A novel particle swarm optimization algorithm based on membrane system for clustering problem

Yanmeng Wei [a], Qingfei Liu [b]

School of Management Science and Engineering, Shandong Normal University Jinan, China

[a]15154122301@163.com, [b]liuqingfei1201@163.com

Abstract

Membrane computing is a class of distributed parallel computing model. In this paper,we propose a new clustering algorithm in which a cell-like membrane system is designed as its computing framework. First, an improved particle swarm optimization is proposed. That is, a well-defined environmental factor is taken into account to inspire the robust behaviors of bird flocking in depth which is called EPSO. Specifically, it not only can carry out effective searching in limited searching space, but also can strengthen the social behaviors of individuals and can accelerate the convergent rate. Second, the cell-like membrane system we designed can realize the communication between several swarms.In fact,the membrane system not only can realize parallel computing but also can avoid trapping into local optimum.Experiment results show that the proposed EPSO algorithm is more superior or competitive.

Keywords

Membrane computing, particle swarm optimization, clustering.

1. Introduction

Membrane computing, initiated by Gh. P˘aun[25], as a new branch of natural computing, aims to abstract computing models from the structure and functioning of living cells. Generally, they are characterized by three elements: (i) membrane structure, (ii) multisets and (iii) evolution rules. Fig.1 presents the structure of membrane system.The multisets of objects are placed in compartments surrounded by membranes, and evolved by some given rules[40].P systems have several interesting features: non-determinism, programmability, extensibility, readability, they are easy to communicate, and many variants have been proposed [23,25,28,29,38].

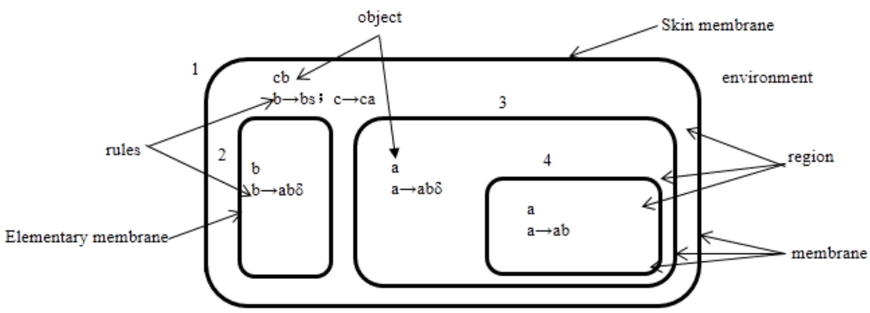

Fig.1. The structure of membrane system

Particle swarm optimization has been proposed originally by Kennedy and Eberhart in 1995,which is a population based stochastic optimization techniques inspried by social behavior of bird flocking or fish schooling[14].Compared with other algorithms, Particle Swarm Optimization (PSO) algorithm uses individual and global particle experience, and has a well-balanced mechanism to exploitation and exploration abilities[8]. A dynamic topology DCluster algorithm which based on two topologies four-cluster and fitness to solve the underlying problem is proposed by Abbas[8]. Liu proposed a population-based clustering technique which integrates different PSO algorithms and k-means algorithm to help particle escape from local optima[16]. A modified coevolutionary multi-swarm optimizer which based on new velocity updating and similarity detection is proposed by Liu [20] to solve dynamic multi-objective optimization problems.

Clustering is an unsupervised classification technique which groups a set of objects based on certain predefined criterion. The partition is done such that objects in the same clusters are more similar to each another than the objects in different clusters. As we all known,the partitional clustering techniques seek to partition a collection of documents into a set of non-overlapping groups, so as to minimum the evaluation value of clustering (Cui et al. 2005). It has been recognized that the partitional clustering techniques are well suited for clustering a large object dataset due to their relatively low computational requirements which makes it widely used. For example, the famous partitional clustering algorithms are K -means algorithm.

