Study of Artificial Immune Clustering Algorithm and Its Applications to Urban Traffic Control

Lei Jia, Licai Yang, Qingjie Kong, Shu Lin

School of Control Science and Engineering, Shandong University,
73 Jingshi Road, Jinan 250061, Shandong Province, P. R. China
yanglc@sdu.edu.cn

Abstract

The vertebrate immune system is a decentralized and adaptive system with high complexity. The information processing mechanism extracted from vertebrate immune system can be used for solving many complex engineering problems. After reviewing the mechanism of immune system, an artificial immune data clustering algorithm based on clone selective principle and immune network theory from the vertebrate immune systems was put forward. This algorithm has been successfully used for making the time of day schemes of urban traffic control systems. It can get over the irrational intervals from manually programming and the hierarchical clustering based on genetic algorithm. This work supplies a new idea for programming time of day intervals and making urban traffic control schemes. The work in this paper proposes a novel idea for automatically making time of day projects and other complex data clustering applications.

1 Introduction

Intelligent transportation system (ITS) is a very important means to avoid the urban traffic jam and make traffic flow in order. Traffic signal control (TSC) is a crucial component in ITS. Among TSC schemes, time of day (TOD) control has been one of the most prevalent methods.

According to the traffic messages, TOD control divides a day into several intervals and runs the optimal control schemes in any interval. The key problem to run TOD control scheme is to program the intervals rationally. Traditionally, traffic engineers collect traffic data for one or two days manually and plot the aggregated traffic volume to determine the TOD break points using their experience. This approach was usually inefficient, because it cannot keep up with the rapid changes in daily traffic. Thus, an adaptive and automated tool, which utilizes a large set of archived traffic data from ITS and produces an optimal TOD plan, could be very useful. To do this, A. Hauser [1] proposed a new method to determine the break points based on hierarchical clustering algorithm and concept of system states. The experiment results show that this method can produce fairly good clusters in some special applications. In other words, Hauser's method can fairly cluster traffic data and automatically divide a day into several intervals according to the traffic condition in some special applications. However, problem arises when unclean cluster or isolated clusters be formed. These clusters do not follow an intuitive TOD scheme as the majority of the clusters do. As a result, traffic engineers have to manually assign to adjacent clusters, or spe-

cial algorithms have to be developed to refine the clusters. B. Park [2] applied genetic algorithm (GA) into TOD control and solved the problem of "unclean clusters" from Hauser's method, but due to the premature phenomena of GA, TOD intervals programming results need to be corrected manually. Park's algorithm needs to be improved.

After artificial neural networks based on principles of brain neural system and GA based on genetic theory have been applied into many fields, artificial immune system (AIS) based on vertebrate immune systems has become a research hot issue. In recent years, researchers persisted in extracting metaphor mechanisms from the vertebrate immune system and applying them into model designing, algorithm realizing and many other engineering applications.

A data clustering algorithm based on AIS was given in this paper. This algorithm does well in reducing the redundant information in clustered data, especially in dealing with the mass data clustering applications, where traditional clustering methods may be inefficient. As a case study, we applied the algorithm given in this paper to recognize traffic patterns, by which we can program TOD intervals automatically. Moreover, it can escape the irrationality of manual intervals separating and the shortcomings of grade clustering using GA.

The paper is organized as following: the theory of immune system is presented in section 2, while an artificial immune clustering algorithm is put forward in section 3. Section 4 applies the artificial immune clustering algorithm to traffic pattern analysis, and section 5 concludes the paper.

2 Theory of Immune System

Artificial immune system based on vertebrate immune systems is an approximate analogy of natural immune system. That vertebrate immune system recognizes the antigen is an evolutionary process to select the proper antibody and destroy the antigen, but it is different from biological evolutionary process [3]. A number of B cells are produced from the marrow and then they enter the lymphoid recurrence. Stimulated by an antigen in the process of combining the excreted antigen with the antibody, B cells produce a series of bio-chemical reactions to destroy the antigen. The closer the combination of antibody and antigen is, the stronger the excitation function is. And there is also stimulation and suppression between antibodies. The intensity is determined by the concentration of the antibodies. When antibodies are excited to some degree, antibody cells are mature and produce the powerful reproduction and differentiation. Most of antibodies are differentiated to be plasma cells, and the remains are differentiated as memory cells. With antigens destroyed, suppressor cells are produced to suppress mature B cells and B cells suppressed will be eliminated gradually. Natural immune system is a complex system producing multilevel defenses through non-specific (innate) and specific (adaptive) immune network. The primary function of immune system is to recognize all cells or molecules inside the body, and identify them as self and non-self, then further classify non-self cells to activate proper defensive systems. Immune system distinguishes external antigens (e.g. bacterium, virus) from self-cell or molecules by evolutionary learning.

