

C. Ratanavilisagul, "Modification fitness function of particle swarm optimization to improve the cluster centroid," in Proc. IOP Conf. Ser.: Mater. Sci. Eng., vol. 965, 2020, Art. no. 012010, pp. 1–10, doi: 10.1088/1757-899X/965/1/012010.

Modification fitness function of particle swarm optimization to improve the cluster centroid

Chiabwoot Ratanavilisagul

Department of Computer and Information Science of Applied Science, King
Mongkut's University of Technology North Bangkok (KMUTNB), Bangkok, Thailand

chaibwoot@hotmail.com, chaibwoot@gmail.com, chaibwoot.r@sci.kmutnb.ac.th

Abstract. Clustering is a classification technique. The clustering is brought to create models. The model is applied with the new input data to identify the class of that data. The goal of clustering is finding the best cluster centroid. The process of searching the best cluster centroid can use Particle Swarm Optimization. Many researches use it and obtain good results. But, searching the best cluster centroid by PSO suffers the problem from calculating the fitness function, because the fitness function is calculated by distances from the cluster centroid to all data inputs. These fitness functions may lead to the bad cluster centroid. So, this paper proposed the fitness function is calculated by the results from classification. The results from classification can lead to the good cluster centroid and create a good model to apply with the new input data. The proposed technique is tested on seven datasets from the UCI Machine Learning Repository and gives more satisfied search results in comparison with PSOs for the data clustering problems.

Keywords: Data Clustering; Particle Swarm Optimization; Centroid Vectors; Fitness Function

1. Introduction

The clustering [1], [2] is optimization problem and classification technique. It is applied with many application areas, such as bioinformatics, machine learning, information retrieval, data analysis, pattern recognition and image processing [3], [4]. It is brought to create models. The model is applied to the new input data to identify the class of that data. So, the goal of clustering is creating a model to identify the class of new objects. The clustering technique composes as follows: The object (data input) comprises many attributes. The attribute can be used to identify the class of the object. The objects that have similarity attributes are in the same class. On the other hand, the objects that have differential attributes are in other classes. Particle Swarm Optimization (PSO) is inspired from the behavior of flying birds and their communication mechanism [5], [6]. PSO is a stochastic global optimization technique. PSO is used for solving optimization problems. The goal of PSO is finding the global optimum of a fitness function or an objective function. PSO has many pros such as rapid convergence, simplicity, and little parameters to be adjusted [7].

In comparison with several other population-based stochastic optimization methods, such as the genetic algorithm (GA) [8, 9] and the evolutionary programming (EP), PSO performs better in solving various optimization problems with fast and stable convergence rates [10 - 12].

C. Ratanavilisagul, "Modification fitness function of particle swarm optimization to improve the cluster centroid," in Proc. IOP Conf. Ser.: Mater. Sci. Eng., vol. 965, 2020, Art. no. 012010, pp. 1–10, doi: 10.1088/1757-899X/965/1/012010.

The well-known technique for applying to clustering is the K-Means clustering (KM) [13]. The cons of KM are the problem of initializing the cluster centers and the problem of trapping in local optimum [3]. To overcome this cons, many researchers used PSO apply with KM. Applying PSO and KM is called that PSO-clustering [14]. PSO-clustering can increase distribution of population. Hence, it can decrease the problem of initializing the cluster centers and the problem of trapping in local optimum. The experiment results of these papers showed that these techniques can get better solutions than KM alone. Technique of PSO-clustering can be concluded as follows: PSO is used to find the cluster centroid and KM is used to calculate the fitness function.

Many researchers tried to improve the searching performance of PSO-clustering. Some researchers improved KM such as X. D. Liu [15] proposed a new clustering criterion function which combines intra-cluster distance and inter-cluster distance. C. C. Hung and H. Pumawan [16] proposed a hybrid rough K-means PSO algorithm. Fengqin Yang [17] proposed the K-harmonic means (KHM) added into the process of PSO. Some researchers improved PSO such as J. Sun and W. B. Xu combined Quantum-behaved Particle Swarm Optimization (QPSO) with KM for solving data cluster [18]. J. Y. Chen and H. Y. Zhang proposed a new K-means cluster algorithm based on improved velocity of PSO cluster algorithm (RVPSO-K) [19]. These techniques can improve the searching performance of PSO-clustering and get better answers.

