined to represent the fuzzy relationship between variables. Then Izakian and Abraham [24] applied FPSO into fuzzy clustering problem. They presented a hybrid between FCM and FPSO abbreviated as FCM-PSO.

In FPSO, the position of particle X_l shows the fuzzy relation from set of objects to set of cluster centers, which is membership matrix:

$$X_l = \begin{bmatrix} u_{11} & \cdots & u_{1c} \\ \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{nc} \end{bmatrix}$$

Since the solution is a matrix, the equations for updating velocities and positions of particles should base on matrix operations.

$$V_l(t+1) = \omega V_l(t) + c_1 r_1 (P\text{best}_l(t) - X_l(t)) + c_2 r_2 (G\text{best} - X_l(t)) \tag{7}$$

$$X_l(t+1) = X_l(t) + V_l(t+1) \tag{8}$$

In order not to violate the constraint (2), it is necessary to normalize position matrix. Firstly, make all negative elements to be zeros, if all elements in a row are zeros, they need to be revalued with random number between $[0, 1]$. Then the matrix will be normalized as follow:

$$X_l = \begin{bmatrix} u_{11}/\sum_{i=1}^{c} u_{1i} & \cdots & u_{1c}/\sum_{i=1}^{c} u_{1i} \\ \vdots & \ddots & \vdots \\ u_{n1}/\sum_{i=1}^{c} u_{ni} & \cdots & u_{nc}/\sum_{i=1}^{c} u_{ni} \end{bmatrix}$$

The fitness function is defined as Eq.(9) to evaluate the generalized solutions.

$$f(X) = \frac{K}{J_{FCM}} \tag{9}$$

Here, J_{FCM} is the objective function of FCM and K is a constant which controls fitness fucntion is at a large value when $1/J_{FCM}$ is too small to compare fitness function values. Fixing K The smaller J_{FCM} is, the better the clustering effect and bigger $f(X)$ is.

According to Izakian and Abraham [25], FCM is much faster than FPSO since it requires fewer function evaluations, whereas, it still has high possibility to fall into local optima. Thus comes the idea of combining FCM with FPSO to overcome the shortcoming. They presented a hybrid between FCM and FPSO abbreviated as FCM-PSO. The steps of FCM-PSO are given by Algorithm 2.

Algorithm 2 FCM-PSO

1: Initialize the parameters of FCM and PSO including particle population size N, inertia weight ω, r_1, r_2 and fuzzy parameter m .
2: Create a swarm with N particles.
3: Initialize membership matrix X, velocity V for each particle and Gbest for the whole swarm.
4: Calculate the cluster center for each particle by Eq.(3)
5: Calculate the fitness value of each particle by Eq.(9).
6: Compare with previous fitness, choose the smaller as new Pbest.
7: Choose the smallest fitness so far as Gbest.
8: Update the velocity for each particle by Eq.(7).
9: Update the position for each particle by Eq.(8).
10: If meet the terminating condition then go to step 11, otherwise, go back to step 4.
11: Calculate cluster center by Eq.(3).
12: Calculate membership by Eq.(4).
13: Calculate the Pbest of each particle.
14: Set the Gbest of the swarm.
15: If FCM met the terminating condition then stop, otherwise, go back to step 11.
16: If met the terminating condition then stop, otherwise, go back to step 4.

III. Proposed methods

A. Particle Swarm Optimization with Worst Position (PSOWP)

In standard Particle Swarm Optimization (PSO), particles only learn from the best positions with Pbest and Gbest, which is an ideal social condition. However, in the real social condition, there also exist some bad individual behaviors that might have effects on around particles. If these bad behaviors can be learned or marked, it will be beneficial to find optimal solutions. Liu and Xu have proposed PSO based on Human Behavior [19] to improve the performance of standard PSO, whereas, they only introduced global bad particles without considering the personal bad particle. Therefore, considering both personal worst and global worst particle behavior, the new method is proposed.

