

# EXTENDED PSO ALGORITHM FOR IMPROVEMENT PROBLEMS K-MEANS CLUSTERING ALGORITHM

Maryam Lashkari<sup>1</sup> and Amin Rostami<sup>2</sup>

<sup>1</sup>Department of Computer Engineering, Ferdows Branch, Islamic Azad University,  
Ferdows, Iran.

<sup>2</sup>Department of Computer Engineering, Ferdows Branch, Islamic Azad University,  
Ferdows, Iran.

## ABSTRACT

*The clustering is a without monitoring process and one of the most common data mining techniques. The purpose of clustering is grouping similar data together in a group, so were most similar to each other in a cluster and the difference with most other instances in the cluster are. In this paper we focus on clustering partition k-means, due to ease of implementation and high-speed performance of large data sets, After 30 year it is still very popular among the developed clustering algorithm and then for improvement problem of placing of k-means algorithm in local optimal, we pose extended PSO algorithm, that its name is ECPSO. Our new algorithm is able to be cause of exit from local optimal and with high percent produce the problem's optimal answer. The probe of results show that mooted algorithm have better performance regards as other clustering algorithms specially in two index, the carefulness of clustering and the quality of clustering.*

## KEYWORDS

*Clustering, Data Mining, Extended chaotic particle swarm optimization, K-means algorithm.*

## 1. INTRODUCTION

Nowadays, usage of data mining observe in most of science, visibly. It's obvious that if don't prepared suitable bedfast for use of this science, we will be away from achieved progress. Clustering is one of the most common data mining tools. That use in most case such as: engineering, data mining, medical science, social science and other items. As for clustering very applications, need to clustering and data mining is necessary in most field for further progress. First time idea of clustering represent in 1935 year and nowadays because of progresses and huge mutation most of researchers pay attention to clustering. The clustering is process of collection grouping form without label's data.

That inner members have most similar to each other in a cluster and least similar to regard as other cluster's members. So, clustering is more ideal when two inner cluster likeness factor is maximum and outside cluster likeness is least. There is other criterions, Such as: Euclidean distance, hamming, for determination level of sample's likeness to each other. That every criterion have further usage in special field.

Purpose's function is convex and non-linear in most clustering problems [1]. It's possible that algorithm place in trap of local optimal and produce the problem's optimal answer. There are several clustering algorithm that grouping to following kinds. Hierarchical clustering algorithm, partition, density based, model and graph based that each of them are more effective regard as other algorithm in special data environment. In all of this algorithms, researchers try to balanced, control or improve parameters to be more effective algorithm that consist of:

-high measurement,- having ability to work with high dimension – having ability to dynamic data clustering, - having ability to work with high distance of problem,- having least need to additional knowledge about problem,- suitable management from noises, and interpretable clusters.

Partition clustering algorithm is one of the most common and most applied from clustering algorithm. That specific data collection to specified partition's number. So that samples in every partition have most similar to each other in a cluster and most difference with samples in other clusters .K-means algorithm of is the famous clustering algorithm in this field [6]. And it's one of the favorite center-pivot clustering algorithms in clustering technique. K-means start with Initialization to cluster's centers and other things with regard to Euclidean distance criterion allocate to one cluster that have least distance to cluster's centers. In every algorithm repetition, perform two chief phase. First, every item in data collection allocate to a cluster that have least distance from cluster's centers. In continue, after that spots grouping to K cluster the new cluster's centers calculate by estimate average from samples of every cluster. And algorithm repeat. The temporal algorithm finish to there is any change in calculation of cluster's centers and or finish the repetition special number [9]. In this algorithm purpose function error square series that goal is to reach a minimum it, that show in equation (1). X indicate to cluster's samples and C indicate to cluster's center.

$$\phi = \sum_{x \in X} \min ||x - c||^2 \quad (1)$$

### **Advantages of K-means algorithm:**

Ease of implementation and high-speed performance, measurable and efficient in large data collection.

