An Improved PSO-GA Hybrid Algorithm Based on P Systems for Data Clustering

Chenggong Qiu[1a], Laisheng Xiang[2b*], Xiyu Liu[1c] and Huaning Yan [1d]

[1]*School of Management Science and Engineering Shandong Normal University, Jinan, 250014, China*
[2]*Shandong Normal University, Jinan, 250014, China*
[a]*chgqiu@163.com,* [b]*xls3366@163.com,* [c]*xyliu@sdnu.edu.cn,* [d]*yanhuaning1124@163.com*

Abstract

Clustering is the process of grouping a series of data objects into multiple groups or clusters so that intra-cluster data are similar and inter-cluster data are dissimilar. It is widely used in many fields, such as machine learning, image pattern recognition, and so on. K-means clustering is one of the most widely-used clustering methods. But its random choice of initial cluster centers may lead to trapping into the local optimum and attaining unstable clustering results. For this problem, P systems is introduced into a modified evolutionary algorithm which combines particle swarm optimization with an improved genetics algorithm (PGHAPS) in this paper. On account of the maximal parallelism feature and the communication rules of P systems, the PGHAPS algorithm owns better global search power and convergence capacity. This algorithm is then applied to improve K-means algorithm by optimizing the initial cluster centers (PGHAPS-K-means). Finally, experiment is tested on two artificial data sets and three true life data sets. By comparing with the results of K-means and PSO-K-means, it is proved that the proposed algorithm is efficient in improving the quality of clustering results.

Keywords: Clustering, particle swarm optimization, P systems, maximal parallelism

1. Introduction

Cluster analysis in general is an unsupervised process of partitioning a set of data objects into subsets. Each subset called a cluster. Therefore, objects in a cluster are similar to each other, yet dissimilar to objects in other clusters [1]. Clustering analysis has been widely used in various fields, such as image processing, data analysis, etc. K-means is a commonly used partitioning clustering method. Its final clustering results may depend heavily on the initial random selection .As a result, the K-means method couldn't be guaranteed to converge to the global optimum and often traps into the local optimum [2-3].

Particle Swarm Optimization (PSO) was first proposed by Kennedy and Eberhart in 1995 [4], as one of the best-known bio-inspired algorithms widely used in optimization problems [5]. Genetics Algorithm (GA) was first developed by Holland in 1975, simulated by genetics selection and natural selection of Darwin's biological evolution process [6]. Like PSO, GA is an optimizer based on population so the hybrid algorithm of these two will improve the global optimization capacity and efficiently prevent the situation from falling into the local optimum.

Membrane computing is a domain of natural computing started by Gh. Păun in 1998, known as P systems or membrane systems [7]. It is a parallel distributed computing model inspired from the functionality and structure of living cells and the intersection of

---

∗ corresponding author: Laisheng Xiang

them in the tissue, organ and neural network. The components of a membrane systems are a cell-like membrane structure. Its regions contain multisets of objects that evolve in a synchronous maximally parallel manner, according to evolution rules associated with the membranes. There are three main classes of P systems investigated: cell-like P systems, tissue-like P systems and neural-like P systems. These computing systems are proved to be so powerful that they are in some ways equivalent with Turing machines. Meanwhile, the powerful global search capacity of membrane computing is also proved [8].

K-means algorithm is widely used in many fields due to its simple operations and well-understood concept. However, its drawbacks like easily trapping into the local optimum may curb its usage. Many clustering methods which combined the K-means clustering with evolutionary algorithm were raised in order to improve the K-means clustering [9, 10, 11]. Shelokar et al. introduced an evolutionary technique based on ACO for clustering analysis [12]. Jing Xiao et al. proposed a quantum-inspired genetic algorithm for K-means clustering (KMQGA) [13]. P systems now is a hot cross-discipline topic which involves computer science, mathematics, biology and artificial intelligence, etc. A. Pavel et al. designed a simulator for numerical P systems (SNUPS) which was applied from ordinary differential equations to autonomous mobile robot [14]. S. Yang et al. introduced PSO into membrane computing (MCBPSO), and then applied it into soft sensor modelling of the components of Texaco gasifier syngas [15]. In order to overcome the premature convergence feature of the GA, H. Peng et al. proposed a novel multi-level threshold method for image segmentation based on tissue P systems to search the optimal thresholds for multi-level threshold based on fuzzy entropy [16].