In proposed method PSO with Worst Position (PSOWP), the personal worst particle Pworst and global worst particle Gworst, whose fitness function value is the worst for l-th particle so far and whose fitness function value in the entire swarm population is worst in each iteration respectively are denoted as follow:

$$I^* = \arg\max_l f(X_l),$$
$$\text{set } P\text{worst} = X_{I^*}, 1 \le l \le N \tag{10}$$

$$I^{**} = \arg\max_l f(P\text{worst}_l),$$
$$\text{set } G\text{worst} = P\text{worst}_{I^{**}}, 1 \le l \le N \tag{11}$$

The process of PSOWP algorithm then could be described as same as standard PSO that particles fly through the solution space towards better position with alternate updating of velocity V_l and position X_l. This is accomplished using follow Eq.(12) and (13):

$$\begin{aligned} V_l(t+1) = \omega V_l(t) &+ c_1 r_1 (Pbest_l(t) - X_l(t)) \\ &+ c_2 r_2 (G\text{best} - X_l(t)) \\ &+ r_3 (P\text{worst}_l(t) - X_l(t)) \\ &+ r_4 (G\text{worst} - X_l(t)) \end{aligned} \tag{12}$$

$$X_l(t+1) = X_l(t) + V_l(t+1) \tag{13}$$

To make full role of Pworst and Gworst, new learning coefficients r_3 and r_4 are introduced. Considered both positive and negative effects on particles, r_3 and r_4 are random numbers which obey the standard normal distribution, that is $r_3, r_4 \in N(0,1)$. There are mainly four kinds of effect with the different signs of r_3 and r_4:

1. When $r_3, r_4 > 0$, the worst positions are considered having positive enhancement on velocity, which can improve particles' exploration capacity.
2. When $r_3, r_4 < 0$, the worst positions are considered having negative enhancement on velocity, which can decrease the flying velocity of particles. It is also beneficial to improve exploration ability.
3. In the case of $r_3 > 0, r_4 < 0$ or $r_3 < 0, r_4 > 0$, these represent the personal worst position and global worst position do not have positive or negative effects at the same time, that still can adjust the exploration ability, which is called neutralized enhancement.
4. Conversely, when $r_3 = r_4 = 0$, it is considered that no worst position has an effect on velocity. This situation is the same as the standard PSO.

Each positive, negative and neutralized enhancement provides a chance for particle's velocity to change directions, which plays a key role in increasing population diversity not be trapped in local optima, improve the convergence speed therefrom. Therefore, the positive/negative/neutralized component performs a proper tradeoff in the exploration.

The algorithm of PSOWP is composed of the following steps:

Algorithm 3 PSOWP

1: Initialize the parameters of PSOWP including particle population size N, inertia weight ω, r_1, r_2, r_3, r_4, c_1, c_2.
2: Create a swarm with N particles.
3: Initialize position X and velocity V for each particle, Gbest and Gworst are also initialized for the whole swarm.
4: Calculate fitness function value of each particle.
5: Compare with previous fitness, choose the smaller as new Pbest, the bigger one as Pworst.
6: Choose the smallest fitness so far as Gbest, the biggest one as Gworst.
7: Update the velocity for each particle by Eq.(12).
8: Update the position for each particle by Eq.(13).
9: If meet the terminating condition then stop, otherwise, go back to step 4.

B. Fuzzy c-Means Hybrid Particle Swarm Optimization with Worst Position (FCM-PSOWP)

The PSOWP algorithm introduces Pworst component for each particle and Gworst for the whole swarm with the goal of improving the performance of PSO. On the other hand, to overcome the disadvantage of FCM that is easily trapped into local optima and in order to lead to more stable clustering results with less convergence time, the algorithm Fuzzy c-Means Hybrid Particle Swarm Optimization with Worst Position (FCM-PSOWP) is proposed.

