409

A Novel Membrane Clustering Algorithm Based on Tissue-like P system

Huaning Yan, Laisheng Xiang, Xiyu Liu, Jie Xue School of Management Science and Engineering Shandong Normal University, Jinan, China *Corresponding author, e-mail: sdxyliu@163.com

Abstract

Clustering is a process of partitioning data points into different clusters due to their similarity, as a powerful technique of data mining, clustering is widely used in many fields. Membrane computing is a computing model abstracting from the biological area, these computing systems are proved to be so powerful that they are equivalent with Turing machines. In this paper, a modified inversion particle swarm optimization was proposed, this method and the mutational mechanism of genetics algorithm were used to combine with the tissue-like P system, through these evolutionary algorithms and the P system, the idea of a novel membrane clustering algorithm could come true. Experiments were tested on six data sets, by comparing the clustering quality with the GA-K-means, PSO-K-means and K-means proved the superiority of our method.

Keywords: membrane clustering algorithm, P system, particle swarm optimization, evolutionary algorithm

Copyright © 2016 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Membrane computing is a recent domain of natural computing started by Gh. Păun in 1998 and is known as P systems or membrane systems [1]. It is a parallel distributed computing model inspired from the functionality and structure of living cells and the intersection of them in the tissue, organ and neural network. There are three main classes of P systems investigated: cell-like P systems, tissue-like P systems and neural-like P systems [2], [3, 4]. These computing systems are proved to be so powerful that they are in some ways equivalent with Turing machines [5], the powerful global search capacity of membrane computing is also proved [6].

Cluster analysis also called clustering, is a typical example of unsupervised learning, it aims at partitioning a set of data objects into subsets, each subset called a cluster, such that objects in a cluster are similar to each other, yet dissimilar to objects in other clusters [7, 8]. Clustering analysis has been widely used in various fields, such as image processing, data analysis, Web searching, market analysis etc. As one of the most classical clustering method based on partitioning, K-means algorithm is an unsupervised clustering algorithm for classifying data points into k distinct groups. It is the most classical clustering method due to its simple principle and easily application [9]. However, the final clustering results of it depend heavily on the initial random selection such that the K-means method is not guaranteed to converge to the global optimum and often trapped into the local optimum [10].

As a novel computing model inspired by biology, the P system has attracted more and more people's attention. Due to the parallel distributed computing ability of the P system, a number of papers have appeared in which P system are hybridized with other evolutionary algorithms like Multi-level thresholding methods [11], Particle Swarm Optimization [12, 13], Genetic Algorithms [14], Ant Colony Optimization [15], Quantum Inspired EAs [16], etc. These modifications obviously made many improvements on the evolutionary algorithm and promoted the development of the P system research.

In order to attain a better clustering result, more improvements have made on current clustering methods. In this paper, a modified inversion particle swarm optimization was proposed, this method and the mutational mechanism of genetics algorithm were used to combine with the tissue-like P system, Through the special designed tissue-like P system and the evolution mechanism and communication mechanism, the idea of the novel membrane clustering algorithm could realized. The rest of the paper is organized as follows. Section 2

describes briefly partition-based cluster algorithm. Section 3 introduces tissue-like P system. Section 4 presents a variant of the PSO algorithm and the proposed membrane clustering algorithm. Section5 illustrates experimental results. Finally, Section 6 makes conclusions.

2. Partition-Based Cluster Algorithm

Cluster analysis is a process of partition data points into different clusters due to their similarity. As a powerful technique of data mining, clustering is widely used in many fields, such as business intelligence, Web searching, biological, safety and so on. Major basic clustering methods can be divided into the following categories: partitioning method, hierarchical method, density-based method, grid-based method [7]. Suppose that dividing the data points into k clusters based on partition-based clustering algorithm, the k cluster centers, z_1 , z_2 , ..., z_k , and the corresponding clustering partitions are C_1, C_2, \ldots, C_k respectively. The data point xi belongs to the cluster p if the distance meets:

$$\Box x_i - z_p \Box \Box x_i - z_q \Box, p, q = 1, 2, \dots, \quad \mathbf{k}, \mathbf{p} \neq q$$
(1)

Once clustering partitions are formatted, new cluster centers are computed by the means of the points in the corresponding cluster using

$$z_{j} = \frac{1}{n_{j}} \sum_{\forall x_{p} \in c_{j}} x_{p}$$
(2)

Where nj is the number of data points in cluster j and Cj is the clustering partition of cluster j. In this partition-based clustering method, the clustering metric M is determined as follows:

$$M(C_{1}, C_{2}, ..., C_{k}) = \sum_{i=1}^{k} \sum_{x_{j} \in c_{i}} \Box x_{j} - z_{i} \Box$$
(3)

Obviously, smaller M value stands for better clustering results, such that this clustering problem can be considered as an optimization process of finding the smaller M.