### **Disadvantages and problems of k-means algorithm:**

- 1- Selection of the first cluster's centers and number of cluster do by user. For this reason clustering results is dependent to first algorithm's selection and if first algorithm's condition don't be suitable, it's possible algorithm place in trap of local optimal.
- 2- Selection number of optimal cluster for problem is difficult.
- 3- This algorithm, because of calculation average from cluster samples for determination cluster's center, have weak management regard as noises and data.
- 4- This algorithm can't be usable in data collection that calculation average is not describable.
- 5- Data clustering is not usable with different forms and density.

In continue we will express techniques for improvement problems that usually this techniques focus on 3 issue that are:

- 1- Determination way of selection first parameters.
- 2- Alternation in basic algorithm.
- 3- Combination clustering algorithm with other initiative algorithms.

And then new solution posed for improvement problems of placing k-means algorithm's result in local optimal and validation evaluate by using 3 real data collection and some indexes and finally, we will have brief comparison from posed techniques. Continue of article organized in following from:

1. Related works
2. Analysis and comparison posed algorithms.
3. The proposed method.
4. Simulation.
5. Conclusion

## **2. RELATED WORKS**

### **K-medidos clustering algorithm:**

This algorithm[1,3] for resolving problem of noises weak management in k-means algorithm and also in perform in case that evaluation average for data collection is non describable. The idea that posed in this algorithm is contemplate most central sample as cluster's center in every cluster rather that selection data's average of one cluster.

Disadvantage: algorithm's temporal complexity is high and it isn't suitable and efficient for large data collection. Result of clustering is sensitive to first condition of algorithm and determination optimal K is difficult for problem.

### **CLARA clustering algorithm:**

For solving the problem of k-medidos algorithm in large data sets, that is, high temporal complexity posed CLARA algorithm [2]. In this algorithm solve the problem of k-medidos algorithm, temporal complexity in large data collection, but there is a problem. Suppose , n is total number of samples and m is most number of samples that this way of clustering can process in objective time . If  $n \gg m$ , often clustering from several small sample of data cause that eliminate some of the data in same groups.

### **K-modes clustering algorithm:**

For clustering nominal data, k-means algorithm isn't suitable for this reason, posed generalized way of k-means algorithm, k-modes. In this algorithm [3] rather that evaluation average, we use of mode every cluster as cluster's centroid and also algorithm first parameters such as k-means, selected randomly for this reason alongside advantage, be suitable for nominal data, it's possible that algorithm's result place in local optimal and also be suitable only for nominal data and it's not efficient for numerical data.

### **Particles Swarm Optimization Clustering algorithm:**

As we said before one of the problem of k-means algorithm is placing algorithm's result in trap of local optimal, because of algorithm local search in region of problem. Clustering algorithm based on PSO algorithm problem posed [4]. For elimination of this clustering algorithm based on pso have better operation rather than k-means algorithm with few dimension for data collection and there is more probability for get all over optimal answer rather than k-means algorithm because of all over research in region of problem but use of pso algorithm lead to much repetitions and slow convergence for data with high volume. For this reason, we often combine this 2 algorithm with each other to be complement and they cover weakness each other.

**Chaotic particle swarm optimization clustering algorithm:**

Two main problem of clustering using PSO method is the convergence to local optimal and slow convergence velocity, which is tried to be solved by using two ideas of chaos theory and acceleration strategy. In the formula of updating velocity of the cluster centers that is mentioned in the (2) updating is done for each particle for relocating the particle to the new position, from the best answer for each particle (Pbest) and the best global solution so far (gbest). In which W Inertia coefficient rate tends to previous velocity of the particle, c1 rates tends to the local best position of the particle, and c2 trends to the best global position of the particle [5].

In (3) replacing cr instead of rr improves PSO algorithm as given:

$$v_{id}^{new} = w \times v_{id}^{old} + c_1 \times r_r \times (Pbest_{id} - x_{id}^{old}) + c_2 \times r_2 \times (gbest_{id} - x_{id}^{old}) \quad (2)$$

$$v_{id}^{new} = w \times v_{id}^{old} + c_1 \times c_r \times (Pbest_{id} - x_{id}^{old}) + c_2 \times (1 - c_r) \times (gbest_{id} - x_{id}^{old}) \quad (3)$$

$$Cr_{t+1} = k \times Cr_{(t)} \times (1 - Cr_{(t)}) \quad (4)$$

In (4), Cr random value is created for each round independently between 0 and 1, which substitutes both r1 and r2, and parameter k is the number of predicted clusters. Using the chaos theory in PSO population generation will result in more diverse of the algorithm.