In this paper, a novel evolutionary algorithm, PSO-GA hybrid algorithm (PGHA) based on P systems (PGHAPS) was proposed. Then the hybrid algorithm is used to optimize the initial cluster centers of the K-means so as to improve the clustering quality. PGHAPS algorithm forms a membrane community solution space and improves the global search capacity. In contrast with the particle swarm optimization and the genetics algorithm, the evolutionary algorithm proposed owns better convergence and global optimization capacity. The rest of this paper is organized as follows. Section 2 introduces K-means clustering, PSO, and Variable Population-size Genetic Algorithm. Section 3 describes briefly P systems. Section 4 presents the hybrid evolutionary algorithm for K-means clustering algorithm. Section 5 illustrates experimental results. Finally, Section 6 makes conclusions.

2. Preliminaries

2.1. K-means Algorithm

The K-means clustering algorithm was developed by Hartigan in 1975. As a simply operation clustering approach, it has been widely used in many fields such as image recognition, Web search, and so on [17]. In this method, k initial cluster centers are chosen randomly, and then data points are classified into clusters based on the Euclidean distance calculation [18]. The procedure of the K-means algorithm is described as follows:

(1) Randomly generate k cluster centroids vectors.

(2) For each data vector, assign the vector to the cluster with the closest centroid vector. The distance of a vector to the centroid is determined as follows:

$$D(x_p, z_j) = \sqrt{\sum_{i=1}^{d}(x_{pi} + z_{ji})^2}$$

(1)

Where x_p denotes the p-th dada vector, z_j is the centroid vector of cluster j, and d presents the number of features of each centroid vector.

(3) Calculate the cluster center, using

$$z_j = \frac{1}{n_j} \sum_{\forall x_p \in c_j} x_p$$

(2)

Where n_j is the number of data vectors in cluster j and C_j is the subset of data vectors that form cluster j,

(4) Repeated the above steps until met stopping criterion.

K-means algorithm is easy to operate and efficient in dealing with large data sets. However, it is liable to fall into the local optimum owing to its initial randomly choice of cluster centers. In this paper, a modified evolutionary algorithm is applied to optimal the cluster center to improve the K-means algorithm.

2.2. Particle Swarm Optimization

Particle Swarm Optimization (PSO) was first proposed by Kennedy and Eberhart as a stochastic optimal algorithm based on population intelligence in 1995 [4]. It is a powerful globally accepted evolutionary swarm intelligence method for solving both linear and non-linear problems [19, 20]. Particle swarm optimization simulates the social behavior of birds flocking or fish schooling in order to achieve a self-evolving system, in PSO, each particle was considered as an individual, these individuals flied at a specific speed in the search space, every particle knows its velocity, its position, the optimal position found so far, the optimal position of the whole particle swarm, they dynamically adjusted their flying velocity according to their own flying experience and their companions'[21, 22, 23].

Suppose that N particles exist in D-dimensional searching space, the particle of the swarm is:

$$X_i, \mathrm{i} = 1, 2, \ldots\ldots N$$

(3)

The position of the i-th particle is:

$$X_i = (X_{i1}, X_{i2}, \ldots\ldots, \ X_{iD}) \ , \ \mathrm{i} = 1, 2, \ldots\ldots, \ N$$

(4)

The velocity of the i-th particle is:

$$V_i = (V_{i1}, V_{i2}, \ldots\ldots, \ V_{iD}) \ , \ \mathrm{i} = 1, 2, \ldots\ldots, \ N$$

(5)

The individual extreme value is:

$$P_i = (P_{i1}, P_{i2}, \ldots\ldots, \ P_{iD}) \ , \ \mathrm{i} = 1, 2, \ldots\ldots, \ N$$

(6)

The optimal position of the whole particle swarm founded is:

$$P_g = (P_{g1}, P_{g2}, \ldots\ldots, \ P_{gD}) \ , \ \mathrm{i} = 1, 2, \ldots\ldots, \ N$$

(7)

The particles update their speed and position according to the following formula are as follows:

$$V_i(t+1) = \omega V_i(t) + c_1 r_1 (P_i - X_i) + c_2 r_2 (P_g - X_i)$$

(8)

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

(9)

In the above formula, V_i is the speed of the i-th particle, the range is $[-V_{max}, V_{max}]$, such perset setting avoids the particle flied out of the search space. X_i is the position of it, t stands for the number of iterations, ω is inertia weight, c_1 and c_2 are learning factors, r_1 and r_2 are random numbers in the range [0, 1]. Generally speaking, the particle's new velocity is calculated used equation (3) and its new position is counted in equation (4). The performance of each particle is measured according to a predefined fitness function, which is usually proportional to the cost function associated with the problem, when satisfied the stopping criteria, the searching process halted.