3. The Tissue-Like P System

Membrane computing is a recent domain of natural computing started by Gh. Păun in 1998. It is also known as membrane computing or membrane systems. In recent years, many different models of P systems have been proposed, such as cell-like P systems, tissue-like P system, and spiking neural P system [17-21]. The obtained computing systems have proved to be so powerful that it is equivalent with Turing machines [22]. The P systems are a class of distributed parallel computing devices of a biochemical type [23], membrane computing now is a hot cross-discipline topic which involves computer science, mathematics, biology and artificial intelligence, etc. One P system mainly consists of three parts: membrane structure, multisets of objects and evolution rules [1]. In tissue-like P system, membranes or cells are placed as the nodes of a graph. The net of nodes deals with symbols and communicates symbols along channels specified in advance. The communication among cells is based on symport/antiport rules [24]. Symport rules move objects across a membrane together in one direction, whereas antiport rules move objects across a membrane in opposite directions [25]. The tissue-like P system which contains m elementary membranes is shown in Figure 1 and described as follows:

$$\prod = (\omega_1, \dots, \omega_m, R_1, \dots, R_m, R', i_0)$$
(4)

(1) ω_i (1 ≤ i ≤ m) is a finite set of objects in elementary membrane i;

(2) R_i (1 ≤ i ≤ m) is a finite set of evolution rules in elementary membrane i;

(3) R' is a finite set of communication rules with the two following forms: antiport rule: (i, Z/Z', j), i, j = 1, 2, ..., m, $i \neq j$. The rule is used to communicate the objects between an elementary

membrane and its neighbor; symport rule: $(i, Z/\lambda, 0), i = 1, 2, ..., m$. The rule is used to communicate the best objects between membranes and the membrane with the environment. (4) i_0 indicates the output region of the system.

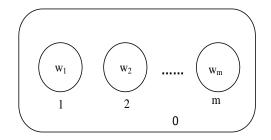


Figure 1. The designed tissue-like P system

The structure and the special mechanism of this tissue-like P system are the key point of the membrane clustering algorithm. The objects of the P system are represent as the possible clustering centers, such that the dimension of each object is $k \ x \ d$, in which the cluster number is k and the dimension of the data points is d. In this tissue-like P systems, each elementary membrane independently work in a maximally parallel way. The computation of a tissue-like P system is a sequence of computing steps, which starts from the m elementary membranes containing w_1, \ldots, w_m . At each calculation step, one or more rules are applied utilizing the multisets of objects. When the computation met the stop criterion; the final results are showed in the output region.

4. The Proposed Membrane Clustering Algorithm

4.1. An Improved Inversion Particle Swarm Optimization

Particle swarm optimization is a population-based stochastic optimal algorithm proposed by Kennedy and Eberhart in 1995. The basic idea of this algorithm simulates bird flocking or fish schooling behavior to build a self-evolving system [18]. As a representative of the swarm intelligence algorithms, particle swarm optimization algorithm is used to solve continuous optimization problems originally [26]. In this algorithm, each particle was considered as a potential solution to an optimal question, these particles flied at a specific speed and then adjust speed with their own and the other particles' flying experience, every particle has an objective function to determine its fitness which was used for evaluate itself, during this iteration course, the particle flied toward the best position, and the optimization problem tends to obtain its optimal solution. The finding process of local best and global best that are computed in every iteration gives the potential to this method to reach the most optimal solution[27]. Particle swarm optimization also has some drawbacks, for example, it is easy trap into the local optimum and be absence of population diversity. Some certain modifications have been made in basic PSO algorithm, for instance, inertia weight, reverse searching, parameter modification and so forth [28]. In this paper, we propose a modified inversion particle swarm optimization (MIPSO) based on the idea of avoiding the worst position in order to improve the exploration capacity and the diversity of the whole swarm [29]. Suppose that N particles exist in D-dimensional searching space, the particle updates its velocity and position in our method as follows:

