TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 16, No. 3, December 2015, pp. 565 ~ 573
DOI: 10.11591/telkomnika.v16i3.9141

Tissue-like P System Based DNA-GA for Clustering

Caiping Hou*[1], Xiyu Liu[2]
School of Management Science and Engineering, Shandong Normal University, Jinan, Shandong, China
*Corresponding author, e-mail: sdnuhcp@163.com[1], sdxyliu@163.com[2]

Abstract

In recent years, DNA-GA algorithm is drawing attention from scholars. The algorithm combines the DNA encoding and Genetic Algorithm, which solves the premature convergence of genetic algorithms, the weak local search capability and binary Hamming cliff problems effectively. How to design a more effective way to improve the performance of DNA-GA algorithm is more worth studying.As is known to all,the tissue-like P system can search for the optimal clustering partition with the help of its parallel computing advantage effectively. This paper is under this premise and presents DNA-GA algorithm based on tissue-like P systems (TPDNA-GA) with a loop structure of cells, which aims to combine the parallelism and the evolutionary rules of tissue-like P systems to improve performance of the DNA-GA algorithm. The objective of this paper is to use the TPDNA-GA algorithm to support clustering in order to find the best clustering center.This algorithm is of particular interest to when dealing with large and heterogeneous data sets and when being faced with an unknown number of clusters. Experimental results show that the proposed TPDNA-GA algorithm for clustering is superior or competitive to classical k-means algorithm and several evolutionary clustering algorithms.

Keyword: *DNA computing, genetic algorithm, tissue-like P system, clustering center*

1. Introduction

This research concerns the capability of the tissue-like P system computational model to provide a framework for DNA-GA algorithm. The general flow of the traditional genetic algorithm is to get the optimal solution scheme in several candidate groups. In each generation,the genetic algorithm is selected according to fitness and genetic reconstruction method to generate new candidate individuals. In the cycle of evolution genetic algorithm, the fitness function facilitates the candidate individual to evolution. It can form new individuals after continuous iterative process and these individuals have improved adaptability compared to the previous individual in general. This performance of genetic algorithm generally use the Genetic Algorithm as the body of algorithm when we will improve the algorithm.

Membrane computing (known as P systems) is a class of distributed parallel computing models, it focuses on the communication in the membranes, cells, tissues or other structures of organisms.In this process, the substance in cell membrane occur changes randomly and in parallel,such as diverting, variation and so on, which makes the algorithm have greatly parallelism, so a large number of operation can be completed at a moment. Compared to these benefits of membrane computing, the tissue-like P system is in a maximally parallel way and can find the best candidate individuals in a shorter time, which must be effective to choose the optimal clustering center.

2. A Brief Outline of Tissue-like P Systems

Tissue-like P systems were presented by Martin-Vide et *al.* in [[1]],which is another type of P system due to the structure of their membrane. Instead of considering a hierarchical structure, membranes are changed as a general a graph. They exist two biological inspirations (see [[2]]): inter cellular communication and cooperation between neurons.

The inter cellular communication is based on symport/antiport rules, which are introduced as the communication rules between cells. Symport rules means that objects

Received June 21, 2015; Revised October 6, 2015; Accepted October 28, 2015

cooperate to traverse a membrane together in the same direction, whereas in the antiport rules, objects across a membrane residing at both sides of it but in opposite directions [[3]].

Formally, a tissue-like P system with input of degree q >0 is a construct:

$\Pi = (A, w_1, ..., w_q, R_1, ..., R_q, R', I, O)$

Where:
(1) A is a finite alphabet, whose symbols are called objects;
(2) w_i $(1 \le i \le q)$ is finite set of strings over A, which represents multiset of objects present in cell i at the initial configuration;
(3) R_i $(1 \le i \le q)$ is finite set of evolution rules in cell i;
(4) R' is finite set of communication rules of the form (i,u/v,j), for $i,j \in \{0,1,2,...,q\}$, $i \ne j$, $u,v \in A$,| u | + |v| is the length of the communication rule (i,u/v, j);
(5) $I \in \{1,2,...,q\}$ is the input cell;
(6) O indicates the output region of the whole system in the environment.