The fitness function is guidance for searching of PSO. So, the main performance of applying PSO-clustering is the fitness function. It leads to classify classes. If the fitness function is defined to be a bad function, the results from searching get bad solutions. Hence, many researchers proposed the fitness function for PSO-clustering [14].

DW van der Merwe and AP Engelbrecht [14] proposed the fitness function uses the quantization error equation. It is the first paper which proposed PSO Clustering. So, it is the standard of PSO with clustering. This technique is called that particle swarm optimization with Clustering or PSO-clustering or PSOC. The results show that the solution of PSOC is better than that of KM. The source code of PSOC is proposed by Augusto Luis Ballardini [20].

Ching-Yi Chen and Fun Ye [21] proposed the fitness function uses the hard c-means which is created with the Euclidean norm. This technique is called PSO-clustering with the hard c-means or PSOCC. The result shows that the solution of PSOCC is better than that of KM and that of Fuzzy C-means.

Pei Zhenkui, Hua Xia, and Han JinFeng [22] proposed the fitness function uses the ratio of the maximum value from K-means to the minimum value from K-means. This technique is called the clustering algorithm based on PSO or CPSO. The results show that the solution of CPSO is better than that of KM.

R. J. Kuo, M. J. Wang, and T. W. Huang [23] proposed the fitness function uses K-means. In my paper, this technique is called that particle swarm K-means optimization algorithm or PSKO. The results show that the solution of PSKO is better than that of KM and that of PSOC.

PSOC, PSOCC, CPSO, and PSKO use the concept of KM which is calculated from distances between the input data and the cluster centroid. The main concept of KM, the cluster centroid vector has the least distances from all data inputs. That point obtains the best fitness value. However, that point may classify classes badly. The good fitness value may classify worse than the bad fitness value.

The goal of using PSO is finding the best point which can classify the most amounts of input data correctly. So, the fitness function which is calculated from distance is unsuitable. If the fitness function is calculated from the results from classification, it will get the point which can classify many amounts of input data correctly.

F-Measure is the results from classification [3]. If a point can classify many data correctly, it will get much value of F-measure. So, F-Measure can be used to guide searching of PSO. It can lead to the best classificatory point and it gets good model to apply with the new input data.

C. Ratanavilisagul, "Modification fitness function of particle swarm optimization to improve the cluster centroid," in Proc. IOP Conf. Ser.: Mater. Sci. Eng., vol. 965, 2020, Art. no. 012010, pp. 1–10, doi: 10.1088/1757-899X/965/1/012010.

From previous mention, this paper proposed the fitness function is calculated from the result from classification or F-measure. The proposed technique is called that modification fitness function by F-measure for particle swarm optimization with clustering or FPSOC. A set of dataset in the UCI Machine Learning Repository [24] is used to compare PSOC [14] by source code from [20], PSOCC [21], CPSO [22], PSKO [23] and the proposed technique. The results show that the solutions quality of the proposed technique is better than other compared algorithms in all tested datasets.

The rest of this paper is organized as follows: Section 2 explains the basic theoretical for this paper such as the K-Means Clustering, the standard PSO and F-Measure. Section3 explains related works (PSOC, PSOCC, CPSO, and PSKO). Section 4 explains the proposed technique or modification fitness function by F-measure for particles swarm optimization with clustering (FPSOC). Section 5 explains the case studies of datasets, the experiment setup and presents the experiment results. Section 6 concludes the paper with a brief summary.