The research on vertebrate immune system is the theoretical basis of AIS. As an excellent parallel and distributed autonomous system, AIS has been successful used for recognition and classification via learning, recollection and associated searches [4]. Its capability to process information is powerful and robust. The immune clustering algorithm given in this paper mainly bases on the immune network theory and clone selective principle [5].

3 Artificial Immune Clustering Algorithm

To describe the algorithm expediently, define follow variables:

***Ag*:** input antigens, a set of input data that will be clustered:

$$\boldsymbol{Ag} = [\boldsymbol{Ag}_1, \boldsymbol{Ag}_2, \mathrm{L}\ , \boldsymbol{Ag}_Q]^{\mathrm{T}} \qquad \boldsymbol{Ag}_j = [Ag_{j1}, Ag_{j2}, \mathrm{L}\ , Ag_{jp}] \in \boldsymbol{R}^p \qquad j = 1,2,\mathrm{L}\ ,Q \tag{1}$$

***Ab*:** antibodies, the N initial net cells:

$$\boldsymbol{Ab} = [\boldsymbol{Ab}_1, \boldsymbol{Ab}_2 \mathrm{L}\ , \boldsymbol{Ab}_N]^{\mathrm{T}} \qquad \boldsymbol{Ab}_i = [Ab_{i1}, Ab_{i2}, \mathrm{L}\ , Ab_{ip}] \in \boldsymbol{R}^p \qquad i = 1,2,\mathrm{L}\ ,N \tag{2}$$

***M*:** a set of memory data, or the n_t memory net cells:

$$\boldsymbol{M} = [\boldsymbol{m}_1, \boldsymbol{m}_2 \mathrm{L}\ , \boldsymbol{m}_{nt}]^{\mathrm{T}} \qquad \boldsymbol{m}_k = [m_{k1}, m_{k2}, \mathrm{L}\ , m_{kp}] \in \boldsymbol{R}^p \qquad k = 1,2,\mathrm{L}\ ,nt. \quad nt << Q \tag{3}$$

***D*$_j$、*AF*$_j$:** distance vector and affinity vector between the antigens and the antibodies;

S*:** similarity vector of the antibodies in ***M;

***Nc*:** number of the clones that are produced by exploding cells;

C*$_j$:** colony of the clones that are produced by exploding cells, will become ***C after affinity maturation;

The artificial immune clustering algorithm based on the immune network theory and the clone choice principle consists of following steps [6,7].

Step1 Randomly choose the N antibodies ***Ab*** as initial net cells.

Step2 Set iterative control parameter N_A.

Step3 Input the clustered data as antigens, and do with the antigens $\boldsymbol{Ag}_j$ (j=1,2,…,Q) as following:

Step3.1 Compute distance vector $\boldsymbol{D}_j$ and affinity vector $\boldsymbol{AF}_j$.

Define the distance between antigen i and antibody j as:

$$d_{ji} = \left\| \boldsymbol{Ab}_i - \boldsymbol{Ag}_j \right\|_2 = \left(\sum_{k=1}^{p} (Ab_{ik} - Ag_{jk})^2\right)^{1/2} \tag{4}$$

According to the characteristics of clustering analysis, the affinity between antigens and antibodies is determined by formula (5)

$$af_{ji} = (1 + d_{ji})^{-1} \tag{5}$$

then:

$$\boldsymbol{D}_j = [d_{j1}, d_{j2}, \mathrm{L}\ , d_{jN}]^{\mathrm{T}} \tag{6}$$

and

$$\boldsymbol{AF}_j = [af_{j1}, af_{j2}, \mathrm{L}\ , af_{jN}]^{\mathrm{T}} \tag{7}$$

Step3.2 According to optional rate *opR* choose *n* antibodies from ***Ab*** to clone. These antibodies have the highest affinities with $\boldsymbol{Ag}_j$. Then we get a set of relevant clones $\boldsymbol{C}_j$.