A tissue-like P system of degree q>0 can be viewed as a set of q cells labeled by 1,2,...,q, each of which consists of an elementary membrane. We will use I to express the input cell and O to denote the output region. In above definition, R_i $(1 \le i \le q)$ is finite set of evolution rules in cell i, whose rule is of the form $u \rightarrow v, u, v \in A$. The application of the rule means that u will be evolved to v. In most of the existing tissue-like P systems and variants, evolution rule of the form is based on string of objects.Moreover, the q cells will be arranged as a loop topology based on the communication rules described below. The communication rule (i,u/v,j) can be applied over two cells labeled by i and j such that u is contained in cell i and v is contained in cell j. The application of this rule means that the objects of the multisets represented by u and v are interchanged between the two cells. Note that if either i = 0 or j = 0 then the objects are interchanged between a cell and the environment [[4]].

The rules of a system like described are used in the non-deterministic maximally parallel manner when there are several possibilities, as usual in the framework of membrane computing.In each step, all cells which can evolve must evolve in a maximally parallel way. That is to say, in each step each object in a membrane can only be used for one rule, but the system applies a multiset of rules which is maximal: no further rule can be added [[5]].

In a tissue-like P system of degree d,a computation is a sequence of steps which start with the cells 1,...,q containing the multisets $w_1,...,w_q$ and where, during each step, one or more rules are applied to the current multisets of symbol objects. The computation starts from the initial configuration and proceeds as defined before; only halting computations (reaching a configuration where no rule can be applied) give a result, and the result is encoded by the multiset of objects. A computation is successful if and only if it halts. When it halts, it produces a final result in output cell.

3. DNA Algorithm Based on Tissue-like P system (TPDNA-GA)

3.1. DNA Coding

The main genetic material of organisms is DNA that contains a wealth of genetic information. The basic elements of the DNA is Nucleotide. Nucleotide contains four bases: A (adenine), G (guanine), C (cytosine) and T (thymine). Where A pairs with T, and C pairs with G.The sequence of the DNA molecules can be abstracted as the string composed by the four bases. DNA encodes use the potential solution of the four base to optimization problem, which is an effective method to express individual and make it more suitable to recombine and variate. Adding DNA encoding to the genetic algorithm is more suitable for expressing complex knowledge, which has higher coding accuracy.It is easy to maintain the diversity of individuals and more easily to introduce the genetic manipulation. A key problem of impact on accuracy and convergence rate of problem solving is the length of the DNA strand. This paper use quanternary encoding in the literature [[6]] instead of $E'=\{0,1,2,3\}^1$,1 is the length of the DNA sequence. This allows to express the solution of the problem as a integer string of quaternary.

Generally, the optimization problem can be described as follows:

$$\begin{cases} \text{Min } f(x_1, x_2, ..., x_n) \\ X\,mini \le X_i \le Xmaxi,\ i=1, 2, ..., n \end{cases}$$

In the practical optimization problems, the variable x_i is an integer string with length L.[$x_{min\ i}$,$x_{max\ i}$] is the boundaries of x_i. $x_{min\ i}$ is a code string of 0 and $x_{max\ i}$ is a code string of 4^l-1. During the encoding process, this area of $x_{max\ i}-x_{min\ i}$ is translated into 4^l segments. Therefore, the accuracy of variable x_i is $(x_{max\ i}-x_{min\ i})/4^l$ and the length of each individual n-dimensional variable is L=n × l[[7]].

3.2. Fitness Function

Genetic algorithm exists such a function that plays decisive role in the evolution direction of genetic algorithm. This function is the fitness function. Fitness may be the objective function and it may also be a function of the objective function related. More over, it can also have nothing to do with the objective function. Genetic algorithm evaluates the candidate by the size of the value of the fitness function. Genetic algorithms can determine the chance of candidate solutions into the next generation of genetic based on the value of the fitness function. It is about whether the quality of candidate solutions may have a better chance of genetic evolution.Also,it is about whether the excellent characteristics of the population can be continued.