2. Background

2.1. Particle Swarm Optimization

In standard PSO, each member of the population is called a "particle". Each particle contains its position and its velocity. Each individual particle executes searching in the search space according to its velocity, the best position found in the whole swarm (GBEST) and the individual's best position (PBEST). PSO starts with randomizing particle positions and their velocities, and the evaluation of the position of each particle is calculated from the objective function of the optimization problem. In each generation, each individual particle updates its position and velocity according to the expression below:

$$V_{id}' = \omega V_{id} + \eta_1 rand() (P_{id} - X_{id}) + \eta_2 rand() (P_{gd} - X_{id}) \quad (1)$$

$$X_{id}' = X_{id} + V_{id}' \quad (2)$$

Where X_{id} is the previous positions of i particle and d dimension, X_{id}' is the current positions of i particle and d dimension, V_{id} is the previous velocity of i particle and d dimension, V_{id}' is the current velocity of i particle and d dimension, P_{id} is PBEST of i particle and d dimension, and G_{id} is GBEST of d dimension. ω is an inertia weight, η_1 and η_2 are acceleration constants, and $rand()$ is the random number generated uniformly in the range $[0, 1]$. A limit velocity calls V_{max} . If calculate velocity of a particle exceeds V_{max} , it is replaced value of V_{max} .

2.2. K-Means Clustering

In standard KM, each data is called an object. The group of objects is called a class or the cluster centroid. Both class and object contain attributes. The attributes of an object and the attributes of classes are used to classify by Euclidean distance technique [14]. The main concept of KM is can be summarized as follows: the smallest Euclidean distances from the attributes of an object to the attributes of one class, this object is assigned into this class. The standard KM starts with randomizing attributes of classes and this step is applied only the first time. Then, the evaluation of each class is calculated from the expression below:

$$d(z_p, m_j) = \sqrt{\sum_{k=1}^{N_d} (z_{pk} - m_{jk})^2} \quad (3)$$

C. Ratanavilisagul, “Modification fitness function of particle swarm optimization to improve the cluster centroid,” in Proc. IOP Conf. Ser.: Mater. Sci. Eng., vol. 965, 2020, Art. no. 012010, pp. 1–10, doi: 10.1088/1757-899X/965/1/012010.

Where N_d is the number of attributes, Z_p is the object p , Z_{pk} is the attribute k in the object p , m_j is the class j , m_{jk} is the attribute k in the class j , and $d(z_p, m_j)$ is the distance between the object p and the class j . Then, each class is recalculated from the expression below:

$$m_j = \frac{1}{n_j} \sum_{z_p \in C_j} z_p \quad (4)$$

Where n_j is the number of object in class j , and C_j is all objects in the class j . This algorithm can be stopped when the maximum number of iterations has been exceeded.

2.3. F-Measure

F-Measure (FM) or $F(i, j)$ is a famous evaluation measure. FM is the harmonic mean which combines the recall or $R(i, j)$ and the precision or $P(i, j)$. FM uses the ideas of precision and recall from information retrieval [3]. The precision is the most important measure of any classification system. It provides information on the probability that a prediction of a given category is correct. The recall is the relationship between the number of correctly detected intrusions and the total number of intrusions. True Positive (T_p) is the model correctly predicts the positive class. True Negative (T_n) is the model correctly predicts the negative class. False Positive (F_p) is the model incorrectly predicts the positive class. False Negative (F_n) is the model incorrectly predicts the negative class. FM, the precision and the recall for each class i of each cluster j is calculated as:

$$P(i, j) = \frac{T_p}{T_p + F_p} \quad (5)$$

$$R(i, j) = \frac{T_p}{T_p + F_n} \quad (6)$$

$$F(i, j) = \frac{2 \times P(i, j) \times R(i, j)}{P(i, j) + R(i, j)} \quad (7)$$

3. Related Works

3.1. PSO with Clustering

The standard PSO was proposed to solve the benchmark function problem [25]. To PSO can apply with the data clustering problem, the objective function and particle position are modified. A single particle represents the class. The fitness function of the particle is measured as (8). In addition to the previous mention, the process of PSO with data clustering [14] is similar to the process of the standard PSO.