Depending on the clone choice principle, put the *n* chosen cells in order according to their affinities from biggest to smallest, and

$$N_c = \sum_{i=1}^{n} Int(\frac{\alpha \cdot N}{i}) \qquad n = opR \times N \tag{8}$$

where *Nc* is the number of clones produced by the *n* antibodies. α is a multiplying factor which is used to control the scale of clones, and *Int*(•) is the operator that rounds its argument toward the closest integers.

Step3.3 Randomly doing aberrance for clone cells with formula (9) in order to realize affinity maturation, by which the antibodies $\boldsymbol{C}$ with higher affinity are gotten:

$$\begin{aligned} &\boldsymbol{C} = Rand(\boldsymbol{C}_R, N_R) \\ &\boldsymbol{C}_R = \boldsymbol{C}_j - \boldsymbol{\mu}(\boldsymbol{C}_j - \boldsymbol{Ag}_j) = [\boldsymbol{Ab}_1^*, \boldsymbol{Ab}_2^*, \mathrm{L}\ , \boldsymbol{Ab}_{Nc}^*] \end{aligned} \tag{9}$$

where $Rand(\boldsymbol{C_R}, N_R)$ is a random function, which denotes that N_R variables are extracted from $\boldsymbol{C_R}$ randomly, and μ is the rate of aberrance for clone cells , which is given by formula (10):

$$\boldsymbol{\mu} = k \times exp(-\boldsymbol{AF}_v / \eta) \tag{10}$$

where $\boldsymbol{AF}_v$ is the standard value of antigenic affinity, *k* is a proportional factor, and η is a attenuation controlling coefficient.

Step3.4 Compute $\boldsymbol{AFc_j}$, which is the affinity between antibodies in $\boldsymbol{C}$ and antigens $\boldsymbol{Ag}_j$.

Step3.5 According to reselect rate *rsR*, choose several antibodies from $\boldsymbol{C}$ as a part of memory cells $\boldsymbol{M_p}$. The antibodies chosen hold the highest affinity with $\boldsymbol{Ag}_j$.

Step3.6 Eliminating the antibodies whose similarities $\boldsymbol{s}_{ij}$ are bigger than the threshold s_1 to produce the new memory set $\boldsymbol{M_k}$. The similarity of antibodies $\boldsymbol{s}_{ij}$ are given by following:

$$d_{ij} = \left\| \boldsymbol{Ab}_i - \boldsymbol{Ab}_j \right\|_2 = (\sum_{k=1}^{p} (Ab_{ik} - Ab_{jk})^2)^{1/2} \quad (11)$$

$$s_{ij} = 1 - d_{ij} \quad (12)$$

where d_{ij} is the distance of different antibodies

Step3.7 Incorporate partial memory cells $\boldsymbol{M}_k$ into memory cell set: $\boldsymbol{M}$ ($\boldsymbol{M} \leftarrow [\boldsymbol{M}; \boldsymbol{M}_k]$).

Step4 Calculate the similarity vector $\boldsymbol{S}$ between memory cells in $\boldsymbol{M}$, and get rid of the antibodies whose similarity $\boldsymbol{s}_{ij}$ are bigger than threshold σ_{s2}. Thus, the net restraint between different clone sets is realized. $\boldsymbol{s}_{ij}$ is calculated by formula (11) and (12).

Step5 According to the worst select rate *wsR*, produce several antibodies randomly to replace the lower affinity ones among initial antibodies.

Step6 Adjust the control variables, and return step3 to continue the next iteration until reaching the requisite numbers of iteration or satisfying the setting requirement of clustering analysis.

After ending the learning process, the output $\boldsymbol{M}$ is the memory data set of initial data, and $\boldsymbol{S}$ is the similarity between antibodies of $\boldsymbol{M}$. According to the distances between initial data and memory cells, or the affinity between antigens and memory cells, it is easy to cluster the antigens.

4 Study of Traffic Pattern Clustering

Programming TOD intervals is a pattern clustering problem, and can be also regarded as an optimization problem for multi-peak values [2].

There are two steps for artificial immune clustering analysis of traffic data. Firstly, consider the traffic data as the antigens in AIS, draw their characteristics according to artificial immune algorithm, and get the immune memory data sets $\boldsymbol{M}$ of traffic flow and its similarity measurement matrix $\boldsymbol{S}$. Secondly, using normal clustering methods, do clustering analysis to the immune memory data sets of traffic flow.

4.1. Pretreatment of Traffic Data

The traffic pattern clustering in TOD schemes is to find several state points of daily traffic flow curve and divide a day into several intervals according to these points. The traffic signal control systems run the special optimal schemes in different intervals.