The rest of the paper is organized as follows. Section 2 introduces the theoretical background. In Section 3, the algorithm process is discussed in detail. Section 4 provides the experiment setup, other algorithms compared, experimental results and analysis. Finally, Section 5 draws the conclusions.

2. Theoretical background

2.1 Conceptions of basic membrane system

Our cell-like system has a two-layer membrane structure, which consists of a skin membrane and n elementary membranes, shown in Figure.1.The n elementary membranes are labeled by 1, 2 and n respectively and the skin membrane is labeled by 0. The cell-like membrane system not only can realize parallel computing but also can communicate best value with the neighboring membrane in every iteration.In fact, the cell-like membrane system can carry out multi-swarms co-evolution and avoid trapping into local optimum.

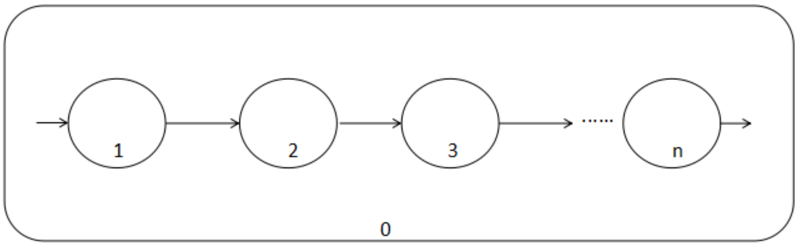

Fig.2. The designed cell-like membrane system

Evolution rules are contained in n elementary membranes. Each elementary membrane contains N objects, while the skin membrane has only one object that expresses the best value F found so far in the whole system during the computation(a global clock is set).Besides,every iteration each elementary membrane deliver their current value to the next membrane respectively.At last,after the computation,the best value stored in the skin membrane.

2.2Conception of particle swarm optimization algorithm

PSO algorithm is an evolutionary computation technique which inspired by social behavior of bird flocking, it uses the physical movements of individual or birds in the swarm. Each bird adjusts to their flight path according to the flying experience of its own and global individual in every iteration. Through this natural phenomenon, Dr.Eberhart and Dr.Kennedy proposed PSO algorithm in 1995[14]. Compared with other intelligence algorithms, the characteristics of PSO algorithm are high efficiency, quick convergence, good robustness, easy implementation and there are few parameters to adjust.

In the PSO algorithm, the population size of particle swarm is , is the dimension of search space, is the fitness function, is the maximum iteration number, each particle has a position vector , and a velocity vector, is the best position of particle in history, is the best particle in the swarm also called global particle.

Algorithm 1: PSO Algorithm	
1	Particle Swarm Population Initialization
2	for i=1 to M (Population Size)
3	Best Particle(i).Position= Particle(i).Position
4	Best Particle(i).Fitness= Particle(i).Fitness
5	End for
6	Global Best Particle=Best (Particle Population)
7	for t=1to Tmax (Max Interation)
8	for i=1 to M
9	$v_{ij}(t+1) = w * v_{ij}(t) + c_1 * r_1 * \left(p_{ij}(t) - x_{ij}(t)\right) + c_2 * r_2 * \left(p_{gj}(t) - x_{ij}(t)\right);$
10	$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1);$
11	if Particle(i).Fitness< Best Particle(i).Fitness
12	Best Particle(i).Position= Particle(i).Position
13	Best Particle(i).Fitness= Particle(i).Fitness
14	End if
15	End for
16	Global Best Particle=Best (Particle Population)
17	End for
18	Output Global Best Particle