In FCM-PSOWP clustering, we need to define a fitness function to evaluate particles' positions. For the purpose of the comparison with the objective function value of FCM, thus the FCM objective function is used as the fitness function of the proposed method.

$$f(X) = \text{minimize} \quad J_{FCM}(U, Y) \tag{14}$$

[H] **Algorithm 4** FCM-PSOWP

1: Initialize the parameters of FCM and PSOWP including particle population size N, inertia weight ω, r_1, r_2, r_3, r_4 and fuzzy parameter m .
2: Create a swarm with N particles.
3: Initialize membership matrix X, velocity V for each particle and Gbest and Gworst for the whole swarm.
4: Calculate the cluster center for each particle by Eq.(3)
5: Calculate the fitness value of each particle by Eq.(14).
6: Compare with previous fitness, choose the smaller as new Pbest, the bigger one as Pworst.
7: Choose the smallest fitness so far as Gbest, the biggest one as Gworst.
8: Update the velocity for each particle by Eq.(12).
9: Update the position for each particle by Eq.(13).
10: If meet the terminating condition then go to step 11, otherwise, go back to step 4.
11: Calculate cluster center by Eq(3).
12: Calculate membership by Eq.(4).

13: Calculate the Pbest and Pworst of each particle.
14: Set the Gbest and Gworst of the swarm.
15: If FCM met the terminating condition then stop, otherwise, go back to step 11.
16: If met the terminating condition then stop, otherwise, go back to step 4.

FCM-PSOWP works by running the PSOWP until it reaches the terminating condition. Then it runs FCM try to find a better solution. When FCM achieves its terminating condition, FCM-PSOWP checks whether its stopping condition was meet or not. If not, it will run PSOWP again.

C. L_1 norm Fuzzy c-Means Hybrid Particle Swarm Optimization with Worst Position (L_1FCM-PSOWP)

Commonly, standard clustering algorithms such as K-Means and Fuzzy c-Means use Euclidean distance in L_2 norm as dissimilarity degree. In that case, the distance is the squared when data individuals are connected in a straight line, which makes it easy to be affected by outliers. The advantage of using L_1 norm distance is increasing robustness to outliers.

When L_1 norm distance is used as dissimilarity degree, the fitness function (the objective function of FCM) can be rewritten like Eq.(15), and the constraint condition is the same as the above algorithm.

$$\begin{aligned} f(X) &= \text{minimize} \quad J_{L_1FCM}(U,Y) \\ &= \sum_{k=1}^{n}\sum_{i=1}^{c}\sum_{j=1}^{p} u_{ki}^{m}|x_{kj}-y_{ij}| \end{aligned} \tag{15}$$

Here, L_1 norm distance is $d_{ki} = |x_k - y_i|$, and the objective function is expanded to every dimension so that it is easier to calculate cluster centers.

By introducing L_1 norm distance, the algorithms of Fuzzy c-Means Hybrid Particle Swarm Optimization with Worst Position in L_1 norm are described as Algorithm 5. According to Miyamoto and Agusta [11] , the component of cluster center can be calculated by a linear search on the derivative of the objective function which is remarkably simple. Thus in order to reduce computational complexity of finding cluster centers in L_1 norm, Algorithm 6 that offered by [11] is used in this paper.

[H] **Algorithm 5** L_1FCM-PSOWP

1: Initialize the parameters of L_1 FCM and PSOWP including particle population size N, inertia weight ω, r_1, r_2, r_3, r_4 and fuzzy parameter m .
2: Create a swarm with N particles.
3: Initialize membership matrix X, velocity V for each particle and Gbest and Gworst for the whole swarm.
4: Calculate the cluster center for each particle by Algorithm 6.
5: Calculate the fitness value of each particle by Eq.(15).
6: Compare with previous fitness, choose the smaller as new Pbest, the bigger one as Pworst.
7: Choose the smallest fitness so far as Gbest, the biggest one as Gworst.
8: Update the velocity for each particle by Eq.(12).
9: Update the position for each particle by Eq.(13).
10: If meet the terminating condition then go to step 11, otherwise, go back to step 4.
11: Calculate cluster center by Algorithm 6.
12: Calculate membership by Eq.(4).
13: Calculate the Pbest and Pworst of each particle.
14: Set the Gbest and Gworst of the swarm.
15: If L_1 FCM met the terminating condition then stop, otherwise, go back to step 11.
16: If met the terminating condition then stop, otherwise, go back to step 4.