$$Fitness = \frac{\sum_{j=1}^{N_c} [\sum_{z \in C_{ij}} d(z_p, m_j) / |C_{i,j}|]}{N_c} \quad (8)$$

Where N_c is the number of classes, and $|C_{i,j}|$ is the number of object belonging to class j

C. Ratanavilisagul, "Modification fitness function of particle swarm optimization to improve the cluster centroid," in Proc. IOP Conf. Ser.: Mater. Sci. Eng., vol. 965, 2020, Art. no. 012010, pp. 1–10, doi: 10.1088/1757-899X/965/1/012010.

3.2. PSO-clustering with the hard c-means

Process of PSO-clustering with the hard c-means (PSOCC) [21] is the same as PSO-clustering except the fitness function is used as follows:

$$J = \sum_{j=1}^{N_c} \sum_{p=1}^N \|Z_p - m_j\| \quad (9)$$

$$Fitness = \frac{k_c}{(J + J_0)} \quad (10)$$

Where N is the number of object, k_c is a positive constant, and J_0 is a small-valued constant. The experimental result shows that PSOCC algorithm has better performance than KM algorithm.

$\|Z_p - m_j\|$ is the Euclidean distance of Z_p to m_j .

3.3. The clustering algorithm based on PSO

The clustering algorithm based on PSO (CPSO) [22] is the same as PSO-clustering except the fitness function is used as follows:

$$Fitness = \frac{\max(\bar{d}_1(Y_i))}{\min(d_2(Y_i))} \quad (11)$$

$$\max(\bar{d}_1(Y_i)) = \max_{a=1, \dots, N_c} \left(\sum_{\forall Z \in C_{ij}} \frac{d(z_p, m_j)}{|C_{ij}|} \right) \quad (12)$$

$$\min(d_2(Y_i)) = \min_{a=1, \dots, N_c, b=1, \dots, N_c, i \neq j} (d(m_i, m_j)) \quad (13)$$

Where $\max(\bar{d}_1(Y_i))$ is the maximum value of average values of distances within same classes in the classification plan which is showed by particle Y_i , $\min(d_2(Y_i))$ is the minimum value of distances between classes in the classification plan which is showed by particle Y_i , m_j is the class j, and m_i is the class i.

3.4. particle swarm K-means optimization

The particle swarm K-means optimization algorithm (PSKO) [23] is the same as PSO-clustering except the fitness function is calculated according to (3). Moreover, the process of recalculating the cluster centroid is added for each particle according to (4). From experimental results, PSKO which integrates both the PSO-clustering and K-means algorithm is able to give the particles both exploration and exploitation capabilities. The experimental results showed that performance of PSKO is better than those of other clustering algorithms.

4. Proposal Works

Duty of PSO is searching the cluster centroid. The process of PSO tries to get better fitness. So, the fitness function is guidance for searching of PSO and the main factor of searching the cluster centroid. For the data clustering problem, if the cluster centroid stays in a suitable area, it can classify data correctly. On the other hand, if the cluster centroid stays in an unsuitable area, it will rarely classify data correctly. These techniques (PSOC, PSOCC, CPSO, and PSKO) use the fitness function which is calculated by the distance from all input data to the cluster centroid. The effect from fitness function, PSO tries to decrease the lowest distance from all input data to the cluster centroid. So, the more fitness value is decreased, the better solution is obtained. It leads to the cluster centroid move to the space of two data groups and the cluster centroids stick together. The sticky cluster centroids are unsuitable to classify data because data of one group may be classified into another group as figure 1 (b). In fact, the cluster centroids should stay central in its group which it can classify data correctly as figure 1 (a).

C. Ratanavilisagul, "Modification fitness function of particle swarm optimization to improve the cluster centroid," in Proc. IOP Conf. Ser.: Mater. Sci. Eng., vol. 965, 2020, Art. no. 012010, pp. 1–10, doi: 10.1088/1757-899X/965/1/012010.