3. The environmental-factor particle swarm optimization algorithm

It is been known that the searching space of bird is limited within a certain period due to the constraint of environment. So the environmental factors can not only strengthen the species features, but also can limit their living space. In order to simulate them better into specific behaviors, we propose a method of bird flocking with the environment constraints. Subsequently, the particles are manipulated by:

$$v_{ij}(t+1) = w * v_{ij}(t) + c_1 * r_1 * \left(p_{ij}(t) - x_{ij}(t)\right) + c_2 * r_2 * \left(p_{gj}(t) - x_{ij}(t)\right) + c_3 * r_3 * \left(pe_{ij}(t) - x_{ij}(t)\right) \tag{5}$$

$$pe_{ik} = \frac{\sum_{j=1}^{N_k} d_{k,j}}{N_k} \tag{6}$$

where c3 is a positive constant; r3 is a random function in the range (0, 1); pei,d denotes environmental factors; the rest of the other parameters are the same to the PSO.where pei,d represents the environmental factors vector of the kth centroid in i particle; dk,j denotes the jth data object vector that belongs to the kth centroid; Nk is the total number of the data objects that belong to the kth centroid. Therefore, the environmental factors of the ith particle can be represented as pei,d = (pei1, pei2,..., peik,..., pein); n is the number of the centroids and d denotes the dimension of the environmental factors.

As we have illustrated the cell-like membrane system in the former section,we initialize n swarms containing the same number.So in every iteration,we will get n global best value at the same time.Based on the communication rules of membranes,neighboring membranes will communicate the global value with each other.What's more,the designed membrane system adopts a simple halting condition, namely, maximum iterations. The membrane system will continue to execute until the halting condition is reached. When the system halts, the global best object stored in the skin membrane is regarded as final computing result.

4. Experimental results and discussion

4.1Datasets used for experiments

To evaluate the effectiveness and efficiency of the proposed EPSO algorithm,three real-word datasets have been widely used in the data mining community and reported comprehensively in the famous UCI machine learning repository.Short descriptions of these datasets are summarized in Table1.

Table 1. Datasets used for experiments

Datasets	Source	Number of records	Data-dimension	Number of clusters
Iris	UCI	150	4	3
Glass	UCI	214	9	6
Wine	UCI	178	13	3

(1)Iris: This dataset consists of 150 points distributed over three clusters: Setosa, Versicolor and Virginica. The dataset was in a four-dimensional space (sepal length, sepal width,petal length and petal width). Two classes (Versicolor and Virginica) showed a large amount ofoverlap while the class Setosa was linearly sep-arable from the other two.

(2)Glass: This dataset has 214 points having nine features (ID number, refractive index, sodium,magnesium, aluminum, silicon, potassium, calcium, barium and iron). There are six categories present in this dataset.

(3)Wine: This dataset has 178 points and 13 features resulting from a chemical analysis of wines grown in the same region in Italy but all of which were derived from three different cultivars. These data have been divided into three clusters.

Table 2. Parameters setting used in the experiment

Parameter	Description	EPSO	PSO
C1	Cognitive parameter	2	2
C2	Social parameter	2	2
C3	Environmental parameter	2	NONE
R1	Random value	[0,1]	[0,1]
R2	Random value	[0,1]	[0,1]
W	inertia weight	Dynamic	Dynamic
T	number of iterations	300	300
N	number of particles	40	40

4.2Comparison and analysis with other algorithm

In this section, some experiments are carried to evaluated the performance of the EPSO algorithm.In order to show convergence characteristics of the proposed method,a comparative experiment of the method is performed.To get the comparing experiment results,we conduct experiment by initial PSO ,the proposed EPSO and K-means respectively.

The experiments are conducted using the Matlab computation platform on a single personal computer with the Win7 operating system.

Table 3. The evaluation of values for two clustering algorithms

Datasets	k(number of clusters)			The evaluation of values	
			EPSO	PSO	K-means
		Max.v	97.7384	99.6368	99.6544
Iris	3	Min.v	96.6799	98.6185	98.5631
		Avg.v	96.8957	98.9976	98.5574
		Max.v	320.179	341.0258	345.032
Glass	6	Min.v	311.272	333.0173	334.011
		Avg.v	317.472	336.5324	339.4682

		Max.v	15891.2	15890.2	15999.8
Wine	3	Min.v	15889.3	15887.7	15907.5
		Avg.v	15890.3	15889.5	15958.3

Table3 shows the performances of these algorithms. Max.v, Min.v and Avg.v represent the maximum, minimum and average values of F respectively. We can see from Table 2 that, in terms of Max.v, Min.v and Avg.v our approach of EPSO achieves the best performance in datasets both Iris and Glass in comparison with other three algorithms. To reveal the stability of EPSO for achieving high F value, Avg.v is further taken into account. Specifically, we calculate the promotions of EPSO in comparison with PSO and K-means.