$$w_{id}^{k+1} = \omega_{id}^{k} - c_{1} \operatorname{rand}((p_{iw} - x_{id}^{k}) - c_{2} \operatorname{rand}((p_{g_{w}} - x_{id}^{k}))$$
(5)

$$x_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1}$$
(6)

where d = 1,2, ..., D; i = 1,2, ..., N, v^{k+1}_{id} is the i-th particle's new velocity at the kth iteration step, the range of Vid is [-Vmax,Vmax], such per-set value prevents the particle from

flying out of the search space; k is the iteration number; xk+1id is the i-th particle's position of the k+1 times, based on its previous position and new velocity at the last iteration; ω is inertia weight; piw is the worst fitness value of the particle itself achieved so far, and pgd is the worst global fitness the whole swarm attained so far; c₁, c₂ are learning factors; rand() are random numbers uniformly distributed in the range [0, 1]. The performance of each particle is evaluated according to a predefined fitness function, which is usually proportional to the cost function associated with the problem. When finding the optimum or meeting the maximal iteration numbers, the searching process halted.

In this method, we consider the particle searching for the best condition by combining the worst flying experience of itself and the worst flying experience of the whole population, the particle adjusts its velocity by flying towards the contrary direction of the worst position, the fitness value is adopted to evaluate the particle. Learning from the worst experience and run away from it is the basis idea of this improved algorithm.

4.2. The Evolution Rules and Communication Rules of the Tissue-Like P System

In each elementary membrane of the tissue-like P system, n number objects are contained. Every object represents the k possible cluster centers, evolution rules are applied to evolve the objects, and the evolution mechanism is designed based on particle swarm optimization, its improved version MIPSO and mutation rules of generation algorithm.

In the proposed tissue-like P system, each elementary owns the same n number objects, during the evolution, these objects are randomly divided into two parts roughly, one of the parts execute standard particle swarm optimization while the other carry out the MIPSO, the two optimize the objects through by finding the best position and avoid the worst position. In this work, mutation mechanism in the binary coding generation algorithm is also used to coevolve the objects, and it makes mutation according to the possible pm, if the value of a mutation point at dimension j is v, after mutation the value becomes that

$$v' = \begin{cases} v \pm \delta v, v \neq 0\\ v \pm \delta, v = 0 \end{cases}$$
(7)

In the above formula, the signs "+" or "-" occur with equal probability, and δ is real number generated with uniform distribution in the range [0, 1]. The optimization process halted until it meets the maximum iteration criteria.

The communication rules in the designed P system mainly have two types: antiport rule: (*i*, *Z*/*Z*', *j*), *i*, *j* = 1, 2, . . . , m, $i \neq j$, including m is the number of elementary membranes, this rule is used to communicate the objects between an elementary membrane and its two neighbor; symport rule: (*i*, *Z*/ λ , 0), *i* = 1, 2, . . . , m, this rule is used to communicate the objects between membrane and the environment, of which m is the number of the elementary membranes. The label 0 stands for the output region namely environment. The role of the communication mechanism is to communicate the optimal object of each membrane and eventually attain the global optimum which stores in the output region. The reference point is the fitness objection that is the M value,

4.3. Description of the Proposed Membrane Clustering Algorithm

In this paper, a new membrane clustering algorithm based on partitioning clustering is proposed through utilizing the rules of evolutionary algorithm, communication rules and the special framework within the distributed parallel computation framework provided by tissue-like P system. The objects of each cell are represented as the possible cluster centers. Particle swarm optimization algorithm and a modified reverse particle swarm optimization algorithm are applied as evolution mechanism to evolve the objects. Mutation mechanism of the generation algorithm is also imported as coevolution rules in order to improve the diversity of the population. Communication rules between cells or between cells and environment is used to exchange the object and arrive to attain the optimum. This new membrane clustering algorithm can automatic searching the optimum cluster centers and arrive at better cluster results, the clustering algorithm is described as follows: IJEECS

Step 1. Generate n initial objects (the possible cluster centers) for each of the m elementary membrane, then distinguish clustering regions according to the existing formula and calculate the new cluster centers.