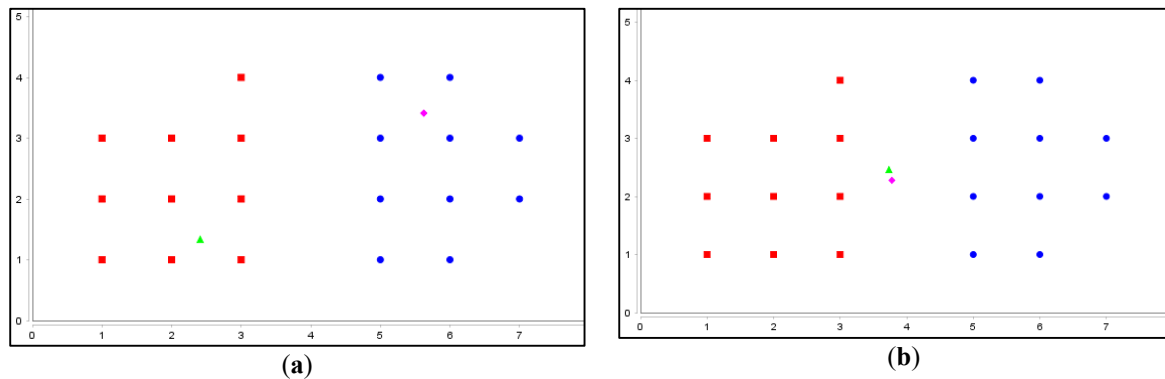


Figure 1. This figure shows the results from searching of PSOC in two dimensions. The red point or the square point is the data of the first group which has ten points. The blue point or the circle point is the data of the second group which has ten points. The green point or the triangle point is the cluster centroid of the first group. The violet point or the diamond point is the cluster centroid of the second group.

From figure 1 (a), the result from calculation by fitness function as (8) gets value that is 1.303. The cluster centroid from this figure can classify data that give correct 100% or 20 points. The value of F-measure in this case is 1. So, this model classifies correctly and has performance and is suitable to be used. From figure 1 (b), the result from calculation by fitness function as (8) gets value that is 1.074. The cluster centroid from this figure can classify data that give correct 65% or 13 points. The value of F-measure in this case is 0.601. Some data in the first group is identified into the second group because they stay close to the cluster centroid of the second group more than that of the first group. So, this model classifies incorrectly and has no performance.

The result from fitness function as (8), PSO selects figure 1 (b) to set GBEST because fitness of figure 1 (b) is less than fitness of figure 1 (a). In fact, PSO should select figure 1 (a) to set GBEST because it is more classificatory than another. This case shows that distance may be unsuitable in some events to calculate fitness of PSO. On the other hand, the fitness function should use the results from classification. In this case, if the fitness function uses the results from classification or F-measure, it selects figure 1 (a) to set GBEST because the amount of correct data from figure 1 (a) is more than that from figure 1 (b). The F-measure is calculated from the results from classification. So, F-measure can be used for guiding searching of PSO or the fitness function. Hence, this paper proposes the fitness function of PSO uses F-measure to solve the data clustering problem or modification fitness function by F-measure for particles swarm optimization with clustering (FPSOC). Pseudo code of FPSOC is shown below:

C. Ratanavilisagul, "Modification fitness function of particle swarm optimization to improve the cluster centroid," in Proc. IOP Conf. Ser.: Mater. Sci. Eng., vol. 965, 2020, Art. no. 012010, pp. 1–10, doi: 10.1088/1757-899X/965/1/012010.

```

Initial position and velocity of each particle
Initial PBEST, GBEST
For gen = 0 to MAXGEN
    For i = 0 to N
        Evaluate particle fitness according to formula (7)
        If  $F(X_i) < F(PBEST_i)$ 
             $PBEST = X_i$ 
        End If
        If  $F(X_i) < F(GBEST)$ 
             $GBEST = X_i$ 
        End If
        Update particle position according to formula (1) and (2)
    End For
Next gen

```

Where N is the number of particles, $F(X_i)$ is fitness of i particle, $F(PBEST_i)$ is fitness from PBEST of i particle, $PBEST_i$ is PBEST of i particle, $F(GBEST)$ is fitness of GBEST, MAXGEN is the maximum generation.