2.3. Variable Population-size Genetic Algorithm

Genetics Algorithm (GA) is an intelligence optimization computational mode which was developed by Holland in 1975 [9]. It is a kind of self-organizing, adaptive artificial intelligence technology, simulated by genetics selection and natural selection of Darwin's biological evolution process [24].

The question to be optimized is usually convert into the form that could be handled by GA. In the ordinary course of events, the GA executes binary code form and described the optimized problem according to genetic code (for example, DNA chromosomes). A population is initialed as the problem is coded as chromosome, each chromosome corresponding to unique individual, the population evolves through biological operations. The population size is determined optionally, in most cases, the size is 100-200. After the population initialization step, the chromosome is evaluated by a given fitness function, the individual with high fitness value will have more chances to reproduce its offspring. It is worth noting that at each generation of the GA, the genetic search operators such as selection, mutation and crossover are applied to make sure their quality are better than their ancestor. It is that, GA operates on a population of potential solutions by applying the principle of the survival of the fittest to produce successively superior approximations to a solution. The process halted until met the termination criterion, and the best entity of the last generation is the optimal solution [25].

Variable population-size according to the natural features of the population-size-changing rule on human beings was introduced into the GA in order to maintain the variety and continuity of the population. The operation is that, after the traditional genetic operation, every generation confronts the certain rate (death or ill rate) which was used to control the number of the generation. If the population number reaches a certain scale, the disease or the war will occur to reduce the scale to the initial size. Obviously, in these processes, the survival is the individual that has higher fitness value [26].

3. P Systems

Membrane computing is also known as P systems or membrane systems; it was first proposed by Pǎun after his years of research about DNA in 1998. Membrane computing aims at abstracting a computational model from the hierarchical structure and the methods of dealing with the compound of the living cell. P systems is a distributed, maximal parallelism and non-determinism computational model, it mainly has three parts: membrane structure, multisets of objects and evolution rules [27, 28, 29]. A P systems of n dimensional can be formalized as follows [7]:

$\Pi = (O, \ T, H, \ \mu, \ w_1,\ldots\ldots,w_m, R_1,\ldots\ldots, \ R_m,i_0)$

Where

(1) O is an alphabet of objects.

(2) $T \subseteq O$ (the output alphabet).

(3) H is the alphabet of membrane labels;

(4) μ is a membrane structure with degree m

(5) $w_1,\ldots\ldots,w_m \in O^*$ are the multisets of objects associated with the m regions of μ;

(6) R_i, $1 \leq i \leq m$, are the finite sets of multiset rewriting rules associated with the m regions of μ;

(7) i_0 is the output membrane.

Membrane structure contains a series of layered membrane, the outermost membrane is skin membrane, it divided the system into two parts, the environment part and the inner part, the interior of a membrane is the region of it. There may be non-elementary membranes or elementary membranes in the inner part, if any membrane was contained in a membrane, it is elementary membrane, or it is non-elementary membrane, in other words, the region of the non-elementary membrane can nest any non-elementary or elementary membrane, but the elementary can't. *Figure 1* shows a P system.
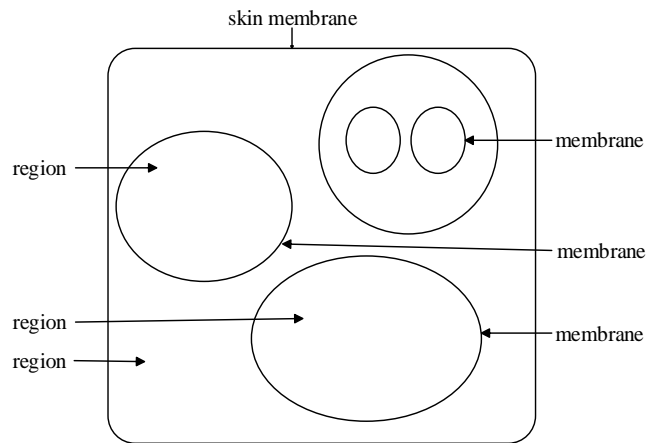
Figure 1. The Structure of a P system