Step 2. Initial a group of individuals randomly, divide the individuals into two portions with the equal scale, initial the velocity and the position, set the mutation probability pm, fitness function f.

Step 3. Execute the particle swarm optimization and the modified inversion particle swarm optimization, mutation rules are applied simultaneously.

Step 4. Communication rules are used to exchange the optimum between the elementary membranes and elementary membrane with environment.

Step 5. Repeat these steps until attain the maximum iteration criteria, and the final global optimum is showed in the environment.

5. Experiments and Results

The proposed membrane clustering algorithm is assessed on six data sets and compared with K-means algorithm, PSO-K-means algorithm and GA-K-means algorithm [30, 31]. For illustrating the results of the proposed algorithm exactly, 30 runs were executed when applying one of the tested algorithms. These six data sets including two artificial data sets named Art1, Art2 and four real life data sets named Iris, Cancer, Vowel, Wine are used. All data sets except Art1 and Art2 are available at ftp://ftp.ics.uci.edu/pub/machine-learningdatabases/. Table 1 summarizes the characteristics of these data sets. Data sets Art 1, Art 2 is illustrate in Figure 2, Figure 3.

Tab	Table 1. Characteristics of the data sets considered				
Name of	NO. of	NO. of	Size of data set		
Date set	classes	features	(size of classes in parentheses)		
Art1	3	3	900 (300, 300, 300)		
Art2	2	3	900 (450, 450)		
Iris	3	4	150 (50, 50, 50)		
Cancer	2	9	683 (444, 239)		
Vowel	6	3	871 (72, 89. 172, 151, 207, 180)		
Wine	6	13	178 (59, 71, 48)		

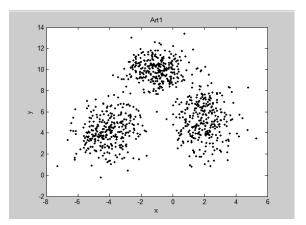


Figure 2. Art 1

413

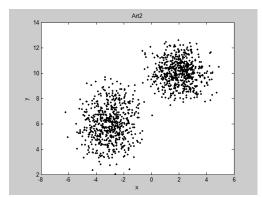


Figure 3. Art 2

The values of the parameter in these tested algorithms are shown in Table 2.

	Table 2. F	Parame	eter set	tting	
Parameter	ω	C ₁	C ₂	Pm	Μ
Value	0.9~0.4	1.2	1.2	0.05	1000

 ω is the inertia weight, the value is between 0.4 and 0.9, small value improves local convergence capacity and the large value improves global convergence capacity; c₁ and c₂ are learning factor; pm is mutation probability; M is the maximum iteration. In order to investigate the effect of elementary membrane numbers, four different tissue-like P system with 4, 8, 16, 20 elementary membranes are designed. In order to overcome the effects of accidental factors, we choose the average value of the every whole 30 times runs. The M values of these P systems are shown in Table 3.

Table 3. Comparison of the M values of different-elementary membrane tissue-like P systems

Data set	4	8	16	20
Art 1	628.3298	627.2565	624.4532	634.0342
Art 2	546.7652	545.3545	539.2868	539.7662
Iris	99.9238	98.7596	98.6324	98.9328
Cancer	3253.3342	3052.5372	3249.6575	3250.0442
Vowel	149315.4096	149311.8741	149309.4863	149309.8223
Wine	16313.6521	16306.4536	16293.5662	16293.7976

Table 3 indicates that tissue-like P system with 16 elementary membranes behaves better than the others with the smaller M value in these six test data sets. For the true life data set of Cancer, the performance of the 16 elementary membrane P system is 3249.6575, superior to the other 3251.3342, 3052.5372 and 3250.0442 obviously.

Generally, more particles in the swarm could enhancing the searching range and may leads better optimization results. To inquiry the appropriate number of the particles, different scale of the swarms are designed for the experiment. Table 4 shows the diverse results of the different particle scales.