5. Experiments and Results

5.1. Datasets

The proposed algorithm is tested on seven datasets [24], listed in Table 1. The results of FPSOC on the datasets are compared with PSOC, PSOCC, CPSO, and PSKO.

Table 1. Details of datasets

Order	Name	Datasets	Attributes	Classes
1	IRIS	150	4	3
2	WINE	178	13	3
3	GLASS	214	9	6
4	HEART	270	12	2
5	CANCER	286	8	2
6	ECOLI	336	7	8
7	YEAST	1484	8	10

5.2. The measures of algorithm performance

The measures of algorithm performance in the experiments are as follows: F-Measure (FM), the average correct number (ACN) is the average of the correct number from the prediction of algorithm in the final generation from all running. FM and ACN indicate the clustering efficiency of an algorithm. The bigger FM and ACN are, the higher the quality of clustering is. Therefore, PSO is the stochastic algorithm. It has the process of randomness. So, FM and ACN may be enough to measure performance of PSO. To measure performance of PSO, this paper adds the standard deviation of ACN (SD). SD indicates the clustering reliability of an algorithm. The bigger SD is, the higher the reliability of clustering is.

C. Ratanavilisagul, “Modification fitness function of particle swarm optimization to improve the cluster centroid,” in Proc. IOP Conf. Ser.: Mater. Sci. Eng., vol. 965, 2020, Art. no. 012010, pp. 1–10, doi: 10.1088/1757-899X/965/1/012010.

5.3. Parameters Setting

Parameters for all experiments are as follows: acceleration constants of η_1 and η_2 are both set to be 1.45, $\omega = 0.75$, the number of population is 100. The number of experiments of each function is 25 runs. The maximum iteration is 1000. The non-PSO parameters for the compared algorithm parameters (PSOCC, CPSO, PSKO and FPSOC) are set according to suggested by the original papers. This research is conducted by a personal computer of AMD FX-8320 with 3.5 GHz CPU, 8 GB RAM and Visual C++ 2010 as the programming language. All datasets are used in experiment are from datasets [24].

5.4. Experiment of the proposed algorithm

Table 2. Comparative F-Measure (FM) of PSOC, PSOCC, CPSO, PSKO and FPSOC on datasets

Datasets	Algorithms				
	PSOC	PSOCC	CPSO	PSKO	FPSOC
IRIS	0.75	0.70	0.72	0.70	0.99
WINE	0.57	0.60	0.60	0.61	0.77
GLASS	0.32	0.33	0.32	0.34	0.54
HEART	0.44	0.46	0.45	0.45	0.73
CANCER	0.53	0.51	0.50	0.51	0.70
ECOLI	0.32	0.28	0.30	0.32	0.60
YEAST	0.29	0.32	0.31	0.23	0.39

Table 3. Comparative the average correct number (ACN) of PSOC, PSOCC, CPSO, PSKO and FPSOC on datasets

Datasets	Algorithms				
	PSOC	PSOCC	CPSO	PSKO	FPSOC
IRIS	117.00	110.47	114.84	111.88	148.48
WINE	115.36	117.44	115.24	117.96	137.68
GLASS	106.44	107.76	109.40	109.52	146.52
HEART	152.92	154.60	153.64	152.92	199.00
CANCER	187.56	188.12	188.80	183.32	215.68
ECOLI	178.04	166.04	169.76	174.00	238.28
YEAST	315.04	411.20	404.00	291.20	779.76

Table 4. Comparative the standard deviation (SD) of PSOC, PSOCC, CPSO, PSKO and FPSOC on datasets