4. The Improved PSO-GA Hybrid Algorithm Based on P Systems for K-means Clustering

As evolutionary algorithms, the genetics algorithm and the particle swarm optimization have something in common, one is that their systems are initialized in a set of randomly generated potential solutions, and next perform the search for the optimal one iteratively. Whereas the GA possess the crossover and mutation processes, combined with the PSO could improve the ability of finding the global optimum. As a model of computation, P systems has significant feature of maximal parallelism and non-deterministic. The objects of the systems also accomplish exchange process according to some certain communication rules.

The main idea of the improved PSO-GA Hybrid Algorithm (PGHA) [27] based on P systems (PGHAPS) is combine the essence of particle swarm optimization and genetics algorithm with membrane computing. This two evolutionary algorithms improve the optimization capacity, and the membrane computing contained in this algorithm is used to overcome the premature convergence feature of the PSO-GA hybrid algorithm. In contrast with the particle swarm optimization and the genetics algorithm, the maximum parallelism of the P system makes this hybrid membrane evolutionary algorithm owns better global optimization capacity. The proposed membrane evolutionary algorithm is used to combine with the K-means by optimizing the initial cluster centers of it. Generally, K-means clustering is easy to implement and convergence fast, but it is easily trapped into the local optimum and hard to find the global optimum, so comparing with the other evolutionary algorithms, the hybrid K-means has a fast convergence, in contrast with the standard K-means, the hybrid K-means could gain better clustering quality.

In the PGHAPS algorithm, the individuals generated stochastic are put in each membrane of the P systems randomly and make sure every membrane has at least one individual. The individuals in one membrane were considered as an independent population, and they were broadly split into two groups, one group executes the PSO, while the other implements the GA simultaneously. The population in every membrane reaction respectively, and then obtained the optimal value through communication between membranes. If the global optimum is founded, the reaction is suspend, or, after T times iteration, the two parts exchange some individuals of them and proceed with the repercussion until satisfies the stop iteration or get the optimization results of the global. Interclass inertia is applied as fitness function to assess the unit so that to choose the better individual as the initial cluster centers of the K-means algorithm. *Figure 2* shows the structure of the PGHAPS.
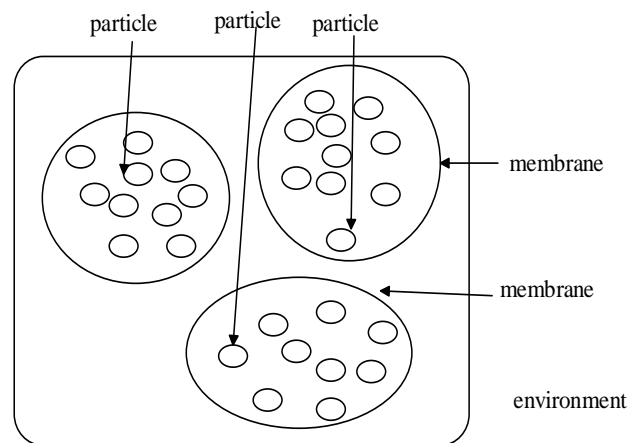
Figure 2. The Structure of the PGHAPS

The procedure of PGHAPS-K-means is as follows:

Step 1 Set M numbers of membranes except the skin membrane. h displays the h-th membrane, $1 \leq h \leq M$.

Step 2 Set N numbers of individuals in one membrane at randomly, in this initial process, make sure each membrane has at least one individual, which is $N \geq 1$. N_h stands for N numbers of particles in the h-th membrane.

Step 3 Stochastic divide the N individuals in one membrane into two parts. One of the two executives PSO and the other carry out the GA simultaneously. If the best position is founded, the reaction suspends, or, after T times iteration, the two parts exchange the optimum of them and proceed with the repercussion until meet termination condition. The optimum is determined by the fitness value comparison according to the objective function. The messages exchanged between membranes operated by communication rules of the P system.

