

DYPSOKM: A Dynamic Union Of PSO And K-Means, A Better Cluster

Qin Ke, Liusheng Huang, Hongli Xu

School of Software Engineering of USTC(Suzhou Institute for Advanced Study, USTC)
Suzhou 215123,China

Keywords: clustering; K-means; PSO; multiple swarms; amplitude limiting of speed

Abstract

As one of the most famous clustering algorithms, K-means is simple and effective but easily falls into local optimal solution. Aimed at this flaw, many methods including PSO had been applied to optimize K-means. As a typical swarm intelligence optimization algorithm, PSO (particle swarm optimization) has better global convergence and robustness. This paper will apply the PSO to optimize the K-means clustering algorithm based on the basic PSOKM. On the one hand, we initialize the particles using dichotomy K-means. On the other hand, this paper in return utilizes the feature of multiple swarms shown in k-means to build the multiple-swarm PSO. In main computational details, we light weight the calculation of multiple-population in order to enhance computational efficiency. Meanwhile, the speed of particles will be limited as a certain way to improve the validity of algorithm. Finally, Experiment results of our algorithm shows better convergence and validity compared with other algorithms mentioned in this paper.

1 Introduction

Clustering is assembling objects into group such that objects within a same group are similar and objects into different groups are dissimilar. The main goal of clustering is to obtain the global optimal clustering. Many approaches [1, 2, 3, 4, 5] have been developed. As a typical clustering algorithm of prototype-oriented in non-supervision machine learning, K-means is simple and effective while processing various data types. But it inevitably exists a lot of drawbacks because of the initial clustering centroids and other factors. Aimed at the random initial centroids, many methods such as dichotomy K-means [6], K-means++ [7] have been proposed to solve the problem. Post-processing techniques [8] are frequently-used ideas to enhance the global convergence of clustering algorithm.

As a typical swarm intelligence optimization algorithm, PSO (particle swarm optimization) originated from the research of simplify social model of birds. Reynolds proposed the model of Boid [9] to simulate the assembling behavior of birds in 1987. Then Kennedy and Eberhart proposed the PSO [10] based on the model until 1995. PSO is a global random heuristic search algorithm and has relatively strong global

convergence and robustness. It has been widely applied in various optimized fields including data mining and machine learning. Omran [11] et al. proposed an unsupervised image classification algorithm based on PSO in 2002, which is the earliest clustering algorithm based on PSO. Then Merwe etc proposed the basic PSO clustering algorithm. Furthermore, they combined the PSO and k-means. Clustering result shown it had better convergence and validity.

What this paper worked on is based on the flowing problems. Firstly, the random initial centroids is one of the most important issues which influence the validity of K-means. So this paper firstly applies the dichotomy K-means replacing K-means to generate a centroid as one of the initial particles. Secondly, in every iteration of PSO, algorithm will conduct clustering according to Cy-pres. Every cluster can be viewed as a swarm. As is known to us, the lose of population diversity is the key factor which leads to PSO falling into local solution. So this paper will utilize the feature of multiple-population emerging in K-means to build the multiple-population PSO. Actually, methods such as DMO-PSO [12,13], KDMSPSO [14] about dynamic multiple-population building have been proposed to optimize PSO, we will blend the idea in the K-means and make proper adjustments. Specially, we proposed a light weight strategy in the calculation of multiple-swarm in order to improve the actual running efficiency. The idea is based on a common truth i.e. the centroid can be viewed as the better leader of a group.

At last, this paper will compare the convergence and validity among our algorithm named as DYPSOKM and other typical mentioned algorithms by experiments. In main computational details, we will make an expression to the setting of some main parameters. Specially, in consideration of the limitation of actual solutions, this paper will limit the speed of particle according to a special way.

2 BPSOKM: Basic union of K-means and PSO

2.1 Basic model of K-means

As a typical clustering algorithm of prototype-oriented in non-supervision machine learning, K-means is widely applied and effective. K-means regards mean value as prototype, then every cluster is the collection of object which is nearest to this cluster. Proximity measure can change with the application scenarios. And we usually select the Euclidean distance as the Proximity measure.

Firstly, Figure 1 gives the flow of the K-means.