Algorithm 6 Update cluster center Y

1: Ascendingly sort objects x_k with each dimension j, recorded as $x_{q(k)j}$.
2: Set $S = -\frac{1}{2}\sum_{k=1}^{n} u_{q(k)j}, r = 0$.
3: While S changes from - to +, $S = S + u_{q(k)j}, r = r + 1$.
4: Update optimal cluster center $y_{ij} = x_{q(k)j}$.

IV. Numerical Experiments

The proposed methods FCM-PSOWP and L_1FCM-PSOWP in this work are compared to FCM and FCM-PSO. Two artificial datasets and three real datasets were employed is this performance experiment. The algorithm are implemented on a computer with an Intel Core i5-6200U, 2.3GHz processor and 8 GB RAM, running a Windows operation system.

A. Clustering Validity Index

1) Adjusted Rand Index: To quantitatively verify the performance of clustering methods, the Adjust Rand Index (ARI) [26] is used. The ARI computes the similarity degree between two clusters by considering all pairs of samples and counting pairs that are assigned in the same or different clusters in the predicted and true clusters.

Let a data set of n objects be classified in $G = \{G_1, G_2, ..., G_c\}$ with c clusters, and $H = \{H_1, H_2, ..., H_r\}$ with r clusters. The original ARI using the permutation model as Eq.(16):

$$ARI = \frac{\sum_{i,j}\binom{n_{ij}}{2} - [\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}}{\frac{1}{2}[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}] - [\sum_i \binom{a_i}{2}\sum_j \binom{b_j}{2}]/\binom{n}{2}} \tag{16}$$

Where n_{ij}denotes the number of objects in common G_i and H_j that is $n_{ij} = |G_i \cap H_j|$. $\sum_{i,j}\binom{n_{ij}}{2}$ is defined as the number of pairs of objects in the same cluster in G_i and same cluster in H_j. a_i and b_j are the number of objects that belong to cluster G_i and H_j respectively. ARI takes the value between

0 and 1, that is ARI$\in [0,1]$. The more similar the two clusters are, the closer to 1 the ARI is.

2) Xie-Beni Index: Xie-Beni index (XB) [27] is an index of fuzzy clustering, which measures the overall average compactness and separation of fuzzy clustering. Consider dataset $X = \{x_1, x_2, ..., x_n\}$ with $Y = \{y_1, y_2, ..., y_c\}$ the cluster centers and $U = (u_{ki})$ as the membership of object x_k belonging to cluster i. Then the definition of XB is presented as Eq.(17).

$$XB = \frac{\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ki}^2 ||x_k - y_i||^2}{n \min\limits_{i,j,i\neq j} ||y_i - y_j||^2} \tag{17}$$

In fact, the numerator of XB equation is the objective function of FCM. It should be noticed that if different objective function is used to be optimized, it may wish to modify the compactness measure so that minimizing XB is compatible with fuzzy clustering objective function. A smaller Xie-Beni index indicates a partition in which all the clusters are overall compact, and separate to each other.

B. Experiments with Artificial Data

Two artificial datasets are used to test the clustering performance of FCM, FCM-PSO, and proposed methods. These two datasets show very different overlapping levels and different class shapes. Dataset 1 has three groups data generated by Gaussian distributions, and Dataset 2 has four groups of data generated by Gaussian distributions. Table I and Table II show the parameters for generating the datasets and Fig.1 and Fig.2 show the shapes of the datasets, respectively.