As the direction of search and evolutionary guiding the TPDNA-GA algorithm, the designing of the fitness function is very important. This paper use a fitness function as follows:

$$F(x) = 1/(1+f(x))$$

Where, f(x) is the objective function of the optimization problem of minimizing. After this treatment, the range of fitness function values is [0,1].

After encoding n individuals of the initial population into quaternary sequence,we sort the fitness values by the fitness function value. Reserve the minimum value of individual fitness (the worst individual) and the maximum individual (the best individual)(a total of two individuals) as limit individuals. The remaining (N-2) individuals conduct TP processing.Reserving the individual of the worst fitness value is in order to maintain the diversity of the offspring, which make it possible for the algorithm to escape from local optimum and avoid falling into premature convergence.

3.3. Tissue-like P Processing

After the operation, the two limit individuals remains in the areas between the basic outer membrane and the surface membrane. The remaining (N-2) individuals are put into m basic membranes equally, which forms a P system.And each individual in each membrane can be viewed as a cell.In this section,the membranes will be arranged as a loop topology.The individuals have some evolution rules to evolve each other in the system, while communication rules between individual membranes are used to exchange and share the information.

In this paper, the tissue-like P system uses a single-layer membrane structure because the layers of the tissue-like P systems are related to the magnitude of population size N. In order to make the running time of each membrane relative to the average of the running time,we define the number of the membranes as follows:

$$Z=\left\lfloor \sqrt{N-2} \right\rfloor$$

Where N is the total number of individuals of initial population, Z is the square root of (N-2) (round down) [[6]]. When N is large enough, we can get the square root of N directly. Using this way to define the number Z can maintain similar limit individuals between intramembrane and extramembrane, which can reduce the overall convergence time.

3.3.1. Rules

At the same time, the evolution rule of the membrane is defined. The role of evolution rules is to evolve the individuals in the membranes to generate new individuals used in the

calculation of fitness. During the evolution, each membrane maintains the same size (the number of individuals). Every individual in each membrane will be evolved after executing the selection, crossover and mutation in turn. When the individuals are evolved,each membrane sends its best individual to the neighboring membranes (such as membrane i-1 and membrane i+1) and by using the communication rule, we retrieve the best individual from the neighboring membranes which constitute a matching pool of the individuals in the next calculation step. A matching pool is made up of all individuals in each membrane and the best individuals from its two adjacent membranes. The individuals in matching pool will be evolved by executing selection, crossover and mutation operations in turn.In order to maintain the size of individuals in each membrane, truncation operation is used to constitute new individual pool according to the fitness function. The individuals in new individual pool will be regarded as the individuals to be evolved in next computing step [[8]].

The special logical structure can bring the following benefits:

(1) The co-evolution of individuals in the m membranes can accelerate the convergence of the proposed TPDNA-GA algorithm.

(2) The individual sharing mechanism of the local neighborhood structure can enhance the diversity of individuals in the entire system.

After all the individuals in the tissue-like P system are evolved, the best individual will be communicated to the environment. They will step to next round operation together with the initial 2 limit individuals. As it involves multiple best individuals in the evolution of the systems, the tissue-like P system can effectively solve the problem of local optimum algorithm.

3.3.2. Selection Operator

In this paper, we adopt the limit individual reservations and the fitness function as the selection operator. Choose the smallest fitness value of the individuals in each generation to compare with the best individual from the previous generation and select the minimum fitness value of the individuals as the limit individuals to participate in selecting of the next generation. It will not release the best individual to the extramembrane until it reaches the conditions of the membrane dissolution (when it reaches the maximum execution step number G). The best individual will carry on the iteration operation outside the membrane. In the outer membrane, we adopt the limit individual reservations as well.

3.3.3. Crossover Operators

In this paper, we use the reconstruction crossover operator. Before the reconstruction of the cross operation, two individuals from the population should be selected as the parents. Since the main purpose of the reconstruction is to change the similarity of the outstanding individual, the choice of the parents is not random, which is to say we need choose the relatively similar individuals. Firstly, choose one individual as one of the parents randomly in the individuals with better fitness. And randomly select two individuals as a candidate parents. Then by comparing the similarity between candidate parents and the known parents, we select the individual that is more similar to the known parents as another parent of the reconstruction operation. Here we define that the similarity between individuals is the distance between individuals. The smaller the distance is, the more similar the two bodies are believed.After the two parents are determined, we define the individual with better fitness as the parent A and another individual is defined as parent B.