Usually, the range of traffic data is very large. For instance, the traffic flows in daytime peak interval is much larger than those in nighttime. It is necessary to stan-

dardize the traffic data before clustering. In this paper we use formula (13) standardizing data:

$$Ag_j = \frac{Ag_j^* - \min\{Ag^*\}}{\max\{Ag^*\} - \min\{Ag^*\}} \quad (13)$$

Where, "*" denotes the initial data of traffic flow.

Experiments show that if clustering traffic data directly using a day traffic flow, the conflict of intervals will arise, i.e. the same interval maybe attach to different clusters in the different workdays (for example, on Monday in two weeks). The conflict of intervals results from the randomicity of traffic systems. To avoid the conflict of intervals, the mean of traffic flow in several days will be used for clustered antigens in this paper. This pretreatment of traffic data accords with the random characteristics of traffic data, and can reduce the computation of AIS by a long way. Using the mean of traffic flow to cluster, p in section 3 will be equal to 1, i.e. $\boldsymbol{Ag}$、$\boldsymbol{Ab}$ 、$\boldsymbol{M}$ and other matrices will be all vectors.

4.2. Case Study of Traffic Pattern Clustering

As a case study, we use the traffic data of Jiefang-Bridge intersection in Jinan to do traffic pattern clustering analysis with AIS algorithm given in this paper.

Firstly, take 96 historical mean traffic flows on Monday as antigens, and carry through immune recognition and memory by the algorithm given in section 3.

Supposes: the number of initial antibodies N=20; opR =20%; α =1; rsR = 20%, wsR =10%, and the biggest iteration time NC = 15; σ_d =0.001. Our simulation indicates that σ_s is in [0.03, 0.08] for the traffic flow clustering problems, which can ensure that the size of the data set is proper and data set can reflect the characteristics of initial data fully. Here, take σ_s as 0.045.

Table 1 depicts the AIS computing process of traffic flow data, concluding 15 iterations. It denotes the mean affinity between memory data and antigen data are bigger and the scale of immune memory data aggregate is moderate when No.14 iteration ends. It is a better clustering result.

Table. 1. AI Clustering calculation of data

Iter Num.	Ave Aff.	Cell Num.	Iter Num.	Ave Aff.	Cell Num.
1	0.6573	17	9	0.6566	18
2	0.6521	16	10	0.6673	18
3	0.6652	17	11	0.6683	18
4	0.6716	16	12	0.6643	17
5	0.6746	15	13	0.6573	18
6	0.6489	15	14	0.6771	16
7	0.6561	17	15	0.6629	18
8	0.6668	15			

Fig.1 is the network structure of the memory data from AIS algorithm using the smallest spanning trees method. The dashed lines denote the branches that can be cut off. In the graph, we can see that the 16 memory data produced by 14 iterations are divided into 6 sorts clearly.

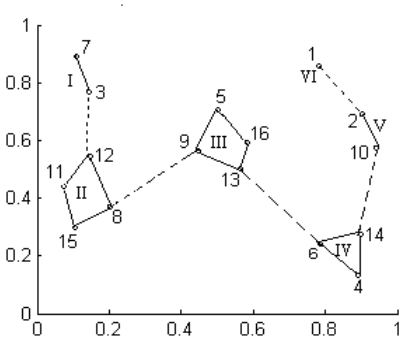


Fig. 1. Network structure of clustered data.

Secondly, calculate the distances between each traffic data using formula (11) and the centrobaric of each memory set using centrobaric method to carry out the second clustering of traffic data.

Corresponding clusters of data with traffic intervals, we get the TOD intervals programming result of case study as Fig.2.

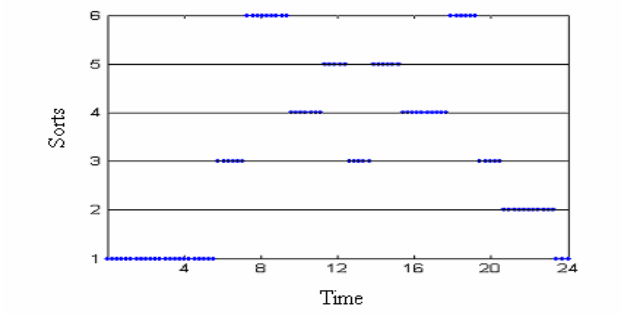


Fig. 2. TOD intervals of case study.