In each method of FCM-PSO, FCM-PSOWP and L_1 FCM-PSOWP, 15 particles are generated and the inertial weight $\omega = 0.9$, acceleration coefficients $c_1 = c_2 = 2.0$, with 100 iterations as stopping condition. All fuzzy parameters m of four methods are set as $m = 2$. Experiments are carried out under the above conditions with the numbers of cluster $c = 3$ for Dataset 1, and $c = 4$ for Dataset 2.

Table III shows the results for Adjusted Rand Index (ARI), Xie-Beni Index (XB), and CPU running time (sec) for two datasets. Best results are highlighted in underline. The tests show that computation time (CPU time) for the methods with PSO and PSOWP on two artificial datasets are significantly lower than that for sole FCM. While proposed FCM-PSOWP and L_1FCM-PSOWP work at higher speed than previous FCM-PSO. The ARI values show the results of FCM-PSOWP and L_1FCM-PSOWP are closer to real labels on two datasets, respectively. XB values of FCM-PSOWP for Dataset 1 and L_1FCM-PSOWP for Dataset 2 indicate a more compact and better-separated clustering partition without considering true labels.

Fig.3 summarizes the effect of varying the number of clusters for proposed methods for Dataset 1. It is expected that ARI should get the biggest value, and XB should reach the smallest value when the number of cluster $c = 3$, which means both of the proposed methods successfully divided data into correct number of groups.

TABLE I
DATASET 1 - PARAMETERS OF GAUSSIAN DISTRIBUTION

Group 1				
$\mu_1 = 0,$	$\mu_2 = 0,$	$\sigma_1^2 = 25,$	$\sigma_2^2 = 15,$	$n = 120$
Group 2				
$\mu_1 = 15,$	$\mu_2 = 5,$	$\sigma_1^2 = 25,$	$\sigma_2^2 = 10,$	$n = 100$
Group 3				
$\mu_1 = 5,$	$\mu_2 = 15,$	$\sigma_1^2 = 25,$	$\sigma_2^2 = 25,$	$n = 100$

TABLE II
DATASET 2 - PARAMETERS OF GAUSSIAN DISTRIBUTION

Group 1				
$\mu_1 = 0,$	$\mu_2 = 0,$	$\sigma_1^2 = 100,$	$\sigma_2^2 = 2,$	$n = 100$
Group 2				
$\mu_1 = 0,$	$\mu_2 = -10,$	$\sigma_1^2 = 100,$	$\sigma_2^2 = 2,$	$n = 100$
Group 3				
$\mu_1 = 5,$	$\mu_2 = 0,$	$\sigma_1^2 = 1,$	$\sigma_2^2 = 100,$	$n = 100$
Group 4				
$\mu_1 = 9,$	$\mu_2 = 0,$	$\sigma_1^2 = 1,$	$\sigma_2^2 = 100,$	$n = 100$