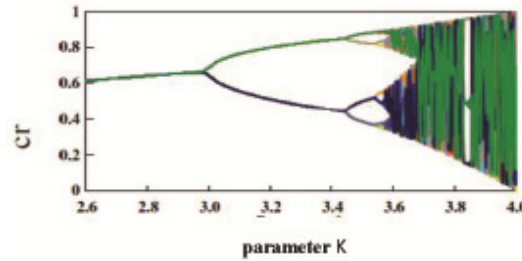


Figure1. Chaos map [5]

As can be seen in Figure 1. To achieve more optimal particle swarm optimization algorithm, chaos theory is applied. And in other change to increase the rate of convergence used acceleration strategy therefore in this mode a number of the population which are the best toward the target - move not all population that it increases the rate of convergence [5].

**Genetic clustering algorithm:**

In this algorithm [7] for exit from trap of local optimal in k-means algorithm we use of genetics optimization algorithm for better data clustering. Because of evolutionary algorithm, such as genetics, have ability for global search in answer, use of them for clustering, decrease probability from placing answer of algorithm in local optimal. And finally produce more optimal answer for clustering.

**Ant colony clustering algorithm:**

Ant colony clustering algorithm [9] is pivot population innovative algorithm that used for solving problem of optimization, such as: clustering. This algorithm is capable to produce optimal answer with high speed in clusters and with complex forms rather than other innovative algorithm. This

algorithm1- for better data clustering and reach to all over optimal answer with more probability of k-means algorithm use.2- Of ant colony algorithm for data clustering process.

### **K-mica compound clustering algorithm:**

This algorithm [11] is combination of colonial competition algorithm with k-means clustering algorithm. In this algorithm after production of primary population, randomly, k-means algorithm perform on available data with distinct numbers. Then obtained final cluster's centers consider as primary population of colonial competition algorithm that is imperialists and perform clustering on them based on extended colonial competition algorithm and allocate colony to suitable imperialists and clustering perform over data based on extended competition algorithm and allocate colonies to suitable colonialisms.

### **Four hybrid strategies for combination continuous ant colony optimization with PSO algorithm for utilizing in clustering process:**

This algorithm [12] posed 4 hybrid strategies for combination PSO algorithm. Their examinations show that utilizing hybrid strategies for clustering is so better than independent utilizing of k-means, PSO, ACOR algorithm for clustering process.

Four Hybrid strategies that used by them are:

- 1: Series combination of 2 algorithm PSO, ACOR
- 2: Parallel combination of 2 algorithm PSO, ACOR
- 3: Series combination of 2 algorithm with one extended chart from pheromone-Particles
- 4: Substitution global best between 2 algorithm.

## **3. ANALYSIS AND COMPARISON OF ALGORITHM**

As you see, we posed and checked several strategies and algorithms for elimination problems and challenges of k-means clustering algorithm each of discussed algorithms have advantages and disadvantages. Some of them expanded for elimination of previous algorithm limitation or they are new strategies for solve the problems of k-means algorithm. Challenges of k-means algorithm are:

- 1: Sensitivity to noise data
- 2: it's limited to numerical data
- 3: Result of algorithm is dependent to primary condition and placing algorithm in local optimal.
- 4: Lack of suitable clustering for clusters with different forms and density.

In continuance, table 1 show the comparison of described algorithm from the point of view of several important parameters. Empty cells of chart show that have any importance about specific algorithm from relevant parameter.