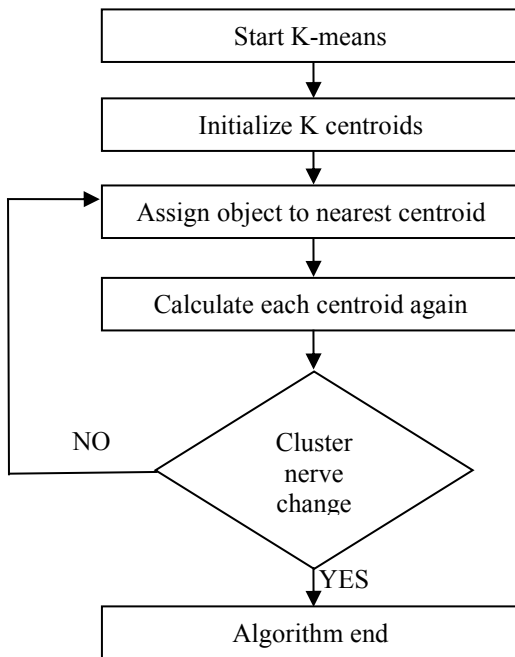


Figure 1: Flowchart of the K-means.

Original K-means randomly initialize the centroids. When assigning points to the nearest centroid, the Euclidean distance is applied extensively while Manhattan distance or Cosine Similarity are also used. This paper will adopt the Euclidean distance as the Proximity measure which defined as follow.

$$dist(c_i, x) = \|c_i, x\| \quad (1)$$

$\|c_i - x\|_2$ denotes the Euclidean norm, c_i denotes the centroid, it will be calculate according to the mean value .

$$c_i = \frac{1}{m} \sum_{x \in C_i} x \quad (2)$$

Here we introduce the most popular objective function i.e. SSE as the evaluation function of clustering. The lesser the SSE is, more valid the clustering is. Other evaluation index will be discussed in the part of experiments.

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} dist(c_i, x)^2 \quad (3)$$

The time complexity of K-means is $O(KnT)$, n denotes the number of objects, K denotes the number of clusters and T denotes the number of iterations. K-means is simple and fast especially processing the big data set because of its linear time complexity. But there also exists some flaws. Firstly, the count of clusters is uncertain. Then the random initialization may cause algorithm to fall into locate solution.

In order to understand the flaw and then improve the performance aiming at it, a small experiment based on the small two-dimension data set of 80 line was conducted. Firstly, this paper will confirm roughly the number of K

according the curve graph of SSE. figure 2 shows the variation trend of SSE when the number K changes.

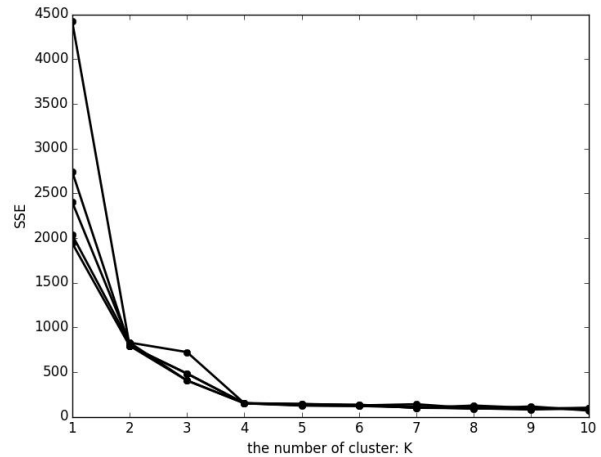


Figure 2 : The SSE curve graph of the various K.

The figure is the result of 5 random experiment ,we can intuitively see that $K = 4$ is the best choice in consideration of the smallest SSE and the smaller number of K.

Secondly, the figure below will give the scatter diagram after we confirm the number of K i.e. 4 and initialize the centroid randomly.