The specific description of the reconstruction crossover operation process is shown in the Figure 1.

At first, we choose one base randomly in the latter of the parent A that is the [L/2, L] and cut the sequence fragment R behind this base from the parent A. Then paste the fragment into the front of the parent B, which form intermediate individual A and intermediate individual B.In the DNA genetic algorithm, the length of the individual is fixed. Therefore it needs to generate a sequence fragment R' in the latter of intermediate individual A which contains the same number of bases with R. At the same time,we cut a sequence fragment from the latter of intermediate individual B that also contains the same number of bases with R.Thereby,it forms two new individuals: progeny A and progeny B.In this paper, we adopt the same probability of crossover operator inside and outside the membrane.

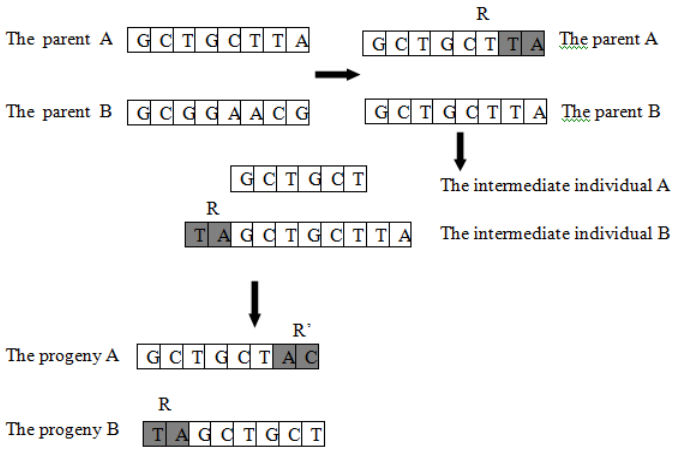

Figure 1. The reconstruction crossover operation

It should be noted that due to the reconstruction of the crossover operation is more complex and the changes are larger for the population individuals, so the execution probability Pr of the reconstruction cross should choose a smaller value. Here we define Pr=0.05.

3.3.4. Mutation Operator

In the genetic algorithm, mutation operator is mainly to ensure the diversity of the groups in a certain degree. In order to fully display the performance of the crossover operator presented in this paper, the mutation operator use the common variants (normal mutation, NM) operator that is commonly used in the DNA genetic algorithm.

This operator is similar to the flip variation of the binary GA,which is to mutate every base in the individual with probability p_m into another base. The probability of performing variation of the whole individual is $p_m \times L$. For example, the base C is replaced by base A in the individual.

The single-point mutation is used to realize the mutations of individuals. If v is a mutation point determined according to mutation probability p_m, its value becomes, after mutating.

$$\begin{cases} v \pm 2\times\delta\times v, & v \neq 0 \\ v \pm 2\times\delta, & v = 0 \end{cases}$$

Where the signs "+" or "-" occur with equal probability, and δ is a real number in the range [0,1], generated with uniform distribution.

3.4. The Implementation Steps of the TPDNA-GA Algorithm

The procedure of TPDNA-GA can be summarized as follows:

(1) Before running the algorithm,we need to set the parameters of the algorithm (initializing the population), including population size N, the largest population evolution algebraic G, reconstruction cross execution probability P, common variant execution probability p_m;

(2) Randomly produce N individuals with the length of L=n x l, which make up the initial population. And the current evolution generation is set to 1;

(3) DNA encoding:calculate the value of each individual's fitness and sorting;

(4) Select and reserve two individuals with the maximum and minimum fitness value. The remaining (N-2) individuals are put into m membranes equally, which forms a tissue-like P systems. We set the maximum execution step number as T1 inside the membranes;

(5) The proposed genetic operations were carried out for each of the individuals within the membranes: selection, crossover and mutation.