Step 4 Execute K-means clustering according to the result attained by the PGHAPS until satisfies the stop criterion.

5. Experiments and Results

5.1. Data Sets

To evaluate the validity of the proposed algorithm, two artificial data sets named Art1, Art2 and three real life data sets named Iris, Crude Oil, Vowel are used. All data sets except Art1 and Art2 are available at ftp://ftp.ics.uci.edu/pub/machine-learning-databases/. *Table 1* summarizes the characteristics of these data sets.

Table 1. Characteristics of the Data Sets Considered

Name of Date set	NO. of classes	NO. of features	Size of data set (size of classes in parentheses)
Art1	3	2	300 (100, 100, 100)
Art2	3	3	210 (70, 70, 70)
Iris	3	4	150 (50, 50, 50)
Crude oil	3	5	56 (7, 11, 38)
Vowel	6	3	871 (72, 89, 172, 151, 207, 180)

(1) Art1: This is a two-featured problem with three unique classes. A total of 300 patterns were drawn from three independent bivariate normal distributions, where classes

were distributed according to N2 ($\mu = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \end{pmatrix}$, $\Sigma = \begin{bmatrix} 0.4 & 0.04 \\ 0.04 & 0.4 \end{bmatrix}$) i=1,2,3, $\mu_{11}= \mu_{12}=0$, $\mu_{21}= \mu_{32}=0$, $\mu_{31}= \mu_{22}=10$, μ and Σ being the covariance matrix. The data set is illustrated in *Figure 3*.

(2) Art2: This is an artificial data set. It is a three-featured problem with three classes and 300 patterns, where every feature of the classes is distributed according to Class1~Uniform (15, 20), Class2 ~Uniform (20, 25), Class3~Uniform (25, 30). The data set is illustrated in *Figure 4*.

(3) Fisher's iris data set consists of three different species of iris flowers: Iris setosa, Iris virginica, and Iris versicolour. For each species, 50 samples with four features each (sepal length, sepal width, petal length, and petal width) were collected.

(4) The Crude Oil data set consists of 56 objects characterized the five features such as vanadium, iron, beryllium, saturated hydrocarbon, and aromatic hydrocarbon content. There are three crude-oil samples from three zones of sandstone (Wilhelm has 7 objects, Sub-Mulnia has 11 objects, and Upper has 38 objects).

(5) The Vowel data set consists of 871 Indian Telugu vowel sounds. The data set has three features corresponding to the first, second, and third vowel frequencies and six overlapping classes, d (72 objects), a (89 objects), i (172 objects), u (151 objects), e (207 objects), o (180 objects).
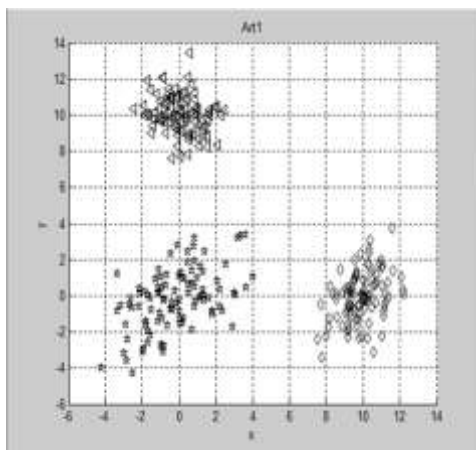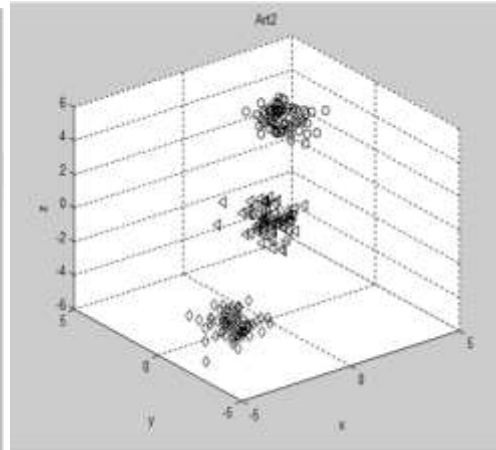

Figure 3. Art 1   **Figure 4. Art 2**

5.2. Parameter Settings

Experiments were conducted in order to compare the performance of PGHAPS-K-means with K-means and PSO-K-means. The values of the parameter in these algorithms are shown in *Table 2*.