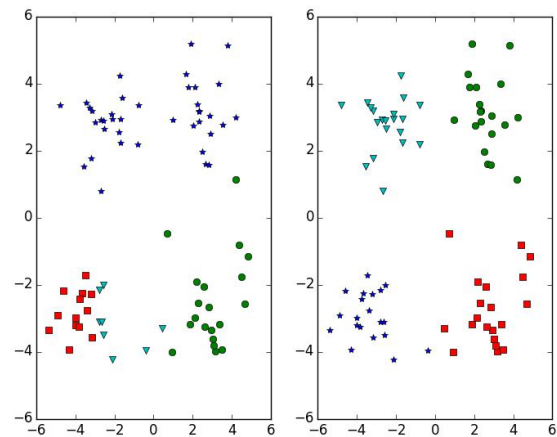


Figure 3 : The scatter diagram of clustering using K-means.

We can easily find that the left one of the figure falls into the locate solution while the other one has fine global convergence performance. The key factor is the random initialization of centroids. Even though we only show the result of several times about the experiment, it exactly exist the probability to fall into locate solution. And it exactly is the power of our studying about it. Several methods have been developed aimed at the random initialize centroid, e.g .dichotomy k-means, k-means++ etc. Inspired by the work of Merwe, this paper will introduce the idea of PSO which

has better global convergence performance and validity while optimizing K-means.

2.2 Basic model of PSO

PSO is a swarm intelligence optimization algorithm originated from the research of simplify social model of birds. It has been widely applied in numerous optimized fields e.g. schedule optimization^[15], economic dispatch^[16], data mining^[17], robot control^[18] etc.

Generally, the problem of optimization can be defined as the following model and the maximization problem can be easily transformed to it.

$$\min F(x) = (f_1(x), f_2(x), \dots, f_m(x))$$

$$\begin{cases} g_i(x) \leq 0, i = 1, 2, \dots, p \\ h_j(x) = 0, j = 1, 2, \dots, q \\ x \in R^n \end{cases} \quad (4)$$

$x = (x_1, x_2, \dots, x_n) \in R^n$ denotes the decision variable;

$F: R \rightarrow R$ denoted the objective function; $g_i: R^n \rightarrow R$

and $h_j: R^n \rightarrow R$ denotes the constraint. In order to global solution and locate solution, we will give the the definition of mathematics as follow.

Definition 1. If $x^* \in D, \forall x \in D, f(x^*) \leq f(x)$ is right, then x^* is called as the global optimized solution.

Definition 2. If there exists a neighborhood

$$N(x^*, \delta) = \{x \mid |x - x^*| \leq \delta, \delta > 0\}$$

when $x \in D \cap N(x^*, \delta), f(x^*) \leq f(x)$ is right, then x^* is called as the locate optimized solution.

Our object is obtaining the first one. As is mentioned before, Kennedy and Eberhart proposed the PSO based on the model of boid. Every particle adjusts its direction and speed of flying according to its history best locate and the best location of the population.

Assume the vector of X denotes the location of particle and V denotes the speed of the particle. They are all can be any dimension. $f(X)$ denotes the minimize fitness function. The history best location of every particle is defined as:

$$pbest_i(t+1) = \begin{cases} pbest_i(t) & \text{if } f(X_i(t+1)) \geq f(pbest_i) \\ X_i(t+1) & \text{if } f(X_i(t+1)) < f(pbest_i) \end{cases} \quad (5)$$

$i = 1, 2, \dots, n$ denotes the number of the particle swarm. The best location of population defined as :

$$gbest(t) = \min\{f(pbest_1), f(pbest_2), \dots, f(pbest_n)\} \quad (6)$$

The equation about the speed and location of particles is defined as:

$$V_i(t+1) = V_i(t) + c_1 r_1 (pbest_i(t) - X_i(t)) + c_2 r_2 (gbest(t) - X_i(t)) \quad (7)$$

$$X_i(t+1) = X(t) + V(t+1) \quad (8)$$

c_1, c_2 are learning factors and r_1, r_2 are random number between 0 and 1. The update of speed is decided by three parts. The first one is its last generation which can balance the ability of global exploration and local exploitation. The second one is its learning part from its history best performance while the last one is its tendency to social best performance. These exactly are the expression of swarm intelligence. Figure 4 gives the flow of PSO.