(6) Run the step of (5) in each of the membrane repeatedly until the membrane operation reaches the maximum evolution generation and produce a best individual which constitutes a matching pool with the individuals in the adjacent membranes. The matching pool consists of all individuals in one membrane and the two best individuals from the adjacent two membranes. The individuals in matching pool will be evolved by executing selection, crossover and mutation operations in turn. In order to maintain the size of individuals in each membrane, truncation operation is used to constitute new individual pool according to the fitness function. The individuals in new individual pool will be regarded as the individuals to be evolved in next computing step.

(7) Run the step of (5) in the matching pool until it reaches the maximum execution step number (with the same membrane execution step number T1 inside the membrane).Then the membranes are dissolved. The final best individual(a total of one) is released to the environment which composes the elite population with the initial two limit individuals.

(8) Elite population conduct the genetic manipulation: crossover and mutation, selection. We set the maximum execution step number as T2 outside the membrane.

(9) Run the step of (7) repeatedly until the operation reaches the termination condition. And the optimal solution is generated.The TPDNA-GA algorithm ends.

4. Application of TPDNA-GA Algorithm in Clustering

4.1. Introduction of Clustering and Clustering Measure

From now on, let us assume that we are concerned with a data set D, which has n sample points and is partitioned into k clusters, $C_1,C_2,...,C_k$. Each of them is represented as an m-dimensional vector of real numbers, say $x_1, x_2,...,x_n$, where $x_i = [x_{i1}, x_{i2},...,x_{in}]$. Denote by $z_1,z_2,...,z_k$ the corresponding cluster centers. If the distances of sample point x_i to cluster centers z_p (p = 1,2,...,k) satisfy:

$$\| x_i - z_j \| \leq \| x_i - z_p \|, p = 1,2,..., k \text{ and } j \neq p,$$

Then sample point x_i is assigned to cluster C_j, i = 1,2,...,n.

Generally speaking, partitional clustering algorithm searches for the optimal cluster centers in the solution space according to some clustering measure in order to solve data clustering problem. A commonly used clustering measure is:

$$M(C_1,C_2,...,C_k) = \sum_{i=1}^{k} \sum_{x_j \in C_i} \| X_j - Z_i \|$$

The smaller the M value, the higher the clustering quality. In this work, the clustering measure is also used to evaluate the individuals of the system during membrane evolution. If the M value of an membrane is the smaller, the membrane is the better, otherwise, it is worse.

Since data set D has k cluster centers and each cluster center is a m-dimensional vector, each membrane in the system is considered as a (k×m)-dimensional real vector of the form:

$$Z = (Z_{11}, Z_{12},\ldots,Z_{1m},\ldots,Z_{i1}, Z_{i2},\ldots,Z_{im},\ldots,Z_{k1}, Z_{k2},\ldots,Z_{km})$$

Where $z_{11},z_{12},...,z_{1d}$ are d components of ith cluster center Z_i, i=1,2,...,k. For simplicity, suppose that each cell has the same number of objects, which is denoted by D [[9]].

Initially, the system will randomly generates m initial objects for each cell.When an initial object z is generated, (k×d) random real numbers are produced repeatedly to form it with the constraint of:

$$A_1 \leq z_{i1} \leq B_1,..., A_j \leq z_{ij} \leq B_j,..., A_d \leq z_{id} \leq B_d$$

Where A_j and B_j are lower bound and upper bound of jth dimensional component of data points, respectively, j = 1,2,...,d.

4.2. Clustering Based on TPDNA-GA

According to the components discussed above, the designed tissue-like P system based DNA can be formally used for clustering to choose the best clustering center. The optimal solution in the environment refers to the optimal clustering that we could obtain from the data given [[10]]. The objective function f(x) will be realized by evaluating the distances between two individuals from each other.

Steps that use TPDNA-GA to solve the problem of clustering are as follows:

(1) First of all, give the data sets that need to be clustered and conduct DNA encoding for data;

(2) Calculate the objective function;

(3) Determine the fitness function according to the objective function, which is used to evaluate the posterity of the genetic algorithms and grasp the evolutionary direction.

(4) Introduce TPDNA-GA algorithm operation and determine the operating operator (select operator, cross operator,mutation operator).

(5) Perform TPDNA-GA algorithm to generate the optimal solution that is the optimal clustering center.

