A CLUSTER ANALYSIS METHOD USING ARTIFICIAL IMMUNE SYSTEM AND ANT ALGORITHM

Chui-Yu Chiu[*] *Chia-Hao Lin* *Chienwen Wu*
Industrial Engineering and Management Department
National Taipei University of Technology, Taipei, Taiwan 106, ROC
E-mail: cychiu@ntut.edu.tw

ABSTRACT

In this paper, we proposed a new algorithm, the immunity-based Ant Clustering Algorithm (IACA) which integrates the artificial immune system with ant algorithm for clustering analysis. A total of 243 data sets generated by Monte Carlo method are used to evaluate the performance of our proposed method and other known methods.

INTRODUCTION

Recently, data mining has attracted a lot of attention in the information industry and academic research due to the wide availability of large amounts of data and the imminent need for turning such data into useful information and knowledge. Data clustering takes the collected data which has similar characteristics into the same cluster and analyzes the relationship among these objects or points.

Ant Algorithm

Ant algorithm [1] is a heuristic method successfully applied to several NP-hard combinatorial optimization problems. ACS--Ant Colony System (ACS) has been introduced to improve the performance of Ant System [2]. Ant System on Clustering Analysis, ASCA, is further applied to clustering problems[3].

Artificial Immune System

Immune Algorithm, is proposed based on the theory of immunity in biology [4][5]. A novel algorithm is further proposed called Immune-based Genetic Algorithm[6].

METHODOLOGY

Let $\boldsymbol{E} = \{O_1, O_2, \ldots O_n\}$ be the data set with n objects, where $\boldsymbol{O}$ indicates the objects(or data, items) collected from the database. And each object has $\boldsymbol{k}$ attributes (dimensions), where $k>0$.

Here we propose a new algorithm called the Immunity-based Ant Clustering Algorithm which uses the concept of the ant algorithm and the artificial immune system. The main procedure of Immunity-based Ant Clustering Algorithm can be denoted as Table 1.

In next procedure, the algorithm applies the immune system for clustering. There are two steps in the procedure ***Immune_System.*** The procedure ***Antigens Recognition*** is to find the critical objects which have high similarity between different clusters and denoted as in Table 2. The procedure ***Antibodies Test*** is to determine the objects in the immune solution space to merge or not and denoted as in Table 3.

Table 1. The procedure of Immunity-based Ant clustering Algorithm

Procedure Immunity-based Ant Clustering Algorithm
Initialize the parameters and pheromone trail
While (termination_criterion_not_satisfied)
Divide_obj(); /*divides the initial cluster into several sub-clusters*/
Agglomerate_obj(); /*agglomerates objects into the suitable sub-cluster */
Agglomerate_clu(); /* merge the similar two sub-clusters into a cluster */
Agglomerate_obj(); /* remove objects from clusters */
Remove();
IF (TWCV doestn't change)
Cluster the non-clustered object to the closest cluster.
Immune_System();
Else repeat the preceding procedure
End IF
End While
Cluster the non-clustered objects to the closest cluster
End Procedure Immunity-based Ant Clustering Algorithm

Table 2. The procedure Antigens Recognition

Procedure Antigens Recognition
Antigens Recognition
This procedure searches the solution space between each object in different clusters.
Let O_j belongs to cluster *k1*, O_j belongs to cluster *k2* , $k1 \neq k2$.
For each T_{k1} and T_{k2}
If ($D(O_i, O_j) \leq \frac{1}{\psi}(D_{mean}(T_{k1}) + D_{mean}(T_{k2}))$
Take O_i and O_j to a temporary cluster T_{immune} as a immune solution space.
Else set objects in the original cluster.
End IF
End Procedure Antigens Recognition

Table 3. The Procedure of Antibodies Test

Procedure Antibodies Test
In $T_{immune.}$
Calculating $\bar{\tau}$ of all the objects
Ant *k* starts at a random object *i* to cluster the other object *j*.
IF $\tau_{ij} \geq \bar{\tau}$ where $i \in T_k$, $j \in \{n - T_k \mid k = 1,2,3,...,\mathrm{m}\}$
Ant k collects the object to a new cluster $T_{new.}$.
Else
Set the objects free
End If
$T_{new.}$ is the new cluster generated by immune system.
Return the **Procedure** Agglomerate_obj
End Procedure Antibodies Test

In order to evaluate the performance of our proposed algorithm and other algorithms, a simulated data set generated from Monte Carlo method is used as shown in Table 4. The experiment is a four-way factorial design with three replications in which 243 data sets are generated. Each factor has three levels arranged in this factorial experiment. The result is shown in Table 5.

Table 4 Factors and levels of this experiment

Factors/Levels	**I**	**II**	**III**
number of clusters	3	5	7
number of dimensions	6	8	10
density level	Equal	10%	60%
Error perturbation level	No	Low	High

Table 5 The result of the experiment

Misclassifi-Cation rates	**SOM**		**ASCA**		**IACA**	
	Original data	**Improved rates**	**Original Data**	**Improved rates**	**Original data**	**Improved rates**
3 clusters	15.64	--	8.86	43.33%	4.51	71.19%
5 clusters	21.20	--	0.53	97.50%	0.47	97.77%
7 clusters	29.77	--	0.48	98.38%	0.09	99.71%
TWCV	**SOM**		**ASCA**		**IACA**	
	Original data	**Improved rates**	**Original data**	**Improved rates**	**Original data**	**Improved rates**
3 clusters	51062.99	--	50855.74	0.41%	47267.46	7.43%
5 clusters	80982.81	--	47887.62	40.87%	47060.93	41.89%
7 clusters	148805.00	--	47010.09	68.41%	45943.84	69.12%

CONCLUSION

In this paper, we proposed the immunity-based ant clustering algorithm, IACA. From the experiment study of simulated data, IACA has better clustering performance and total within cluster variance than Self-Organizing Map method and Ant System-based Clustering Algorithm.

REFERENCES

[1] Dorigo, M., Maniezzo, V. and Colorni, A., The Ant System: Optimization by a colony of cooperating agents, *IEEE Transactions on Systems, Man and Cybernetics-Part B*, Vol.26, No.1, pp.1-13, 1996.

[2] Dorigo, M. and Gambardella, L. M., Ant Colony System: A Cooperative Learning Approach to the Traveling Salesman *Problem, IEEE Transactions on Evolutionary Comp., Vol.*1, No.1, April, 1997.

[3] Kuo, R. J., Chiu, C. Y., Wang, H. S. and Chio, S. H., Application of Ant System on Clustering Analysis in Data Mining, submitted to *IEEE Transactions on Evolutionary Computing*.

[4] Dasgupta, D. and Attoh-Okine, N., Immunity-Based Systems A Survey, *Proceeding of the IEEE Transactions on Systems, Man and Cybernetics*, Vol. 1, pp. 369 –374, 1997.

[5] Dote, Y., Soft Computing (Immune Networks) in Artificial Intelligence, *Proceeding of the IEEE Transactions on Systems, Man and Cybernetics*, Vol. 2, pp.1382 –1387, 1998.

[6] Wang, L. and Jiao, L., A novel genetic algorithm based on immunity, 2000 *IEEE International Symposium on Circuits and Systems*, pp. 385-388, 2000.