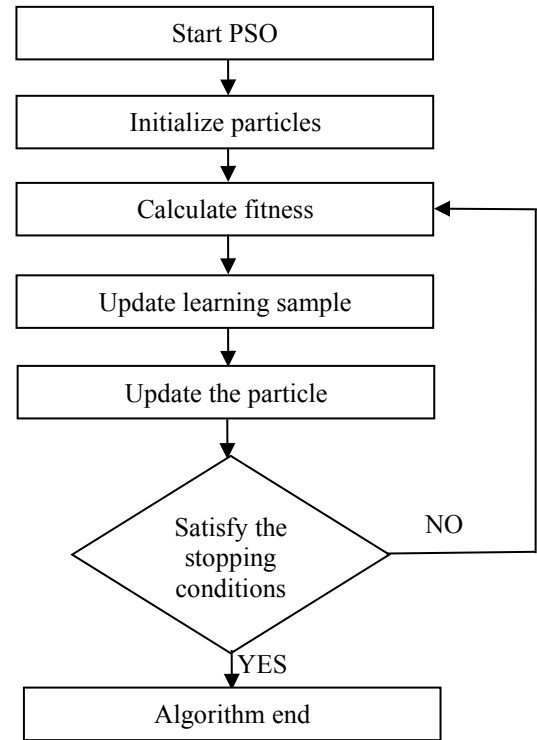


Figure 4: Flowchart of the basic PSO.

In order to show the global convergence of PSO, this paper chooses a function^[19] which has 4 local minimal location as the fitness function to test PSO.

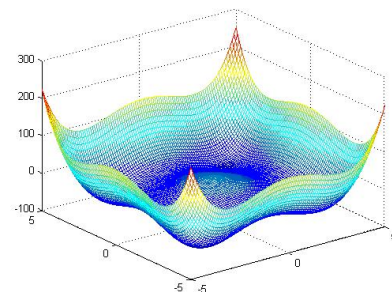


Figure 5: The graphic of tested function.

$$f = \sum_{i=1}^2 (x_i^4 - 16x_i^2 + 5x_i) / 2 + \cos(x_1 + x_2) \quad X = [-5, 5]^2 \quad (9)$$

Above we give the definition and graphic of f , we can find that it exactly has 4 locate solution and one global solution near to the point of (3.0,3.0). Then this paper solves this problem using PSO. The solution is about (-2.8838823,-2.90938893), the figure 6 gives the convergence curve.

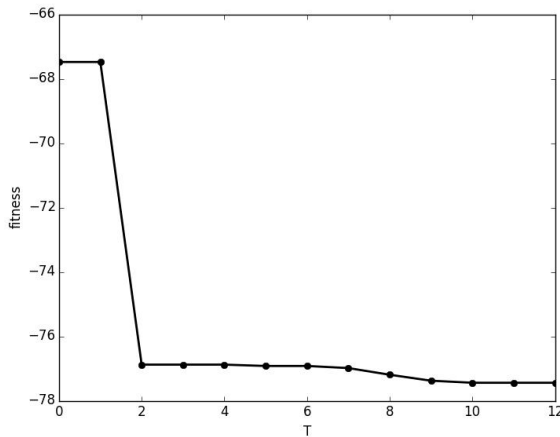


Figure 6: the convergence curve using PSO.

It has proved the global convergence performance of PSO. Spontaneously, we can use basic PSO to optimize the clustering problem. The experiment will be shown in the part of experiments. Then this paper introduces the basic combination of PSO and K-means.

2.3 Basic combination of PSO and K-means

As is mentioned in the part of introduction, Merwe [20] etc proposed the basic PSO clustering algorithm based on the work of Omran. Every particle of the particle swarm includes a data vector which denotes the centroid of the cluster and it represents a candidate partition. The fitness Merwe used is defined as follow.

$$J_e = \frac{\sum_{j=1}^{N_c} [\sum_{z_p \in C_j} d(z_p, m_j)] / |C_j|}{N_c} \quad (10)$$

$m_j, d(z_p, m_j)$ denote the mean value of cluster and Euclidean distance, which is equivalent to the definition in equation (1) and (2). N_c, C_j denotes the count of cluster and the number of objects in every cluster. This fitness is actually similar to SSE, which both are the indication of condensation degree of clustering. This paper called the methods as **BPSOKM** proposed by Merwe et al. In order to be more convenient and give the flow of algorithm as flow.