Table 2. Parameter Setting

ω	c_1	c_2	P_c	P_m	G_m	M	N	M
0.9~0.4	1.2	1.2	0.8	0.05	200	4	100	200

ω is inertia weight, c_1 and c_2 are learning factor, The paper adopts single-point crossover, and the crossover rate is P_c, P_m is mutation ratio, G_m is evolution generation, m is the number of the elementary membrane, N represents the number of the individuals, M is the maximum iteration.

5.3. Results

In addition to the proposed algorithms, PSO-K-means and original K-means algorithm are also tested, 30 runs were executed independently when applying one of the tested algorithms. The termination condition is the predetermined number of function evaluation to be reached. The performance of the three algorithms is evaluated and compared using the following two criteria:

The sum of intra-cluster distances. The distance between each data vector in a cluster and the centroid of that cluster, as defined in equation (1). The smaller the sum of the distance means the better performance of clustering

Error rate (ER). It is defined as the number of misplaced points over the total number of points in the dataset, as shown in equation (5):

$$\text{Error rate} = (Q/e) \times 100\% \tag{10}$$

Where e is the total number of data points and Q is the total number of misplaced points. The error rate is defined as 0 when the value is smaller than 10-6. *Table 3* shows the intra-cluster distances compared.

Table 3. Comparison of Intra-Cluster Distances for K-Means, PSO-K-Means and PGHAPS-K-Means

Data set	Criteria	K-means	PSO-K-means	PGHAPS-K-means
Art 1	Average	426.05	418.47	416.45
	Std	(15.71)	(0.59)	(0.00)
	best	420.78	418.42	416.45
Art 2	Average	442.41	434.46	433.36
	Std	(18.47)	(9.04)	(0.03)
	best	434.76	433.74	432.03
Iris	Average	106.05	96.76	96.06
	Std	(14.11)	(0.07)	(0.03)
	best	97.33	96.66	96.06
Crude Oil	Average	287.36	277.77	276.07
	Std	(25.41)	(0.33)	(0.03)
	best	279.20	277.45	277.64
Vowel	Average	159242.87	149375.70	149373.60
	Std	(916)	(155.56)	(150.01)
	best	149422.26	149206.10	149203.10

Table 3 illustrates the average, standard deviation and best intra-cluster distances from five datasets. As can be seen clearly, PGHAPS-K-means outperform PSO-K-means and K-means in these datasets and obtains the optimal intra-cluster distances in each run. For Art1, the average of the fitness for PGHAPS-K-means is identical to the best distance, but the standard deviation of the fitness for this algorithm is 0, smaller than that of PSO-K-means and significantly smaller than that of K-means. For the three real life data sets, PGHAPS-K-means performs well in the average distance, the best distance and the best intra-cluster distances. In these three cluster methods, K-means has the largest standard deviation, performs worse than PSO-K-means, the PGHAPS-K-means does the best performance. It shows that the PGHAPS-K-means outperform the PSO-K-means and K-means by better capacity of finding the global optima.

The standard K-means algorithm is widely used owning to its simple concept and easy operation. With randomly choosing k data points as initial cluster centers, the K-means algorithm is easy trapped into the local optimum and the clustering results is unstable. In particle swarm optimization, the particles moves according to some certain rules, when combining with genetics algorithm and P systems, the hybrid algorithm will has a better

optimizing capacity and he error rate will decline in the meantime. *Table 4* shows the average error rates, standard deviations, and the best solution of the error rates from the 30 simulations.

Table 4. Comparison of Error Rates for K-means, PSO-K-means and PGHAPS-K-means (%)

Data set	Criteria	K-means	PSO-K-means	PGHAPS-K-means
Art 1	Average	19.23	17.85	0.00
	Std	(0.69)	(0.00)	(0.00)
	best	13.45	12.98	0.00
Art 2	Average	15.51	14.34	0.00
	Std	(6.40)	(0.00)	(0.00)
	best	8.93	8.20	0.00
Iris	Average	17.8	10.20	10.20
	Std	(10.72)	(0.21)	(0.16)
	best	10.67	10.00	10.00
Crude Oil	Average	24.46	24.29	23.42
	Std	(1.21)	(0.72)	(0.40)
	best	23.21	23.21	23.21
Vowel	Average	44.26	42.24	42.17
	Std	(2.15)	(0.95)	(0.83)
	best	42.02	40.64	40.64