Datasets	Algorithms				
	PSOC	PSOCC	CPSO	PSKO	FPSOC
IRIS	15.44	17.55	15.30	7.20	0.51
WINE	8.65	10.13	10.00	9.83	3.73
GLASS	4.03	4.19	6.66	5.39	2.42
HEART	6.67	8.73	7.35	6.47	0.00
CANCER	15.20	10.10	16.54	20.31	9.10
ECOLI	0.09	0.07	0.08	19.39	0.06
YEAST	71.70	80.35	81.94	73.47	66.94

C. Ratanavilisagul, "Modification fitness function of particle swarm optimization to improve the cluster centroid," in Proc. IOP Conf. Ser.: Mater. Sci. Eng., vol. 965, 2020, Art. no. 012010, pp. 1–10, doi: 10.1088/1757-899X/965/1/012010.

From the experimental results of Table 2 and Table 3 show that the quality solution of FPSOC is better than that of PSOC, PSOCC, CPSO, and PSKO because of its highest FM and ACN all datasets. From the experimental results of Table 4 show that the reliability of FPSOC is better than that of PSOC, PSOCC, CPSO, and PSKO because of its lowest SD all datasets. Hence, FPSOC outperforms PSOC, PSOCC, CPSO, and PSKO in case of reliability and quality solution. Hence, FPSOC outperforms PSOC, PSOCC, CPSO, and PSKO in case of quality solution. The experiments can prove that the fitness function affects searching of PSO and searching the centroid vector. PSOC, PSOCC, CPSO, and PSKO use the fitness function that is calculated from the distance. These techniques are calculated by the same concept so they get a few different solutions both the reliability and quality of solutions in all datasets. The solutions from searching of FPSOC are much better than solutions from searching of PSOC, PSOCC, CPSO, and PSKO. Hence, F-Measure can use the fitness function of PSO (FPSOC) to search the cluster centroid and it has performance better than the fitness function that is calculated by the distance (PSOC, PSOCC, CPSO, and PSKO). In the case of HEART dataset, SD of FPSOC equals 0 although the results of FPSOC are better than the results of other techniques. It shows that FPSOC encounters the problem of trapping in local optimum. In the future this problem should be solved to improve the searching performance of PSO.

6. Conclusions

The clustering is brought to create models. The model is applied to new input data to identify the class of that data. The clustering depends on the cluster centroid. PSO is brought to search the cluster centroid. The fitness function is the determinant of direction searching of PSO and the main factor of searching the cluster centroid. Many papers proposed that the fitness function is calculated by the distance from all input data to the cluster centroid. However, these methods may lead to bad solutions. To improve performances of PSO for searching the cluster centroid, this paper proposed the fitness function is calculated by the results from classification or F-Measure. The proposed technique is called that FPSOC. From the experimental results show that the proposed FPSOC outperforms PSOC, PSOCC, CPSO, and PSKO with regard to the reliability and quality of solutions in all datasets.

References

- [1] P.N. Tan, M.Steinbach, and V.Kumar. "Introduction to data mining," Boston: AddisonWesley. pp.487–559, 2005.
- [2] Jiawei han and Michelin Kamber. "Data mining concepts and techniques," Elsevier. 2010.
- [3] F. Yang, T. Sun, C. Zhang, "An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization," Expert Syst., pp.9847–9852, 2009.
- [4] C.D. Manning, P. Raghavan, H. Schütze, "Introduction to Information Retrieval," Cambridge University Press, Cambridge, Vol. 1, 2008.
- [5] J. Kennedy and R. Eberhart, Particle swarm optimization, "Proceeding of IEEE International Conference on Neural Networks (ICNN'95)," Perth, Western Australia, IEEE, vol. 4, pp.1942–1947, 1995.
- [6] J. Kennedy and R. Eberhart, "A new optimizer using particle swarm theory," Sixth International Symposium on Micro Machine and Human Science, pp.39-43, 1995.
- [7] Qinghai, B., "Analysis of Particle Swarm Optimization Algorithm," Computer and Information Science, pp.180–184, 2002.
- [8] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan press, Ann Arbor, pp.211, 1979.
- [9] D. E. Goldberg, *Genetic Algorithm in Search, Optimization and Machine Learning*, Addison-Wesley, pp.372, 1989.
- [10] K. E Parsopoulos, V. P. Plagianakos, G. D. Magoulas, M. N. Vrahatis, *Objective function stretching to alleviate convergence to local minima*, Nonlinear Analysis TMA 47, pp.3419-3424, 2003.