Algorithm	Advantages	Disadvantages	Temporal complexity	Suitable for data sets	Result of algorithm	Sensitivity to noise	Kind of algorithm search
K-medidos	Better managemen t of noises and pert data it's suitable for data sets which evaluation	High temporal complexity in large data sets. It is not suitable for clusters with different forms and density. Result of algorithm is	$O(k(n-k)^2)$	Numeric al	Most central members in each cluster	Have not	Local

	of average isn't describable in it.	related to first algorithm condition. And there is high probability for placing result of algorithm in local optimal. It's hard to determine optimal (k) for problem. Utility of this algorithm is lesser and it is implementation is more complex rather than k-means algorithm.					
CLARA	This algorithm can solve the problem of k-medoids algorithm, that is high temporal complexity in large data sets and also is suitable for massive data sets	Having weakness in operation of clustering. It is not suitable for clusters with different forms and density. The result of algorithm is related to first algorithm condition and there is high probability for placing result of algorithm in local optimal. It's hard to determine optimal(k) for problem	$O(k(40+k)^2 + k(n-k))$	Numerical	Most central members in each cluster	Have not	Local
K-modes	It's suitable for clustering of nominal data sets.	Having weakness in clustering of numerical data. It is not suitable for clusters with different forms and density. The result of algorithm is related to first algorithm condition and there is high probability for	$O(n)$	Nominal	Mood of each cluster	Have not	Local

		placing result of algorithm in local optimal. It's hard to determine optimal (k) for problem					
Clustering based on PSO algorithm	Probability for reach to all over optimal answer and exit of local optimal is more than k-means algorithm because of all over research in area of problem	Use of pso algorithm lead to much repetitions and slow convergence for data with high volume and it's suitable for data sets with low volume. First copy from pso algorithm is very related to problem parameters. And for this reason, algorithm place in local optimal.	-	Numerical	Centers of first clusters for k-means.	-	Global
chaotic particle swarm optimization clustering algorithm	There is increase of population variation and increase of convergence speed in pso clustering algorithm. And there is more probability for reach to all over optimal answer rather than pso clustering algorithm.	There is higher fiscal complexity than pso.	-	Numerical	Centers of first clusters for k-means algorithm or correction of formed clusters by k-means	-	Global
Clustering algorithm based on GA algorithm	Exit from trap of local optimal with high percent and there is probability for reach to	Low convergence speed and increase of fiscal complexity	-	Nominal	Centers of clusters	-	Global

	all over optimal answer for clustering.						
Clustering algorithm based on ant colony algorithm	It produce optimal answer with higher speed and complex forms in clusters rather than other algorithms. Better data clustering and reach to all over optimal answer with more probability rather than k-means algorithm	It's possible that algorithm place in trap of local optimal and produce optimal answer because of randomly things selection by ants and numbers of repetition	-	Numerical	Optimal centers of clusters	-	Global
Compound clustering algorithm (PSO+ACO+K-means)	There is improvement in problem of first condition selection for k-means algorithm. Increase of convergence speed to all over optimal answer and there is more probability for close to all over optimal answer rather than other evolutionary algorithm	High fiscal complexity	-	Numerical	Optimal centers of clusters	-	Global



## 4. PROPOSED METHOD

### 4.1. Introduction of standard PSO algorithm and its problem

Particle swarm optimization (PSO) is a population-based stochastic search process, modeled after the social behavior of a bird flock. The algorithm maintains a population of particles, where each particle represents a potential solution to an optimization problem. In the context of PSO, a swarm refers to a number of potential solutions to the optimization problem, where each potential solution is referred to as a particle. The aim of the PSO is to find the particle position that results in the best evaluation of a given fitness (objective) function. Each particle represents a position in  $N$ -dimensional space, and is "flown" through this multi-dimensional search space, adjusting its position toward both the particle's best position found thus far and the best position in the neighborhood of that particle. Each particle  $i$  maintains the following information:

$x_i$  : The current position of the particle;

$v_i$ : The current velocity of the particle;

$y_i$  : The personal best position of the particle.