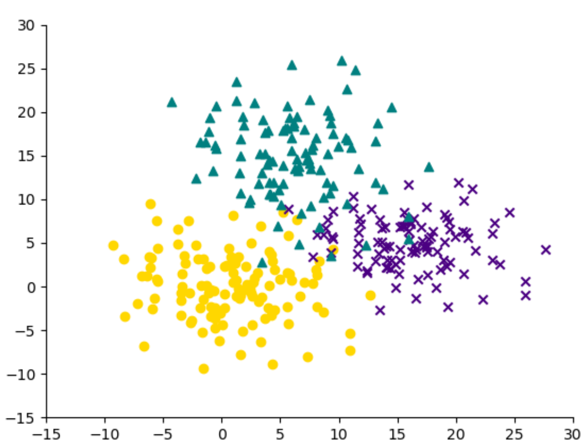

Fig. 1. Dataset 1 - overlapping data generated by Gaussian distributions

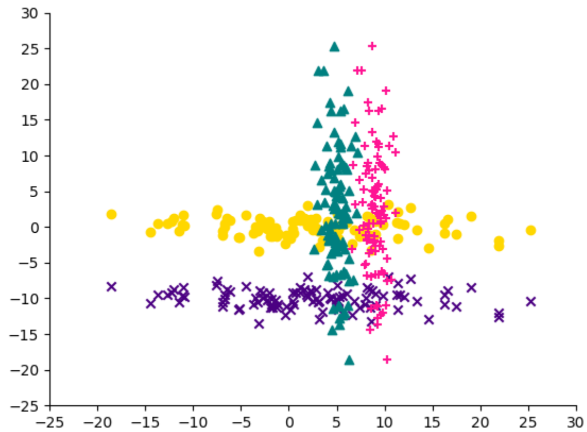

Fig. 2. Dataset 2 - overlapping data generated by Gaussian distributions

TABLE III
RESLUTS OF TWO ARTIFICIAL DATASETS

		FCM	FCM-PSO	FCM-PSOWP	L_1FCM-PSOWP
Dataset 1	ARI	0.8237	0.8336	0.8336	0.8238
	XB	0.0978	0.0962	0.0962	0.0183
	CPU time(sec)	0.9980	41.0899	15.1014	26.0181
Dataset 2	ARI	0.2089	0.1988	0.2128	0.3515
	XB	0.1841	0.3485	0.1815	0.3611
	CPU time(sec)	1.9378	72.5749	25.8641	33.7611

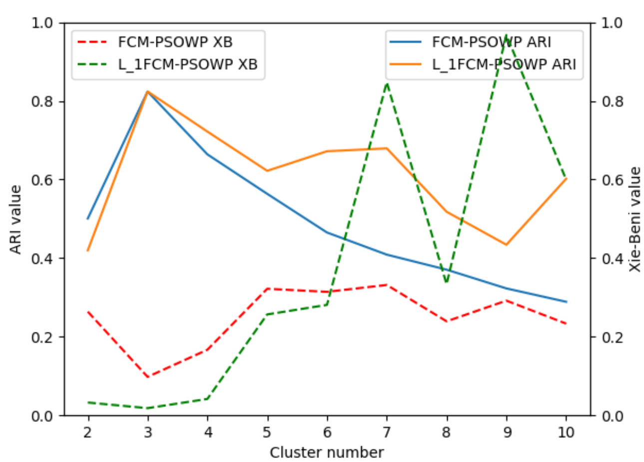

Fig. 3. ARI and XB curves on variable number of clusters of proposed methods on Dataset 1

C. Experiments with Real Data

Three datasets are selected from UCI dataset repository :

- Fisher's Iris Dataset consists of three species of iris flowers. There are 50 objects with four features in each species.
- Wine Dataset consists of 178 objects with 13 features in 3 types of wine.
- Breast Cancer Wisconsin (Diagnostic) Dataset (BCWD) consists of 569 samples and 30 features in 2 types that 212 were malignant and 357 began.

Table IV shows the results for Adjusted Rand Index (ARI), Xie-Beni Index (XB) and CPU running time (sec) for three real datasets. Best results are highlighted in underline. Since result differences among FCM, FCM-PSO, and FCM-PSOWP are small, the decimal points are token to the last four digits. L_1FCM-PSOWP shows the best ARI values associated with the lowest XB values for Iris and Wine datasets. FCM-PSO and FCM-PSOWP get the same well performance on BCWD dataset, while FCM-PSOWP is much faster than FCM-PSO. However, overall it cannot be denied that FCM is still the fastest one among four methods.