The purpose of algorithm is lowering the probability of falling into the locate solution. Experiments by Merwe had shown that method enhance the validity and convergence speed to some extent.

“
”

Algorithm BPSOKM (N,K,T)

- 1. Initialization:** initialize one particle using K-means and other particles randomly.
 - 2. Repeat:**
 - Aiming at each particle: Assign object to nearest centroid.
 - Calculate fitness of particle according to J_e
 - Update learning sample
 - Update the particle
 - Until:** satisfy the stopping conditions
 - 3. Output:** the location of optimized particle
-

Based on the working of Merwe etc, some methods which combine the PSO and K-means have been developed and applied. Next, this paper will make some adjustments based on BPSOKM aimed at improving the validity and convergence further on.

3. DYPSONKM: Our version about the combination of PSO and K-means

3.1 Problems and Ideas

Based on the BPSOKM, we can easily find that the bottleneck mainly includes two aspects. One is the the initialization of the particle, and the another is the performance of PSO itself. A better initialized particle swarm which has good topological structure [18] will have better performance generally.

Firstly, since there exists many methods to produce the initialized centroids and dichotomy K-means is simple and more validity than K-means, this paper will use it to replace the K-means.

Secondly, the lose of population diversity is the key factor leading to PSO falling into local solution, this paper will utilize the multiple swarm feature in K-means to maintain the population diversity. YanMing, Liu[14] et al. have proposed a version of PSO which directly divides several subgroup from all particles by K-means. what this paper do is building the same number of subgroup with the count of particle i.e. K. It is feasible when PSO is applied in turn for clustering. As it is obvious that a cluster is a population .

The next challenges mainly include two aspects. How should we build the subgroup and the frequency of building. In each iteration, an updated particle represents a clustering, so we can view the produced centroids as a subgroup. But frequent building will cause the loss of optimal information even enormously increase the cost of running. So this paper develops a light weight way to replace the multiple swarm strategy but still have the effect of multiple swarm. As is known to us, although the leaders necessarily do best in knowledge, but they must be the kernel of the group. So it is reasonable to use the centroids to replace the performance of group especially in PSO which is intelligent with K-means and does not reply on the part seriously. we can choose the mean value of the centroids as the population information.

“

3.2 Methods

Firstly, the flow of dichotomy K-means [6] will be introduced as follow. It is based on a simple idea. In order to obtain K clusters, we can split all objects into two clusters and then choose one to split until K clusters have been generated. The algorithm of **DICHOTOMYKM** is defined as follow.

Algorithm DICHOTOMYKM (K)

1. **Initialization:** initialize the clusters table using the Cluster which includes all objects.
 2. **Repeat:**
 Choose one cluster from the clusters table in way.
for $i = 1$ to count of experiment **do**
 dichotomia the selected cluster using K-means.
end for
 Select the two clusters which have the minimum total of SSE.
 Add the two clusters to clusters table.
Until: satisfy the stopping conditions
 3. **Output:** K clusters.
-

The speed of every particle can be decided by self-knowledge, subgroup-knowledge and social-knowledge. It adds a item of subgroup based on equation (7).

$$\begin{cases} V_i^{old} = V_i(t) + c_1 r_1 (pbest_i(t) - X_i(t)) + c_2 r_2 (gbest(t) - X_i(t)) \\ V_i^{new}(t+1) = V_i^{old} + c_3 r_3 (cbest_i(t) - X_i(t)) \end{cases} \quad (11)$$

$cbest_i(t)$ denotes the history best location of the every subgroup. The other part s of PSO will keep the traditional flow. To sum up, the final version of our method will be give below.

Algorithm DYPSONKM (N,K,T)

1. **Initialization:** initialize one particle using **dichotomy K-means** and other particles randomly.
 2. **Repeat:**
 Aiming at each particle: Assign object to nearest centroid.
 Calculate fitness of particle according to J_e
 Calculate the mean of centroids aimed at particle
 Update learning sample include subgroup part.
 Update the particle according to the new equation (11)
Until: satisfy the stopping conditions
 3. **Output:** the location of optimized particle
-

Figure 7 gives the flow chart of our method based on BPSOKM.