Using the above notation. A particle's position is adjusted according to

$$v_{i,k}(t+1) = wv_{i,k}(t) + c_1r_{1,k}(t)(y_{i,k}(t) - x_{i,k}(t)) + c_2r_{2,k}(t)(y_k(t) - x_{i,k}(t))$$

$$X_i(t+1) = X_i + v_i(t+1)$$

Where  $w$  is the inertia weight,  $c_1$  and  $c_2$  are the acceleration constants,  $r_{1,j}(t)$ ,  $r_{2,j}(t) \sim U(0,1)$ , and  $k = 1 \dots N$ . The velocity is thus calculated based on three contributions:

a fraction of the previous velocity. The cognitive component which is a function of the distance of the particle from its personal best position. The social component which is a function of the distance of the particle from the best particle found thus far (i.e. the best of the personal bests). Important issue in standard PSO algorithm is rate of its fast convergence that possibly lead to placing result of algorithm in local optimal. It's clear that use of more informational sources increase the space of search and distribution of algorithm and improve problem of PSO. Therefore, in a suggestive algorithm that we called it, ECPSO1, briefly, try for increase utility of PSO algorithm implement changes in movement Particle function for Improvement from utility of algorithm. In this equivalence, two random functions ( $rand1, rand2$ ) determined according to recent posed strategies based on chaos map. That this function have hidden order rather than random numbers that are disordered and this change cause improvement of utility from PSO algorithm in clustering.

As you see in 20 equivalence, the new Particle speed in primary PSO text calculate according to local best situation of Particle and global best situation from all of the Particle. Entrance of global best situation in speed of Particle, caused intense movement in displacement of Particles for going to new situation and also caused fast convergence of PSO algorithm and increase probability placing algorithm in local optimal.

Because if this situation be a incorrect and deviant situation caused that particle have intense deviation in its movement. That is, one misled leader deviate all the population and algorithm place in trap of local optimal and can't reach to all over optimal answer. Therefore, in our suggestive approach for solving this problem, consider ( $k$ ) global best for all the population that number of this global best determine according to population and during performance of

---

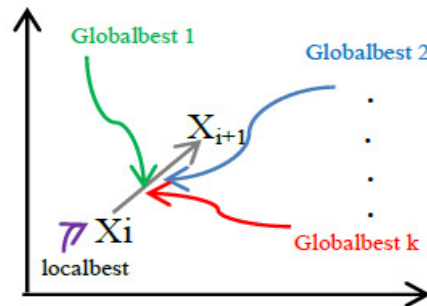
<sup>1</sup> Extended chaotic particle swarm optimization

algorithm amounts of (k) global best up to date according to steady distribution then in calculation of new particle speed average of them and we have speed of average difference of several situation (global-best) from particle situation.

This way is caused moderation particle movement, and increase the probability of reach to all over optimal answer and decrease the probability of placing algorithm in trap of local optimal and we consider  $c_1$ ,  $c_2$  in form comparative in algorithm that in order determine tendency to local best situation and global-best of particle and (w) determine tendency to previous particle speed, that is, any amount approach to last algorithm repetitions. Decrease rate of this variables.

Extended speed equivalence is:

$$V_{new} = W \times V_{old} + c_1 \times c_r \times (localbest - x_i) + c_2 \times (1 - c_r) \times (\text{mean}(globalbest1, globalbest2, \dots, globalbest k) - x_i)$$



## 5. SIMULATION

New algorithm perform by using of MATLAB software. Then, for evaluation and measurement from suggestive way in comparison with 4 clustering algorithm (GA-PSO-PSO+K-means-CPSO) we use of 3 real and standard informational base from UCI site. Table 2 shows the characteristics of these categories.

Table2.Data sets used

Data set	Number of samples	Number of classes	Number of characters
iris	150	3	4
seeds	210	3	7
glass	214	6	10

In this article we use of 4 criterion and index, that are:

Number of algorithm repetition to reach to termination bet. Number of calculation from competence function in algorithm Exactness of cluster or purity criterion Index of validity In continuance we express the operation of every index.

### 5.1. Analysis of 4 criterion and posed indexes for comparison of mentioned algorithms operation:

Number of algorithm repetition to reach to termination bet:

We use of this algorithm for evaluation of algorithm convergence speed. Number of calculation from fitness in algorithm:

In this index, we calculate the number of calculation from fitness in algorithm and whatever the number of this index be further, that will imply further fiscal complexity from that algorithm. On the other hand, with this index, we evaluate the fiscal complexity from clustering algorithm index of

### 5.3. Clustering carefulness with purity index

This index evaluate data clustering carefulness by clustering algorithm. The extent of this index is between 0 and 1 and whatever extent of this algorithm be closer to 1, the clustering carefulness is higher and more desirable. Criterion of purity evaluation for all obtained clusters from clustering algorithm. That calculation of this criterion for each cluster, calculate with equation (8):

$$p_j = \frac{1}{n_j} \text{Max}(n_j^i)$$

That is for each cluster we consider maximum similarity between each cluster to all the available classes in data collection.