C. Ratanavilisagul, "Modification fitness function of particle swarm optimization to improve the cluster centroid," in Proc. IOP Conf. Ser.: Mater. Sci. Eng., vol. 965, 2020, Art. no. 012010, pp. 1–10, doi: 10.1088/1757-899X/965/1/012010.

- [11] R. Eberhart and Y. Shi, *Comparison between genetic algorithms and particle swarm optimization*, Proc. Seventh Annual Conference on Evolutionary Programming, pp.611-619, 1998.
- [12] J. Kennedy and R. Eberhart, *Swarm Intelligence*, Morgan Kaufmann, San Mateo, 2001.
- [13] J. Nayak, B. Naik, H.S. Behera, "Fuzzy C-Means (FCM) clustering algorithm," a decade review from 2000 to 2014. In: Jain, L.C. et al. (eds.) Comput. Intell. Data Min. vol. 2, Smart Innov. Syst. Technol. 32, vol. 2, pp.133–149, 2014.
- [14] D.W.Van der Merwe. and A.P. Engelbrecht, "Data clustering using particle swarm optimization," Proceedings of IEEE Congress on Evolutionary Computation, pp.215-220, 2003.
- [15] X.D. Liu,Q.F. Sha andYK Liu et.al, "Analysis of classification using particle swarm optimization," Computer Engineering, vol.32 (6), pp.201-202+217,2006.
- [16] J. Komorowski,Z.Pawlak and L.Polowski,et al.,Rough sets: a tutorial, a new trend in decision making,Springer,Singapore Pte Ltd,1998.
- [17] F. Q. Yang, T. E. L. Sun and C. H. Zhang, "An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization," Expert Systems with Applications, vol.36, pp.9847- 9852, 2009.
- [18] J. Sun, W. B. Xu and B. Ye, "Quantum-behaved particle swarm optimization clustering algorithm," Lecture Notes in Computer Science, Vol.4093, pp. 340-347, 2006.
- [19] J. Y. Chen and H. Y. Zhang, "Research on application of clustering algorithm based on PSO for the web usage patter," International Conference on Wireless Communications, Networking and Mobile Computing, pp.3705-3708, 2007.
- [20] Augusto Luis Ballardini, "A tutorial on Particle Swarm Optimization Clustering," arXiv:1809.01942, pp.1–13, 2016.
- [21] Ching-Yi Chen, and FunYe, "Particle Swarm Optimization Algorithm and Its Application to Clustering Analysis", Proceedings of the 2004 IEEE International Conference on Networking, Sensing & Control Taipei. Taiwan, pp. 789 - 794 ,March 21-23, 2004.
- [22] Pei Zhenkui, Hua Xia, and Han Jinfeng, "The Clustering Algorithm Based on Particle Swarm Optimization Algorithm," 2008 International Conference on Intelligent Computation Technology and Automation, pp. 148- 151, 2008
- [23] R. J. Kuo, M. J. Wang, and T. W. Huang, "An application of particle swarm optimization algorithm to clustering analysis, " Soft Comput, vol. 15, pp. 533–542, 2009.
- [24] Department of Information and Computer Science. University of California. "UCI Machine Learning Repository,"[Online].Available: <http://www.ics.uci.edu/mllearn/MLRepository.html>.1998.
- [25] Molga, M., & Smutnicki, C. Test functions for optimization needs (2005). Retrieved June 2013.