4.3Analysis of convergence behaviour

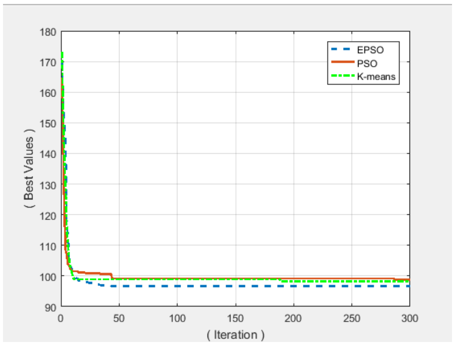

Fig. 3. Convergence of the EPSO,PSO and K-means algorithm(Iris)

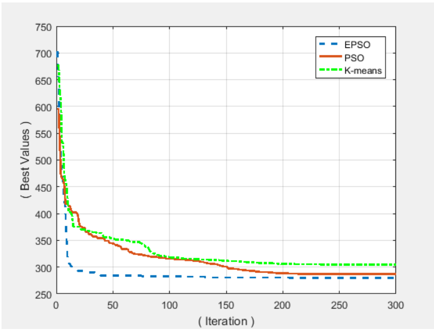

Fig. 4. Convergence of the EPSO,PSO and K-means algorithm(Glass)

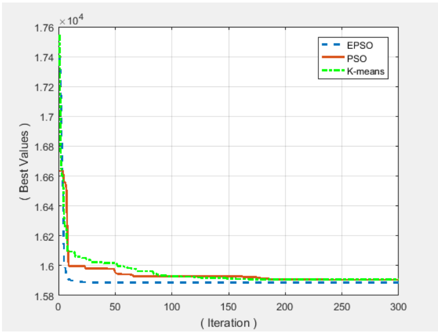

Fig. 5. Convergence of the EPSO,PSO and K-means algorithm(Wine)

Fig.3, Fig.4 and Fig.5 all illustrated the comparative results clearly.It is evident that the convergent rate of our EPSO is faster than PSO and K-means in the three datasets in almost every iteration because the proposed EPSO can effectively avoid trapping into local optimum.As pointed in the figure,the convergent rate of EPSO is faster than PSO and K-means.And the environmental factor we proposed can make best use of the best value of every iteration for velocity and position evolution.What's more,as we can see the best value of the EPSO is smaller than the others.

If you follow the "checklist" your paper will conform to the requirements of the publisher and facilitate a problem-free publication process.

5. Conclusion

Clustering is one of the key tasks of exploratory data mining and the subject of active research in several research fields, including finance, information retrieval, network management, biology, and medicine. Swarm intelligence (SI) is a relatively new interdisciplinary field of research, which has gained huge popularity in the data mining area. Swarm intelligence (SI) methodologies, such as particle swarm optimization (PSO), have been recently successful in a number of real world and synthetic clustering problems.

Inspired by the relationship between the species and the correlated environment, the EPSO is proposed in this study. In EPSO algorithm, the historic best personal information of the particle, the global best information of the swarm and the environmental factors are all used in the velocity update equation to achieve the multilocal optimums and improve its global search capacity. This approach can effectively overcome the shortcoming of initial PSO which is prone to trap into local minimal solution.Besides,the cell-like membrane system has great parallelism and carry out multi-swarm co-evolution in limited time.In our experiments, EPSO is compared with the initial PSO and K-means clustering algorithms, and the experimental results show that EPSO can achieve better performance in most cases. In the future work, we will continue concentrating on the issue of how to enhance the EPSO algorithm and apply it to some highly dimensional datasets.