Total purity evaluation by equation (9):

$$\text{purity} = \sum_{j=1}^m \frac{n_j}{n} p_j$$

In this relation ( $n_j$ ) is size of cluster, ( $j$ ) and ( $m$ ) are number of cluster and ( $n$ ) is number of sample.

### 5.4. Index of validity

This index evaluation proportion of intra-cluster distance sum and distance between clusters, And clustering is more desirable if intra-cluster distance sum is less and outside-cluster distance sum is more therefore the less amount of this index imply to better data clustering. Termination bet form all of the mentioned algorithm is Convergence criterion. That is in two continuous repetition of clustering, if amount of validity index be less than 0.0001, algorithm finished because we reach to convergence in clustering.

Table3. Result from posed indexes in foresaid algorithm after evaluation average of 10 times accomplishment in iris data sets.

Algorithm	Number of	Number of	Clustering	Index of
PSO	5.8	23.2	0.71	25.61 -6.8
CPSO	5.6	22.4	0.75	23.74 -5.01
PSO+K-means	16.8	47.2	0.76	33.2 -5.01
GA	5.1	49.8	0.71	33.2 -14.44
<b>Ecpso</b>	<b>6</b>	<b>24</b>	<b>0.79</b>	<b>21.29 -</b>

As shown in Table 3 new posed algorithm has more suitable convergence speed for reach to all over optimal answer rather than PSO+K-means compound algorithm and it have a little difference with PSO, CPSO, GA algorithm of this criterion. From the point of view of index 2,

algorithm fiscal complexity is less than GA, PSO+K-means algorithm and have high difference with them. Because of carefulness and quality are two important factor in data clustering procedure that cause discovery of important and exact information in primary raw data. And even have higher priority in applications from clustering such as: medical and engineering, clustering quality and carefulness. So, the algorithm that can increase carefulness in clustering procedure, can be more suitable algorithm for clustering from the point of view of this index posed algorithm is superiority rather than other algorithms and have high clustering carefulness.

This remedial results obtain for algorithm because of particle move toward optimal answer with cautionary, further carefulness and more suitable speed. Finally, from the point view of index 4, that is, quality of clustering which its goal is growth of compaction and aggregation in one cluster and further separation between different clusters, posed algorithm have higher clustering quality rather than other algorithms. This domination demonstrate with numerical difference 2.53 from algorithm validity index which it is optimal amount is 18.76 in iris data collection.

Table4. Result from posed indexes in foresaid algorithm after evaluation average of 10 times accomplishment in seeds data sets.

Algorithm	Number of algorithm repetition to reach to termination bet	Number of calculation from fitness in algorithm	Clustering carefulness with purity index	Index of validity
PSO	5.7	22.8	0.69	42.26 -3.86
CPSO	5.8	23.2	0.73	39.10 -0.7
PSO+K-means	11	41	0.74	42.78 -4.38
GA	4.2	41.6	0.73	56.5 -18.1
<b>Ecpso</b>	<b>7</b>	<b>28</b>	<b>0.77</b>	<b>37.8 -0.6</b>

As shown in Table 4, and we show in previous experiment, our posed algorithm according to result of index that show number7 become convergence to termination bet, more fast pso+ k-means algorithm and also such as pervious experiment, fiscal complexity of new algorithm is less than PSO-K-means and GA algorithm about clustering carefulness index or purity that expressed matters about its importance in clustering and we also show in previous experiment, this algorithms which illustrate with number 0.77. about clustering quality index or validity from this algorithm with numerical difference 0.6 in proportion with original amount from this index in seeds collection that is 38.40 consider data clustering more suitable and with more quality other than other algorithms.

Table5. Result from posed indexes in foresaid algorithm after evaluation average of 10 times accomplishment in glass data sets.