In above case study, a day is divided to following intervals according to traffic data: traffic valley value interval :23.15-5.30, corresponding to sort Ⅰ in Fig.1; traffic peak value intervals: 5.30-7.00, 17.00-18.30, corresponding to sort Ⅵ; peak value intervals at noon 11.00-12.15, 13.30-15.00, to sort Ⅳ; normal traffic intervals: 5.00-7.00, 12.15-13.30, 18.30-19.45, to sort Ⅲ; and traffic transitional intervals 19.45-21.00, to sort Ⅱ.

Traffic engineering practice shows that the traffic interval scheme programmed automatically using the algorithm given in this paper corresponds with that made by traffic engineers manually. Running traffic control soft programmed via above algorithm, traffic engineers can make TOD intervals automatically. Moreover, it can escape the irrationality of manual intervals separating and the shortcomings of grade clustering using GA.

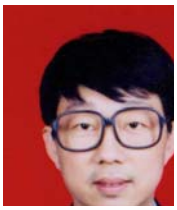
5 Conclusions

An artificial immune data clustering algorithm from metaphor based on the vertebrate immune systems was put forward. The algorithm can reduce the redundant information of the clustered data efficiently. The clustering algorithm given in this paper especially fits for the mass data clustering applications where traditional clustering methods cannot deal with effectively. As an example, the algorithm given in this paper was applied to urban traffic pattern recognition, which successfully programmed TOD schemes automatically in ITS.

Having too many adjustable parameters is one of the characteristics of the clustering algorithm given in this paper. Setting the ranges of the parameters rationally can meet the needs of different clustering applications. On the other hand, it may influence the effects of the program running and data clustering if these parameters are set irrationally. Using the engineering experience of special applications and transiting them into bacteria of AIS to improve the AIS clustering algorithm will be our future work.

References

1. T.A. Hauser, W.T. Scherer: Data Mining Tools for Real Time Traffic Signal Decision Support and Maintenance. Proc. of The IEEE International Conference on Systems, Man ,and Cybernetics, vol.3, (2001) 1471-1477.
2. B.B. Park, Do-Hoon Lee, Hsoo Yun: Enhancement of Time of Day Based Traffic Signal Control. Proc. of The IEEE International Conference on Systems, Man, and Cybernetics, vol.4. (2003) 3619-3624.
3. Xueyan Lin, Ling Zhang: Modern Cellular and Molecule Immunology. Beijing Science Press (1999). (In Chinese).
4. Zhonghua Li, Jianping Wu, Zongyuan Mao: Application of Artificial Immune Algorithm in the Dynamaic Zoning of Elevator Traffic. Proc of the 5th World Congress on Intelligent Control and Automation, (2004) 2222-2226.
5. Licheng Jiao, Haifeng Du: Development and Prospect of the Artificial Immune System. ACTA ELECTRONICA SINICA, Vol.31, No.10, (2003)1540-1548. (in Chinese).
6. J.Timmis, M Neal:An Artificial Immune Systems for data Analysis. BioSystems, Vol.55, (2000) 143-150.
7. T. Knight, J. Timmis: An Immmunological Approach to Data Mining. Proc of IEEE International Conference on Data Mining, (2001) 297-304.



Lei Jia is currently the Head of the School of Control Science and Technology, Shandong University, China. He received his B.Eng. degree in Automation and the M.Eng degree in Control Engineering from Shandong University of Technology, China, in 1982 and 1988, respectively, and the Ph.D degree in Control Theory and Control Engineering from Zhejiang University, China, in 1991. From 1992 to 1994, he was an Associate Professor, in Shandong University of Technology. Sine 1995, he has been a Professor in Shandong University. In 1998, he was appointed as the Head of the School of Control Science and Technology, Shandong University. His current research interests include Artificial Intelligent and Intelligent Control, Robust Control, Intelligent Transportation Systems, Control Theory and Applications.

Li-Cai Yang received his B.Eng. degree in Automation and the M. Eng degree in Control Engineering from Shandong University of Technology, China, in 1983 and 1990, respectively, and the Ph.D degree in Control Theory and Control Engineering from Shandong University, China, in 2005. From 1991 to 1995, he was a Lecturer in Shandong University of Technology, China. Sine 1996, he has been an Associate Professor in Shandong University, China. His current research interests include Artificial Intelligent and Intelligent Control, Intelligent Transportation Systems, Biomedical Engineering, Control Theory and Applications.