4 Experiments and Results

4.1 Computational Details

At last, this paper will compare the performance of PSO, K-

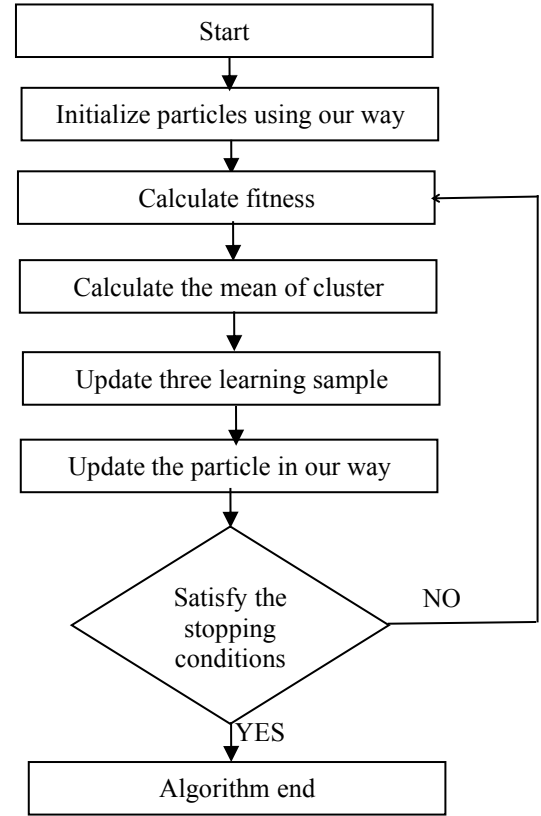


Figure 7: The flow of DYPSONKM.

means, BPSOKM, DYPSONKM. It mainly refers to two aspects including the convergence performance and validity. Before working on the experiments, we firstly choose the appropriate data sets and set suitable value of some important parameters. Table 1 gives the information of data Sets used in this paper and they have been widely applied in Cluster Analysis. Table 2 shows the setting of parameters about PSO, it is based on the studying of other scholars.

Data set	Count	dimension	Cluster label
Test_Set	80	2	NO
4k2_far	400	2	YES
Iris	150	4	YES

Table 1: Information of data set.

parameter	value
c1	2.0
c2	2.0
c3	1.0
r1、r2、r3	random(0,1)
T	15
N	15
K	4
vk	0.4

Table 2: Setting of parameters.

In Table 1, Cluster label denotes whether the data set has cluster label. Although clustering is a unsupervised leaning, the label information can be utilize to evaluate the validity of clustering.

In Table 2, T denotes the number of iteration, N denotes the number of particle and K denotes the default value of K-means.

In computational details, we must prevent the particle flying out of the area of clustering. So this paper solute this problem by lit the speed of particle according to the flowering aspect.

If the searching space is $[-x_{\max}, x_{\max}]$, then maximum speed named V_{\max} is defined as $v_k * V_{\max}$. So the update of speed based on (11) must add the flowering step:

$$V_i(t+1) = \begin{cases} \min\{V_i(t+1), V_{\max}\} & \text{if } V_i(t+1) > 0 \\ \max\{V_i(t+1), -V_{\max}\} & \text{if } V_i(t+1) < 0 \end{cases} \quad (12)$$

On the other hand, we will keep the same conditions of experiments except the necessary difference of the algorithms.

4.2 Results

Firstly, this paper gives the mean convergence curve of the four algorithms based on 10 experiments. To keep the consistency, $T=0$ denotes the mean SSE of the data set and $T=1$ denotes the initialized time. And we use the Je defined in equation (10) as the dependent variable. It is one kind of condensation degree of clusters. Figure 8 gives the convergence curve based on Iris.