Based on the proposed TPDNA-GA, the best object in the environment is regarded as the system output, which is the found optimal cluster centers.

5. Experiment Results and Analysis

In this section, the proposed TPDNA-GA algorithm is evaluated for clustering on the five real-life data sets provided in UCI [[11]] (including the Iris, Breast Cancer, Wine, New Thyroid and Liver Disorders) and compared with classical k-means algorithm and several clustering algorithms, including GA [[12]] and PSO [[13]]. In order to test the robustness of these clustering algorithms, we repeat the experiments 50 times for each data set. The M value is also used to measure the clustering quality of each clustering algorithm. That is to say, if the M value of one individual is the smaller than others,the fitness values will be smaller, which means the individual will be better than others. So the cluster centers will also be better. These algorithms are implemented in Matlab7.1.

Table 1. The performance comparisons of tissue-like P systems of different degrees

Data sets	4 cells	8 cells	16 cells	20 cells
Iris	96.84 ±0.0751	96.81 ±0.0435	96.75 ±0.0428	96.77 ±0.0361
Breast Cancer	2974.24 ±1.5431	2971.14 ±1.5287	2970.24 ±1.1225	2969.06 ±1.0970
Wine	16309.01 ±2.5053	16303.42 ±1.9595	16292.25 ±0.1529	16301.97 ±2.8563
New Thyroid	1885.69 ±14.3773	1870.37 ±1.7355	1869.29 ±0.9215	1871.18 ±2.2496
Liver Disorders	9860.54 ±5.7239	9859.02 ±0.5116	9851.78 ±0.0347	9857.08 ±0.1043

In the experiments, we use four kinds of tissue-like P systems with degrees 4, 8, 16 and 20 respectively. The purpose is to evaluate the effects of the number of cells (different degrees) on clustering quality. The four tissue-like P systems are applied to find out the optimal cluster centers for the five data sets respectively. The average values are used to illustrate the average performance of the algorithms while standard deviations indicate their robustness. As we can see, Table 1 provides experimental results of the tissue-like P systems of four degrees on five data sets respectively. It can be further observed that the tissue-like P system with degree 16 obtains the smallest average values and standard deviations on all of data sets, which illustrate that the tissue-like P system with degree 16 has good clustering quality and high robustness.

For this reason, we will do experiment on TPDNA-GA, k-means, GA, and PSO with degree 16. In the experiments, the parameters of TPDNA-GA are set as follows. The population size is 200 (N = 200). The largest population evolution algebraic G is set as 300, which is to say that the maximum evolution general inside A1 and outside A2 is 150. The parameters of the mutation operator is set as pm=0.01. And the crossover probability is set as Pr=0.05.The number of membranes is 15(m= 15). In order to study performances of tissue-like P systems of different degrees, four cases are considered in the experiments: q= 4; 8; 16; 20. And for each test problem, we run 50 times for the five algorithms of TPDNA-GA, k-means, GA and PSO. The results are shown in Table 2.

Table 2. The results obtained by the algorithms for 50 runs on the five data sets

Data sets	TPDNA-GA	GA	PSO	k-means
Iris	96.80	99.83	97.23	104.11
	±0.1436	±5.5239	±0.3513	±12.4563
Breast Cancer	2997.36	3249.26	3050.04	3251.21
	±2.1345	±229.734	±110.801	±251.143
Wine	16292.25	16298.42	16292.25	16312.43
	±0.1530	±2.1523	±0.1531	±9.4269
New Thyroid	1870.36	1875.11	1872.51	1886.25
	±0.9215	±13.5834	±11.0923	±16.2189
Liver Disorders	9851.81	9856.14	9851.73	9868.32
	±0.0348	±1.9523	±0.0356	±7.9274

In order to further evaluate clustering performance, the proposed TPDNA-GA algorithm is compared with GA-based and PSO-based clustering algorithms as well as classical k-means algorithm. Table 2 gives the comparison results of the tissue-like P system of degree 16 with other four clustering algorithms on the five data sets, respectively. From the results, we can see that the TPDNA-GA provides the optimum average value and smallest standard deviation in compare to those of other algorithms.For instance, the optimum value is 96.80 which is obtained in most of runs of TPDNA-GA algorithm, however, other three algorithms fail to attain the value even once within 50 runs. And the results on the Wine also show that the TPDNA-GA algorithm provides the optimum value of 16292.25 while the PSO, GA and k-means obtain 16298.42, 16292.25 and 16312.43 respectively. In addition, the tissue-like P system obtains smallest standard deviation on each data set in compare to other three algorithms, which illustrates that it has high robustness. This is a strong evidence to prove the significant superiority of the proposed algorithm.