As is indicated in *Table 4*, for Art1 and Art2, the average, the standard deviation, and the best solution of the error rates for PGHAPS-K-means are 0.00%, which are superior PSO-K-means and K-means. For Iris data set, the average and the best solution of PSO-K-means are equal to those of PGHAPS-K-means but the standard deviations is outperform the other two by 0.05 and 10.56. Generally, PGHAPS-K-means does not compare favorably with PSO-K-means and K-means for Iris. For Crude Oil, the proposed method is superior to the other two in average and standard by 0.87, 0.94, 0.32 and 0.81. For Vowel. The value is 0.07, 0.9, 0.12 and 1.32, such that the hybrid K-means clustering has a better clustering quality and a stable clustering results. It should be noted that the error rate is not proportional to the intra-distance. The actual data distribution was not regular and a smaller intra-cluster distance does not mean a lower error rate, as there is no absolute correlation between them. Conclusions could be arrived that, in error rate, the proposed clustering method does better than the two being compared on the whole.

6. Conclusions

The proposed algorithm is characterized by employing the formal framework of P systems, the evolutionary principles of particle swarm optimization and improved variable population-size genetics algorithm in each membrane (PGHAPS). The maximal parallelism and non-determinism of P systems also do help in improving global search capacity. This hybrid evolutionary algorithm is then used to improve the K-means by optimizing the initial cluster centers. By comparing the clustering performance of the hybrid K-means, the K-means and the PSO-K-means based on five tested data sets, proved the better clustering quality and the stable clustering results of the hybrid K-means. In this paper, simulation experiment is tested on data sets within dimension less than five. The following research will focus on dealing with high-dimensional data sets.

Acknowledgments

References

[1] J. Han and Kamber M, "Data Mining: Concepts and Techniques", Morgan Kaufmann, San Francisco, **(2001)**.

[2] R. Maitra, A. Peterson, A. Ghosh, "A systematic evaluation of different methods for initializing the K-means clustering algorithm", IEEE Trans. Knowl. Data Eng., **(2010)**.

[3] A. Harbi and S. R. Smith, "Adapting K-means for supervised clustering", Applied Intelligence, vol. 24, no. 3, **(2006)**, pp. 219–226.

[4] J. Kennedy and R. Eberhart, "Particle swarm optimization, International Conference on Neural IEEE Networks", **(1995)**, pp. 1942–1948.

[5] M. Couceiro and P. Ghamisi, "Particle Swarm Optimization", Springer Briefs in Applied Sciences and Technology, **(2015)**, pp. 1-10.

[6] T. Krinkand and M. Løvbjerg, "The Lifecycle Model: Combining Particle Swarm Optimization", Genetic Algorithms and Hill Climbers, **(2002)**, pp. 621-630.

[7] G. Pǎun, Rozenberg and A. Salomaa, "The Oxford Handbook of Membrane Computing", Oxford University Press, New York **(2010)**.

[8] G. Pǎun, G. Rozenberg and A. Salomaa, "Membrane Computing", Oxford University Press, New York, **(2010)**.

[9] D. E.Goldberg, Genetic Algorithms in Search, Optimization & Machine Learning, Addison-Wesley Professional, **(1989)**.

[10] P. W. Tsao, "Combining PSO cluster and nonlinear mapping algorithm to perform clustering performance analysis: take the enterprise financial alarming as example", Quality & Quantity, vol. 45, no. 6, **(2011)**, pp. 1291-1302.

[11] M. A. Rahman and M. Z. Islam, "A hybrid clustering technique combining a novel genetic algorithm with K-Means", Knowledge-based Systems, vol. 71, **(2014)**, pp. 345-365.

[12] P. S. Shelokar, V. K. Jayaraman and B. D. Kulkarni, "An ant colony approach for clustering", Analytica Chimica Acta, vol. 509, no. 2, **(2004)**, pp. 187–195.

[13] J. Xiao, Y. P. Yan, J. Zhang and Y. Tang, "A quantum inspired genetic algorithm for k-means clustering", Expert Systems with Applications, vol. 37, **(2010)**, pp. 4966–4973.