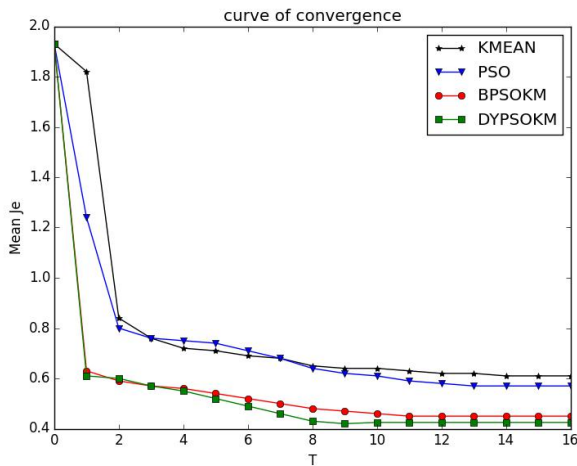


Figure 8: The curve of convergence

We can easily observe from above that DYPKOKM has faster convergence performance and lower Je .

Secondly, this paper will pay attention to the validity of clustering which usually includes unsupervised evaluation and supervised evaluation. When thinking about unsupervised evaluation, SSE and SSB are usually applied. They are respectively the condensation degree and separation degree

of the whole clusters. SSE has been defined in equation (3), and SSB is defined as follow.

$$SSB = \sum_{i=1}^K m_i \text{dist}(c_i, c)^2 \quad (13)$$

There m_i denotes the count of cluster i , c_i denotes the cent of cluster i and c denotes the cent of whole data set. The larger SSB is, the better separation of clustering is. And the SSE and SSB satisfy the flowering equation:

$$TSS = SSB + SSE \quad (14)$$

TSS denotes the sum of squares. So the minimum SSE corresponds to the maximum SSB, which both represent the validity of clustering. On the other hand, this paper will choose entropy^[21] and purity^[22] as the supervised evaluation. They are also showing the extent that clusters include the objects belonging to single class. This needs extra label information of the data sets. Entropy is defined as:

$$e_i = - \sum_{j=1}^L p_{ij} \log_2 p_{ij} \quad (15)$$

p_{ij} computes the probability that the objects of cluster i belong to the cluster j , which is defined as:

$$p_{ij} = m_{ij} / m_i \quad (16)$$

m_i and m_{ij} denotes the count of cluster i and the count of objects of cluster i which belonging cluster j . The total entropy is defined as:

$$e = \sum_{i=1}^K \frac{m_i}{m} e_i \quad (17)$$

The purity of cluster i and the total purity is defined as:

$$p_i = \max_j p_{ij} \quad (18)$$

$$\text{purity} = \sum_{i=1}^K \frac{m_i}{m} p_i \quad (19)$$

Table 3 gives the results of mean validity of clustering while running each algorithms 10 times based on 4k2_far.

Algorithm	SSE	SSB	TSS	Entropy	Purity
KMEAN	138.11	4236.77	4374.88	0.2480	0.9222
PSO	137.00	4237.88	4374.88	0.2491	0.9222
BPSOKM	135.81	4239.07	4374.88	0.2496	0.9225
DYPKOKM	134.70	4240.18	4374.88	0.2498	0.9226

Table 3: Validity of clustering.

Smaller SSE and lager SSB show that the result of clustering has better condensation degree and separation degree. Meanwhile, higher Entropy and Purity indicate that clusters produced by our version of algorithm include the objects belonging to originate single cluster more likely.

At last, this paper will give the scatter graph of algorithms. Because of the random initialization, we choose the following graphic based on 4k2_far as the typical scatter which respectively named I, II, III, IV.

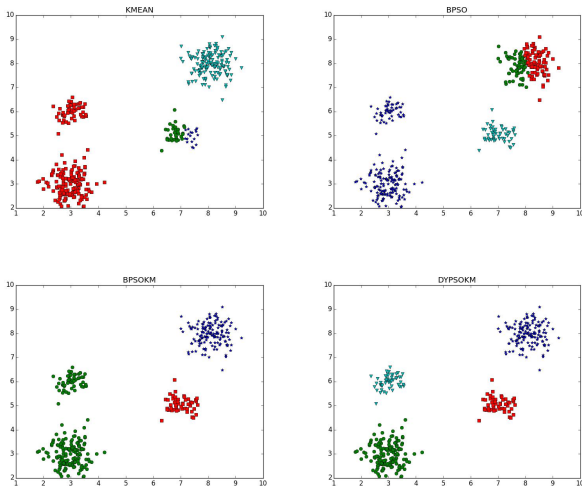


Figure 9: Typical scatter of clustering

Then we statistic the frequency of emerging the above-mentioned clustering results based on 10 experiments each algorithm.