Algorithm	Number of	Number of	Clustering	Index of
PSO	6.8	27.2	0.75	10.6 -
CPSO	6.6	26.4	0.73	12.7 -
PSO+K-	33.9	63.9	0.78	7.1
GA	7.6	69.8	7.6	13.8 -
Ecpso	7.1	28.4	0.80	9.8 -0.93

As shown in Table 5, new algorithm have suitable convergence speed other than PSO+K-means and GA algorithms and this algorithm about fiscal complexity is less than GA and PSO+K-means algorithms. About clustering carefulness index as previous experiments that we show in this experiment, posed algorithm have high clustering carefulness other than other algorithms.

About clustering quality index or posed algorithm validity with numerical difference 0.93 in proportion with original amount from this index in glass data collection, that is 8.87, this algorithm is more suitable and more quality other than other algorithms. We can inference from results of posed experiments that suggestive algorithm is better than other algorithms and have good operation in index of clustering carefulness and clustering quality other than four other algorithms.

## 6. CONCLUSIONS

Data mining is one Helpful technology which used and develop by expansion of data base technology. We can say that this technique is one current instrument for analysis and in for ration extraction among a lot of data. We can discover remedial patterns without interference of user by it. Clustering is one of the most common data mining techniques. Which have application in very cases. In this paper we evaluate and express some of the techniques and researches that posed in current years for improvement problems k-means clustering algorithm. As you see, some of this technique are for improvement problems from previous posed algorithms and some of them are new algorithm for improvement problems from k-means algorithm. After review to history of research, we posed new algorithm for improvement problem from placing algorithm result in local optimal. And evaluation it by 3 real data sets and some of the validity index. We show that posed algorithm have clustering carefulness and clustering quality. We can use of this algorithm when clustering carefulness and quality is more important from the time.

## REFERENCES

- [1] Shalini S Singh, N C Chauhan, K-means v/s K-medoids: A Comparative Study, National Conference on Recent Trends in Engineering & Technology, 2011.
- [2] Chih-Ping Wei, Yen-Hsien Lee and Che-Ming Hsu, Empirical Comparison of Fast Clustering Algorithms for Large Data Sets, 2000, Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000.
- [3] Zhexue Huang, Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values, 1998.
- [4] Suresh Chandra Satapathy, Gunanidhi Pradhan, Sabyasachi Pattnaik, JVR Murthy, PVGD Prasad Reddy, Performance Comparisons of PSO based Clustering, 2009, InterJRI Computer Science and Networking
- [5] Li-Yeh Chuang a, Chih-Jen Hsiao b, Cheng-Hong Yang, Chaotic particle swarm optimization for data clustering, Elsevier ad-hoc, 2011.
- [6] Sandeep Rana, Sanjay Jasola, Rajesh Kumar, A hybrid sequential approach for data clustering using K-Means and particle swarm optimization algorithm" ,2010, International Journal of Engineering, Science and Technology, 2010.
- [7] L.E. Agustín-Blas, S. Salcedo-Sanz, S. Jimenez-Fernandez, L. Carro-Calvo, J. Del Ser, J.A. Portilla-Figueras, A new grouping genetic algorithm for clustering problems, Elsevier ad hoc 2012.
- [8] DW van der Merwe, AP Engelbrecht, Data Clustering using Particle Swarm Optimization, Conference Publications, 2003.
- [9] P.S. Shelokar, V.K. Jayaraman, B.D. Kulkarni, ' An ant colony approach for clustering, Elsevier, 2003.
- [10] Taher Niknam, Babak Amiri, An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis, Elsevier Engineering Applications of Artificial Intelligence, 2010.
- [11] Taher Niknam a,n, Elahe Taherian Fard b, Narges Pourjafarian b, Alireza Roustae, " An efficient hybrid algorithm based on modified imperialist competitive algorithm and K-means for data clustering, Elsevier Engineering Applications of Artificial Intelligence, 2011
- [12] Cheng-Lung Huang, Wen-Chen Huang, Hung-Yi Chang, Yi-Chun Yeh, Cheng-Yi Tsai, " Hybridization strategies for continuous ant colony optimization and particle swarm optimization applied to data clustering, Elsevier Engineering Applications of Artificial Intelligence, 2013.