[14] A. Pavel, O. Arsene and C. Buiu, "Enzymatic numerical P systems-a new class of membrane computing systems", IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2010) Liverpool, **(2010)**, pp. 1331–1336.

[15] S. Yang and N. Wang, "A novel P systems based optimization algorithm for parameter estimation of proton exchange membrane fuel cell model", International Journal Hydrogen Energy, vol. 37, **(2012)**, pp. 8465–8476.

[16] H. Peng, J. Wang, M. J. Pérez-Jiménez and P. Shi, "A novel image thresholding method based on membrane computing and fuzzy entropy", Journal of Intelligent & Fuzzy Systems, vol. 24, no. 2, **(2013)**, pp. 229–237.

[17] M. Y. Cheng, K. Y. Huang and H. M. Chen, "K-means particle swarm optimization with embedded chaotic search for solving multidimensional problems", Applied Mathematics and Computation, vol. 219, **(2012)**, pp. 3091-3099.

[18] S. Kalyani and S. Swarup, "Particle swarm optimization based K-means clustering approach for security assessment in power systems", Expert Systems with Applications, vol. 38, no. 9, **(2011)**, pp.10839-10846.

[19] J. C. Feng and C. Y. Wang, "A self-generating fuzzy system with ant and particle swarm cooperative optimization", Expert Systems with Applications, vol. 36, **(2009)**, pp. 5362–5370.

[20] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization", IEEE International Congress on Evolutionary Computation, vol. 3, **(1999)**, pp. 101–106.

[21] S. Naka, T. Genji, T. Yura and Y. Fukuyama, "A hybrid particle swarm optimization for distribution state estimation", IEEE Transactions on Power Systems, vol.18, no. 1, **(2003)**, pp. 60-68.

[22] P. J. Angeline, "Using Selection to Improve Particle Swarm Optimization", IEEE Congress on Evolutionary Computation, **(1998)**, pp. 84-89.

[23] J. Zhang and Z. Zhang, "Improved Particle Swarm Optimization Algorithm and Its Application to Global Optimization for Complex Function", Intelligent and Soft Computing, vol. 143, **(2012)**, pp. 683-690.

[24] C. F. Juang, "A Hybrid of Genetic Algorithm and Particle Swarm Optimization for Recurrent Network Design. IEEE Transactions on Systems", vol. 34, no. 2, **(2004)**, pp. 997-1006.

[25] T. Santhanam and M. S. Padmavathi, "Application of K-Means and Genetic Algorithms for Dimension Reduction by Integrating SVM for Diabetes Diagnosis", Procedia Computer Science, vol. 47, **(2015)**, pp. 76 – 83.

[26] X. Shi, Y. C. Liang, H. P. Lee, C. Lu and L. M. Wang, "An improved GA and a novel PSO-GA-based hybrid algorithm, Information Processing Letters, vol. 93, **(2005)**, pp. 255–261.

[27] G. Păun, "Computing with membranes", Journal of Computer and System Sciences, vol. 1, no. 61, **(2000)**, pp. 108-143.

[28] C. I. Vasile, A. B. Pavel, I. Dumitrache and G. Păun, "On the power of enzymatic numerical P systems, Acta Informatica, vol. 49, no. 6, **(2012)** pp. 395-412.

[29] Y. Zhao, X Liu and J. Qu, "The improved hierarchical clustering algorithm by a P system with active membranes", WSEAS Transactions on computers, vol. 12, **(2013)**.

Authors

Chenggong Qiu, he is a Ph. D candiadate in Management Science and Engineering from Shandong Normal University. Now he is the lecture teacher in Shandong Normal University. His research interest is mainly in the areas of membrane computing, computational intelligence, nonlinear analysis.

Laisheng Xiang, he received his Ph.D degree in Management Science and Engineering from Northeastern University, postgraduate tutor, professor of Shandong normal university. His research activities are focused on circular economy, and university reform.

Xiyu Liu, professor and doctoral supervisor in Shandong Normal University. His research activities are focused on Information management and e-commerce, computational intelligence, nonlinear analysis. He has published many research papers in scholarly journals in the above areas.

Huaning Yan, she is a Master candidate of Shandong Normal University. Her research interest is mainly in the areas of membrane computing, clustering analysis and computational intelligence.