Table 4. Comparison of the M values of the different particle scale swarms

Data set	20	50	100	200	400
Art 1	629.2378	624.9662	624.5342	624.4532	624.4532
Art 2	550.4574	546.6753	539.7662	539.2868	539.2868
Iris	104.2438	101.7456	99.9458	98.6324	98.6324
Cancer	3257.5664	3053.8698	3252.6572	3249.6575	3249.6575
Vowel	149320.4033	149314.6541	149310.6222	149309.4863	149309.4863
Wine	16317.3496	16311.4581	16295.2803	16293.5662	16293.5662

Table 4 shows that when the number of particles attains 200, the clustering quality will be better than the small-scale swarms. For the Wine data set, the M value is 16293.5662 with the particle scale is 200, obviously smaller than the smaller scale with the value 16317.3496, 16311.4581, 16295.2803. Too many particles may does no help in improving the clustering quality, the clustering results have no change when the number rose from 200 to 400, and thus the appropriate number could define 200 in our membrane clustering method.

Table 5. Comparison of the M value of the tested algorithms on the six data set	s (%)

GA 649.9533 542.0232	PSO 639.2434 541.6982	K-means 649.1573 552.3472	MCA 623.3662 539.0848
542.0232			
	541.6982	552.3472	530 0848
			333.0040
98.9734	98.0531	106.1347	97.0324
3249.8713	3048.2311	3257.6725	3248.5472
149310.0908	149309.7492	149336.9646	149307.4223
16293.3142	16292.0419	16318.4763	16290.7976
	149310.0908 16293.3142		

According to the results of the chart in Table 5, we could conclude that, in most of runs of these algorithms, the proposed membrane clustering method is superior to other tested algorithm. For Art 1, the value of the MCA is 623.3662, much less than the PSO which is 639.2334, the GA which is 649.9533 and the K-means algorithm which is 649.1573. The M value on the Vowel is 149307.4223, smaller than the K-means, GA-K-means and PSO-K-means with 149336.9645, 149309.7492 and 149310.0908, respectively. 30 times runs of each tested algorithm lows the contingency of the experiment and enhance the persuasion of the test. Through comparing and analyzing, the better clustering quality of the proposed membrane clustering algorithm could be proved.

6. Conclusions

At present there are many novel techniques committed to improving the data clustering. This paper focuses on realizing a membrane clustering algorithm based on a designed tissue-like P system. With the structure of the tissue-like P system, particle swarm optimization and an modified inversion particle swarm optimization are applied as the evolution rules and the mutation mechanism of the generation algorithm is used as the supplement of the evolution rules and enhance the diversity of the population simultaneously, communication rules which contain antiport rules and symport rules are employed to communicate the optimum, our proposed membrane clustering algorithm behaves superior to the some other clustering algorithm, the perfect performance result of the simulation experiment prove this claim. The further work concentrate upon using more evolutionary algorithm based on membrane computing in order to improve the clustering methods and enhance the clustering quality.

Acknowledgements

Project supported by National Natural Science Foundation of China (61472231, 61170038, 61502283), Jinan City independent innovation plan project in College and Universities, China (201401202), Ministry of education of Humanities and social science research project, China (12YJA630152), Social Science Fund Project of Shandong Province, China (11CGLJ22).

References

- [1] G Pǎun, Rozenberg and A Salomaa. *The Oxford Handbook of Membrance Computing*. Oxford University Press, New York. 2010.
- [2] G Păun. Computing with membranes. *Journal of Computer and System Sciences*. 2000; 61(1): 108-143.
- [3] C Martin-Vide, J Pazos, Gh Păun, A Rodriguez-Paton. Tissue P systems. *Theoretical Computer Science*. 2003; 296(2): 295—326.