Acknowledgements

Project is supported by National Natural Science Foundation of China (61472231, 61502283,61640201), Ministry of Education of Humanities and Social Science Research Project, China (12YJA630152), Social Science Fund Project of Shandong Province, China (16BGLJ06, 11CGLJ22).

References

[1] A. Hatamlou, Black hole: A new heuristic optimization approach for data clustering. Inf. Sci(222)175–184 (2013).

[2] Alok A K, Saha S, Ekbal A: A new semi-supervised clustering technique using multi-objective optimization. Applied Intelligence43(3):633-661(2015).

[3] B. Chen, Y. Lin, W. Zeng, D. Zhang, Y.-W. Si:Modified differential evolution algorithm using a new diversity maintenance strategy for multi-objective optimization problems. Appl. Intell 1–25 (2015).

[4] Bao Z, Watanabe T: Mixed constrained image filter design using particle swarm optimization. Artificial Life and Robotics, 2010, 3(15):363-368.

[5] Ben Niu:A population-based clustering technique using particle swarm optimization and k-means.Nat Comput16:45–59 (2017).

[6] Bharti K K, Singh P K:Opposition chaotic fitness mutation based adaptive inertia weight CSPSO for feature selection in text clustering. Applied Soft Computing43:20-34(2016).

[7] Cheng, Yongming, M. Jiang, and D. Yuan:Novel Clustering Algorithms Based on Improved Artificial Fish Swarm Algorithm. International Conference on Fuzzy Systems and Knowledge Discovery IEEE141-145(2009).

[8] Çomak E:A modified particle swarm optimization algorithm using Renyi entropy-based clustering. Neural Computing Applications27(5):1381-1390(2016).

[9] Ding, Yi, and X. Fu:Kernel-based fuzzy c-means clustering algorithm based on genetic algorithm.Neurocomputing (188):233-238(2016).

[10] G. Xu, Y.q. Yang, B.B. Liu, Y.h. Xu, A.j Wu: An efficient hybrid multi-objective particle swarm optimization with a multi-objective dichotomy line search. J. Comput. Appl. Math (280)310–326 (2015).

[11] G. Zhang, M. Gheorghe, L. Pan, M.J. Pérez-Jiménez:Evolutionary membrane computing: a comprehensive survey and new results. Inf. Sci. (279)528–551 (2014).