Algorithm	I	II	III	IV	OTHER
KMEAN	0.2	0.1	0.1	0.5	0.1
PSO	0.0	0.2	0	0.6	0.2
BPSOKM	0.0	0.1	0.1	0.8	0
DYPSONM	0.0	0.0	0.0	0.9	0.1

Table 4: Scatter statistics of clustering.

We can easily observe that DYPSONM occupies the largest rate of type IV which indicates the best clustering.

Acknowledgements

This work was supported by Laboratory of network information security in Suzhou research institute of USTC. I will express my great appreciation to my tutor and senior fellow apprentice.

References

[1] Agrawal R, Gehrke J, Gunopulos D, Raghavan P. "Automatic subspace clustering of high dimensional data for data mining applications", (1999).
 [2] Nagesh H, Goil S, Choudhary A. "Efficient and scalable subspace clustering for every large data sets", (1999).
 [3] Sheikholeslami G, Chatterjee S, Zhang A. "Wavecluster: A multi-resolution clustering approach for very large spatial databases", Proc. **24**, pp. 428-439, (1998).
 [4] Kaufman L, Rousseeuw p. "Finding groups in data in :An introduction to Cluster Analysis", (1990).
 [5] Sneath P H A, Sokal R R. "Numerical Taxonomy. The Principles and Practice of Numerical Classification", (1973).
 [6] Hartigan J A, Wong M A. "A K-Means Clustering Algorithm ", Royal Statistical Society, **28**, pp. 100-108, (1979).

[7] Bahmani B, Mouseley B, et al. "Scalable k-means++", VLDB Endowment, **5**, pp. 622-633, (2012).
 [8] Wagstaff K, Cardie C, et al. "Constrained k-means clustering with background knowledge", ICML, pp. 577-584, (2001).
 [9] Reynolds C W. Flocks, herds, schools. "A distributed behavioral model", Computer Graphics, **21**, pp.25-34, (1987).
 [10] Kennedy J, Eberhart R C. "Particle swarm optimization", ICNN, pp. 1942-1948, (1995).
 [11] O Marang M, Salman A, Engelbrecht A P. "Image classification using particle optimization",APCSEL, pp. 370-374, (2002).
 [12] Zhao S Z, Liang J J, Suganthan P N. "Dynamic Multi-Swarm Particle Optimizer with Local Search for Large Global Optimization", World Congress on Intelligent Control and Automation, pp. 3138-3142, (2010).
 [13] Liang J J, Suganthan P N, Deb K. "Dynamic Multi-swarm Particle Optimizer", IEEE Swarm Intelligence Symposium (2005).
 [14] YM Liu, CL Shui et al. "Dynamic multi-swarm optimizer based on K-means clustering and its application", Control and Decision, pp. 53-58, (2011).
 [15] Allahverdia A, AL-Anzib F S. "A PSO and a tabu search heuristics for the assembly scheduling problem of two-stage distributed database application". Computer & Operations Research, **33**, pp. 1056-1080, (2006).
 [16] Swarup K S, Kumar P R. "A new evolutionary computation technique for economic dispatch with security constrains", Electrical Power and Energy Systems, **28**, pp. 36-42, (273-283).
 [17] Sousa T, Silva A, ET AL. "Particle swarm based on data mining algorithms for classification tasks", Parallel Computing, **30**, pp. 767-783, (2004).
 [18] Dong H K, Jin I P. "Intelligent PID controller tuning of AVR system using GA and PSO", International Conference on Intelligent Computing, pp.366-375, (2005).
 [19] C Malherbe, E Contal, N Vayatis. "A Ranking Approach to Global Optimization", Machine Learning, (2016).
 [20] Merwe D W, Engelbrecht P A. "Data clustering using particle swarm optimization", Proceedings of the IEEE Conference on Evolutionary Computation, pp. 215-220, (2003).
 [21] R Jenssen, K E Hild, D Erdogmus. "Clustering using Renyi's entropy", Neural Networks, (2003).
 [22] A Hotho, S Staab, G Stumme. "K-means clustering algorithm with meliorated initial center", Data Mining ICDM, (2003).