- [4] M Ionescu, Gh Păun, T Yokomori. Spiking neural P systems. *Fundamenta Informaticae*. 2006; 71(2-3): 279—308.
- [5] G Păun, Rozenberg G and Salomaa A. *Membrane Computing*. Oxford University Press, New York. 2010.
- [6] G Zhang, J Cheng, M Gheorghe and Q Meng. "A hybrid pproach based on differential evolution and tissue membrane ystems for solving constrained manufacturing parameter optimization problems". *Applied Soft Computing Journal*. 2013; 13(3): 1528–1542.
- [7] J Han, M Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco. 2001.
- [8] N Kinoshita, Y Endo. On Objective-Based Rough Hard and Fuzzy c-Means Clustering. J. of Advanced Computational Intelligence and Intelligent Informatics (JACIII). 2014; 19(3): 29-35.
- [9] ZTang, H Zhang. Improved K-means Clustering Algorithm Based on Genetic Algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2014; 12(3): 1917 - 1923.
- [10] R Maitra, A Peterson, A Ghosh. A systematic evaluation of different methods for initializing the Kmeans clustering algorithm. *IEEE Trans. Knowl. Data Eng.* 2010.
- [11] H Peng, J Wang, MJ Pérez-Jiménez, P Shi. A novel image thresholding method based on membrane computing and fuzzy entropy. *Journal of Intelligent & Fuzzy Systems*. 2013; 24(2): 229–237.
- [12] W Pan –Tsao. Combining PSO cluster and nonlinear mapping algorithm to perform clustering performance analysis: take the enterprise financial alarming as example. Quality & Quantity. 2011; 45(6): 1291-1302.
- [13] MY Cheng, KY Huang, HM Chen. K-means particle swarm optimization with embedded chaotic search for solving multidimensional problems. *Applied Mathematics and Computation*. 2012; 219: 3091-3099.
- [14] Md Anisur Rahman, Md Zahidul Islam. A hybrid clustering technique combining a novel genetic algorithm with K-Means. *Knowledge-based Systems*. 2014; 71: 345-365.
- [15] GX Zhang, JM Cheng, M Gheorghe. An approximate algorithm combining P ystems and ant colony optimization for traveling salesman problems. Proceedings of the 8th rainstorming Week on Membrane Computing, ISBN 978-84-614-2357-6, 321-340.
- [16] GX Zhang, M Gheorghe, CZ Wu. A quantum-inspired evolutionary algorithm based on p systems for a class of combinatorial optimization. *Fundamenta Informaticae*. 2008; 87: 93–116.
- [17] TO Ishdorj, A Leporati, L Pan, X Zeng, and X Zhang. "Deterministic solutions to QSAT and Q3SAT by spiking neural P systems with pre-computed resources". *Theoretical Computer Science*. 2010; 411(25): 2345–2358.
- [18] M Couceiro, P Ghamisi. Particle Swarm Optimization. Springer Briefs in Applied Sciences and Technology. 2015: 1-10.
- [19] L Pan and C Martin-Vide. "Solving multidimensional 0-1 knapsack problem by P systems with input and active membranes". *Journal of Parallel and Distributed Computing*. 2005; 65(12): 1578– 1584.
- [20] TO Ishdorj, A Leporati, P Linqiang, and J Wang. "Solving NP-Complete problems by spiking neural p systems with budding rules". in Proceedings of the 10th International Conference on Membrane Computing. 2009: 335–353.
- [21] L Pan, X Zeng, X Zhang, and Y Jiang. "Spiking neural P systems with weighted synapses". Neural Processing Letters. 2012; 35(1): 13–27.
- [22] G Păun, Rozenberg, A Salomaa. Membrane Computing. Oxford University Press, New York. 2010.
- [23] L Pan, TO Ishdorj. P systems with active membranes and separation rules. Journal of Universal Computer Science. 2004; 10(5): 630-649.
- [24] R Freund, G P^{*}aun and MJ P^{*}erez-Jim^{*}enez. Tissue-like P systems with channel-states. *Theoretical Computer Science*. 2005; 330(1): 101-116.
- [25] A Păun, G Păun. The power of communication: P systems with symport/antiport. *New Gen Comput.* 2002; 20: 295–305.
- [26] J Kennedy, R Eberhart. Particle swarm optimization. International Conference on Neural IEEE Networks. 1995: 1942–1948.
- [27] JJ Jamian, MW Mustafa, H Mokhlis, MA Baharudin. Implimentation of Evolutionary Particle Swarm Optimization in Distributed Generation Sizing. *International Journal of Electrical and Computer Engineering (IJECE)*. 2012; 2(1): 137-146.
- [28] G Zhang, F Zhou, X Huang et al. A hybrid multi-swarm particle optimization to solve constrained optimization problems. *Journal of Universal Computer Science*. 2005; 8(13): 1821-1841.
- [29] CM Yang, D Simon. A new particle swarm optimization technique. In: Proceedings of the 18th international conference on systems engineering (ISCEng'05). 2005: 164–169.
- [30] YT Kao, E Zahara and IW Kao. "A hybridized approach to data clustering". *Expert Systems with Applications*. 2008; 34(3): 1754–1762.
- [31] U Maulik and S Bandyopadhyay. Genetic algorithm based clustering technique. *Pattern Recognition*. 2000; 33(9): 1455-1465.