6. Conclusion

In this paper, a new DNA-GA algorithm based on tissue-like P system (TPDNA-GA) is proposed by combining DNA-GA with tissue-like P system for the first time and we use it for clustering. Distinguished from the existing evolutionary clustering techniques, two inherent mechanisms of tissue-like P system are exploited to realize the TPDNA-GA algorithm, including evolution and communication mechanisms. And the two inherent rules are used between membranes that contain the average number of individuals.Moreover, the communication rules imply realize a local neighborhood structure, namely, each membrane exchanges and shares the best objects with its two adjacent membrane. Under the control of two rules and the fitness function of individuals, the TPDNA-GA algorithm is able to search for the optimal cluster centers for a data set to be clustered.In addition, the local neighborhood structure can guide the exploitation of the optimal object and enhance the diversity of evolution objects. Therefore, the TPDNA-GA algorithm presented in this paper can be viewed as a successful instance for data clustering.

After repeated verification, the TPDNA-GA algorithm performs well when the population size is large enough. The future work is to improve the algorithm efficiency when the population size is small.

References

[1] Martin-Vide C, Pazos J, Paun Gh, Rodriguez-Paton A. A new class of symbolic abstract neural nets: Tissue P systems. In: Ibarra OH, Zhang L. *Editors*. COCOON. Lecture Notes in Computer Science. Springer. 2002: 290-299.
[2] Martin-Vide C, Paun Gh, Pazos J, Rodriguez-Paton A. Tissue P systems. *Theoretical Computer Science*. 2003; 296(2): 295-326.
[3] Peng H, Zhang J, Wang J, Wang T, Pérez-Jiménez MJ, Riscos-Núñez A. *Membrane Clustering: A Novel Clustering Algorithm under Membrane Computing*. Twelfth Brainstorming Week on Membrane Computing (BWMC2014). 2014: 311-328.
[4] Carnero J, Diaz-Pernil D, Gutierrez-Naranjo MA. *Designing tissue-like P systems for image segmentation on parallel architectures*. Ninth Brainstorming Week on Membrane Computing. 2011: 43-62.
[5] Freund R, Păun G, Pérez-Jiménez MJ. Tissue P systems with channel states. *Theoretical Computer Science.* 2005; 330(1): 101-116.
[6] Wang K, Wang N. A novel RNA genetic algorithm for parameter estimation of dynamic systems. *Chemical Engineering Research and Design.* 2010; 88(11): 1485-1493.
[7] Zhao S, Liu X. Research on a New DNA-GA Algorithm Based on P System*. Frontier and Future Development of Information Technology in Medicine and Education*. Springer, Netherlands. 2014: 1691-1698.
[8] Escuela G, Gutiérrez-Naranjo MA. *An application of genetic algorithms to membrane computing*. Eighth Brainstorming Week on Membrane Computing. 2010: 101-108.
[9] Bakar RBA, Watada J, Pedrycz W. DNA approach to solve clustering problem based on a mutual order. *Biosystems.* 2008; 91(1): 1-12.
[10] Bakar RBA, Watada J, Pedrycz W. A proximity approach to DNA based clustering analysis. *International Journal of Innovative Computing, Information and Control.* 2008; 4(5): 1203-1212.
[11] http://archive.ics.uci.edu/ml/
[12] S Bandyopdhyay, U Maulik. An evolutionary technique based on k-means algorithm for optimal clustering in RN. *Inf. Sci*. 2002; 146: 221-237.
[13] YT Kao, E Zahara, IW Kao. A hybridized approach to data clustering, Expert Systems with Applications. 2008; 34(3): 1754-1762.