TABLE IV
RESLUTS OF REAL DATASETS

		FCM	FCM-PSO	FCM-PSOWP	L_1FCM-PSOWP
Iris	ARI	0.7294	0.7294	0.7294	0.8479
	XB	0.1371	0.1371	0.1326	0.0854
	CPU time(sec)	0.4791	17.8320	9.6507	12.0329
Wine	ARI	0.3539	0.3711	0.3711	0.4063
	XB	0.1257	0.1226	0.1217	0.1056
	CPU time(sec)	0.5110	21.2645	14.7281	18.2903
BCWD	ARI	0.6829	0.6888	0.6888	0.5835
	XB	0.4793	0.4731	0.4731	0.9258
	CPU time(sec)	1.0830	47.0621	20.5710	32.4029

V. CONCLUSIONS

This paper introduced a new concept that is the worst position into standard PSO, which gives particles a chance to change their flying directions and has contributed to the improvement in the exploration ability of particles to jump out of the local optima, which is beneficial to the global optimization, called PSOWP. On the basis of PSOWP, new hybrid algorithms for fuzzy clustering both in L_2 norm and L_1 norm, FCM-PSOWP, and L_1 FCM-PSOWP are proposed.

The clustering performances of FCM-PSOWP and L_1 FCM-PSOWP are estimated in comparison with two previous methods, FCM and FCM-PSO, with two artificial and three real datasets. Experimental results showed PSOWP-based methods proposed in this paper achieved higher accuracy than FCM and better clustering performance with faster computation speed than FCM-PSO. It can be believed that proposed methods solved the problem of time-consuming convergence of FCM-PSO, and can obtain better results with a short convergence time than FCM.

On the other hand, in spite of better performances of proposed methods than pervious methods consider in the comparison, on the experiments for real datasets with plenty of attributes and groups such as Wine data, they tend to achieve inconspicuous clustering results. Moreover, another weakness of the proposed methods is the limitation of analyzing linear inseparable data.

Concerning the above two situations, more extensions of methods in this paper should be encouraged to overcome these problems. Therefore, following future work directions will be included: (i) The adaption of proposed methods for linear inseparable data, for example, introduce kernel function to proposed methods. (ii) Improvement of proposed methods in dealing with high-dimensional data. (iii) Extend the proposed method to tackle outliner problems so that utilized in fields such as image segmentation.

ACKNOWLEDGMENT

This study was supported by JSPS KAKENHI Grant Number JP17K00332.

REFERENCES

[1] Shi Yuhui, and Russell C. Eberhart, Empirical study of particle swarm optimization, *IEEE Proceedings of the 1999 Congress on Evolutionary Computation-CEC99(Cat. No. 99TH8406)*, Vol. 3, pp. 1945-1950, 1999.

[2] Aarts E, Korst J., Simulated annealing and Boltzmann machines, 1988.

[3] Koza J. R., Genetic programming, 1997.

[4] Bezdek, James C., Robert Ehrlich, and William Full, FCM: The fuzzy c-means clustering algorithm, *Elsevier Computers & Geosciences*, Vol. 10, No. 2-3, pp. 191-203, 1984.

[5] Ruspini E. H., A new approach to clustering, *Information and control*, Vol. 15, No. 1, pp. 22-32, 1969.

[6] Eberhart, R. and Kennedy, J, Particle swarm optimization (PSO), *Proc. IEEE International Conference on Neural Networks*, pp. 1942-1984, 1995.

[7] Dorigo, M., and Di Caro, G, Ant colony optimization: a new meta-heuristic, *IEEE Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406)*, Vol. 2, pp. 1470-1477, 1999.

[8] Runkler, T. A., and Katz, C, Fuzzy clustering by particle swarm optimization, *2006 IEEE international conference on fuzzy systems*, pp. 601-608, 2006.

[9] Bobrowski L. and Bezdek J. C., C-Means clustering with the L_1 and L_∞ norms, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 21, pp. 545-554, 1991.

[10] Jajuga K., L_1-norm based fuzzy clustering, *Fuzzy Sets and Systems*, Vol. 39, No. 1, pp. 43-50, 1991.

[11] Miyamoto Sadaaki and Agusta Yudi, An efficient algorithm for l_1 fuzzy c-means and its termination, *Control and Cybernetic*, Vol. 24, pp. 421-436, 1995.