[12] Gagolewski M, Bartoszuk M, Cena A. Genie: A new, fast, and outlier-resistant hierarchical clustering algorithm. Information Sciences, 2016, 363:8-23.
[13] K.Z. Gao, P.N. Suganthan, T.J. Chua, C.S. Chong, T.X. Cai, Q.K. Pan: A two-stage artificial bee colony algorithm scheduling flexible job-shop scheduling problem with new job insertion.Expert Syst. Appl42 (21) 7652–7663(2015).
[14] Kennedy J, Eberhart R. Particle swarm optimization. IEEE International Conference on Neural Networks, 2002, (4):1942-1948.
[15] Lai HP, Visani M, Boucher A, Ogier JM: A new interactive semi-supervised clustering model for large image database indexing. Pattern Recogn Letter 37:94–106 (2014).
[16] Liu X, Liu H, Duan H: Particle Swarm Optimization based on Dynamic Niche Technology with Applications to Conceptual Design[J]. Computer Science38(10):668-676(2006).
[17] Liu X, Xue J:A Cluster Splitting Technique by Hopfield Networks and Tissue-like P systems on Simplices[J]. Neural Processing Letters1-24(2017).
[18] Liu X, Xue J:Spatial Cluster Analysis by the Bin-Packing Problem and DNA Computing Technique[J].Discrete Dynamics in Nature and Society (5187):845-850(2013).
[19] Liu XY, Zhao Y, Sunb M:An Improved Apriori Algorithm Based on an Evolution-Communication Tissue-like P System with Promoters and Inhibitors. Discrete Dynamics in Nature and Society (1):1-11(2017).
[20] Liu, Gang, and Z. Guo:A clustering-based differential evolution with random-based sampling and Gaussian sampling. Elsevier Science Publishers B. V. (205):229-246(2016).
[21] Liua X, Zhao Y, Sunb M: An Improved Apriori Algorithm Based on an Evolution-Communication Tissue-like Tissue-like P system with Promoters and Inhibitors[J]. Discrete Dynamics in Nature and Society2017(1):1-11(2017).
[22] Lv Y, Ma T, Tang M, et al. An efficient and scalable density-based clustering algorithm for datasets with complex structures. Neurocomputing171(C):9-22(2015).
[23] M. Ionescu, Gh. Paun, T. Yokomori:Spiking neural P systems, Fundam. Inform. 71 (2-3) 279–308(2006).
[24] Netjinda N, Achalakul T, Sirinaovakul B. Particle Swarm Optimization inspired by starling flock behavior. Elsevier Science Publishers B. V.35(C):411-422(2015).
[25] PAUN GH., ROZENBERG G., SALOMAA A. eds., The Oxford Handbook of Membrane Computing. Oxford University Press, New York (2010)
[26] PENG H., WANG J., PEREZ-JIMENEZ M.J., WANG H., SHAO J., WANG T:Fuzzy reasoning spiking neural P system for fault diagnosis, Information Sciences (235)106–116 (2013).
[27] Qu BY, Suganthan PN, Das S: A distance-based locally informed particle swarm model for multimodal optimization. IEEE T Evolut Comput 17(3):387–402 (2013).
[28] R. Freund, Gh. Paun, M.J. Pérez-Jiménez: Tissue-like P systems with channel-states. Theoret. Comput. Sci330 (1) 101–116(2005).
[29] Shahrivari S, Jalili S. Single-pass and linear-time k-means clustering based on Map Reduce. Information Systems 60(C):1–12(2016).
[30] WANG J., SHI P., PENG H., PEREZ-JIMENEZ M.J., WANG T.:Weighted fuzzy spiking neural P systems. IEEE Transactions on Fuzzy Systems21(2) 209–220 (2013).
[31] WANG J., ZHOU L., PENG H., ZHANG G.:An extended spiking neural P system for fuzzy knowledge representation.International Journal of Innovative Computing, Information and Control 7(7A) 3709–3724 (2011).
[32] Wei Song:Particle swarm optimization algorithm with environmental factors for clustering analysis.Soft Comput(21)283-293(2017)
[33] X. Zeng, L. Xu, X. Liu, L. Pan: On languages generated by spiking neural P systems with weights. Inf. Sci. 278423–433 (2014).
[34] X. Zhang, Y. Liu, B. Luo, L. Pan: Computational power of tissue P systems for generating control languages. Inf. Sci. (278) 285–297 (2014).

[35] Y.-B. Zhong, Y. Xiang, H.-L. Liu:A multi-objective artificial bee colony algorithm based on division of the searching space.Appl. Intell. 41 (4) 987–1011 (2014).

[36] Yamina Mohamed Ben Ali:Unsupervised Clustering Based an Adaptive Particle Swarm Optimization Algorithm.Neural Process Lett 44:221–244 (2016).

[37] Yuwono M,Su SW,Moulton BD, Nguyen HT(2014) Data clustering using variants of rapid centroids estimation. IEEE Trans Evol Comput,2014, 18(3):366-377.

[38] Z. Zhang, X. Yi, H. Peng: A novel framework of tissue membrane systems for image fusion.Bio-Med. Mater. Eng 24 (6) 3259–3266 (2014).

[39] Zhao X, Lin W, Hao J, et al. Clustering and pattern search for enhancing particle swarm optimization with Euclidean spatial neighborhood search. Neurocomputing171(C):966-981 (2016).

[40] Zhao Y, Liu X, Wang W: Spiking Neural Tissue-like P systems with Neuron Division and Dissolution[J]. Science China11(9)(2016).