[12] Hathaway R. J. and Bezdek J. C., Generalized Fuzzy c-Means Clustering Strategies Using L_p Norm Distances, *IEEE transactions on Fuzzy Systems*, Vol. 8, No. 5, pp. 578-582, 2000

[13] Shi Yuhui and Eberhart Russell, A modified particle swarm optimizer, *1998 IEEE international conference on evolutionary computation proceedings, IEEE world congress on computational intelligence (Cat. No. 98TH8360)*, pp. 69-73, 1998.

[14] L. Cagnina and S. Esquivel and R. Gallard, Particle swarm optimization for sequencing problems: a case study, *Proceedings of the 2004 Congress on Evolutionary Computation*, Vol. 1, pp. 536-541, 2004.

[15] WEN Z, Li R, Fuzzy C-means clustering algorithm based on improved PSO. *Application Research of Computers*, Vol. 7, 2010.

[16] Mehdizadeh E., Sadi-Nezhad S. and Tavakkoli-Moghaddam R., Optimization of fuzzy clustering criteria by a hybrid PSO and fuzzy c-means clustering algorithm, *Iranian Journal of Fuzzy Systems*, pp. 1-14, 2008.

[17] Silva Filho, T. M., Pimentel, B. A., Souza, R. M., Hybrid methods for fuzzy clustering based on fuzzy c-means and improved particle swarm optimization, *Expert Systems with Applications*, pp.6315-6328, 2015.

[18] Clerc Maurice, Discrete particle swarm optimization, illustrated by the traveling salesman problem, *New optimization techniques in engineering*, pp.219-239, 2004.

[19] Liu Hao, Xu Gang and Ding Gui-yan and Sun Yu-bo, Human behavior-based particle swarm optimization, *The Scientific World Journal*, Vol. 2014, 2014.

[20] Janson Stefan and Middendorf Martin, A hierarchical particle swarm optimizer for dynamic optimization problems, *Workshops on Applications of Evolutionary Computation*, pp. 513-524, 2004.

[21] Selvakumar A. I. and Thanushkodi K., A new particle swarm optimization solution to nonconvex economic dispatch problems, *IEEE transactions on power systems*, Vol. 22, No. 1, pp. 42-51, 2007.

[22] Pang Wei, Wang Kang-ping, Zhou Chun-guang and Dong Long-jiang, Fuzzy discrete particle swarm optimization for solving traveling salesman problem, *IEEE The Fourth International Conference onComputer and Information Technology*, pp. 796-800, 2004.

[23] Flood Merrill M., The traveling-salesman problem, *Operations research*, Vol.4, No. 1, pp. 61-75, 1956.

[24] Izakian Hesam and Abraham Ajith, Fuzzy C-means and fuzzy swarm for fuzzy clustering problem, *Expert Systems with Applications*, Vol. 38, No. 3, pp. 1835-1838, 2011.

[25] Izakian Hesam, Abraham Ajith and Snasel Vaclav, Fuzzy clustering using hybrid fuzzy c-means and fuzzy particle swarm optimization, *IEEE 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, pp. 1690-1694, 2009.

[26] Collins Linda M. and Dent Clyde W., Omega: A general formulation of the rand index of cluster recovery suitable for non-disjoint solutions, *Multivariate Behavioral Research*, Vol. 23, No. 2, pp. 231-242, 1988.

[27] Xie, Xuanli Lisa and Gerardo Beni, A validity measure for fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, pp. 841-847, 1991.

[28] Rand William M., Objective criteria for the evaluation of clustering methods, *Journal of the American Statistical association*, Vol. 66, No. 336, pp. 846-850, 1971.

[29] R. A. Fisher., The Use of Multiple Measurements in Taxonomic Problems, *Annals of Eugenics*, Vol. 1.7, No. 2, pp. 179-188, 1936.

[30] Forina, M. Wine DataSet, https://archive.ics.uci.edu/ml/datasets/wine

[31] Dr. William H. Wolberg, W. Nick Street and Olvi L. Mangasarian. Breast Cancer Wisconsin (Diagnostic) DataSet, http://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)