Application of Immune Genetic Algorithm on Standard Library Building

Chunxia YUAN, Ke YAO
(Information and Communication College, National University of Defense Technology, P. R. China)
81180767@qq.com

Abstract: Based on the analysis of standard system construction current situation, this paper proposes an improved genetic algorithm by combining the fuzzy neural network and the immune genetic algorithm, which can be used to solve the standard feature vector clustering, and better overcomes the premature phenomenon of general intelligent algorithms. Furthermore, this paper also builds a service-oriented standards library, which takes standard feature vectors as the main clustering index. The practical experiment shows that this smart genetic algorithm, with strong robustness, reliability and accuracy, is of certain reference value to the standard system construction innovation and development.
Key words: Fuzzy; Immune genetic algorithm; Standard library

1 Introduction

At present, with the development of science and technology and deepening of the military system reform, the scientific and technological content and technical standard of the weapon and equipment system are becoming higher and higher. More and more artificial intelligence techniques are applied to the acquisition of equipment. The versatility and compatibility of the military and civil dual-use technologies own more intersection, and knowledge of all standards correlates more and more. Especially in the face of a large number of specific standards, how to use the knowledge correlation between standards to cluster and classify, update supplement standards in time, and give full play to the maximum economic benefits and practical value of each standard becomes a problem. Therefore, a comprehensive platform of standards library should be constructed for evaluation, analysis and optimization. But in practice, the standard features quantity or characteristic attribution is usually very complicated, and the constructed standard feature matrix usually has thousands or more dimensions, which makes the extraction work very difficult. With the development of information technology, facing the high dimension and sparsity of data, the intelligent clustering, as an effective tool for big data mining, can be used to accurately and efficiently obtain the implied knowledge correlation degree from a large number of standards, and to provide a better smart algorithm for the formation of the standard library.

2 Standard status analysis

The standard is a kind of normative document, a basic rule, and a system project. The implementation of standards is to facilitate standardization and standardize various parameters. But presently, there are many kinds of standards, including international standards, national standards, local standards and industry standards. According to statistics, there are more than 30,000 national standards, nearly 40,000 industry standards, and nearly 30,000 local standards [1]. A standard system covering the three major industries and social undertakings has been established, which has played a positive and important role in serving the economic and social development [2-3]. China's national standard network platform database shows that there are 34485 national standards in China up to November 2017. According to different standardizing objectives, standards can be divided into technology standard, management standard, and work standard. With the development of national standard system reform and innovation strategy in various industries, the classification model of the existing standard system in China is becoming more and more complicated, which brings new challenges to the establishment and application of standards. With the help of artificial intelligence, big data and other technologies, through the construction of intelligent algorithm model, the standard is classified accurately, the correlation degree between standards is divided, and the resource allocation and optimization integration of the standard system are realized. It will inevitably become the future trend of innovation and development in standard system construction.

3 Analysis of fuzzy immune genetic algorithm

3.1 Basic principle

Fuzzy immune genetic algorithm (FIA), which has the characteristics of neural network, immune and genetic algorithm, is established on the basis of fuzzy neural network and immune genetic algorithm. The fuzzy neural network [4] has the ability of nonlinear adaptive information processing, which, as an intelligent system, can simulate the information processing of human brain neural system. The network has strong learning and training ability and can acquire external knowledge to solve the computer difficulties. Its basic idea [5] is to construct a fuzzy inference system with mixed algorithm according to the existing input and output data sets, in which the conditions are obtained. The algorithm is used as backpropagation algorithm and the linear least square estimation algorithm is used to adjust the conclusion parameters. In view of the characteristics of poor stability of general network and easy falling into local minima, the optimal square estimation of error is used to approximate the error function. Combined with the optimal square estimation algorithm, the convergence to the global minimum point in the parameter space is guaranteed, thus the training speed is improved and the learning time is shortened. Immune genetic algorithm [6-11] combines the advantages of immune algorithm and genetic algorithm. It is a genetic algorithm based on medical immune function. In genetic algorithm, immune mechanism can be fused, immune system mechanism can be imitated, and antigen and antibody recognition can be imitated. In addition, the process of combining, producing, effectively utilizing the diversity of immune system and memory ability makes the performance of genetic algorithm more efficient. Specific fuzzy immune genetic algorithm flow, as shown in figure 1.

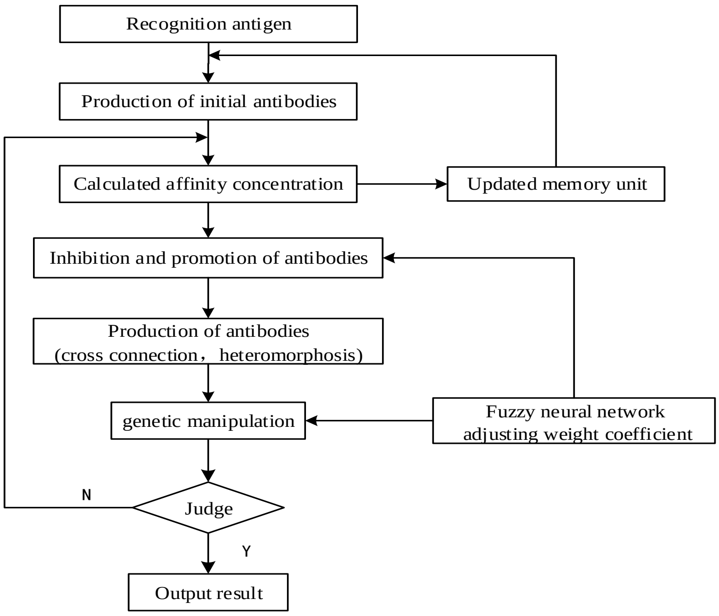

Fig. 1 Flow of fuzzy immune genetic algorithm

3.2 Specific steps

Step 1: To determine the antigen. The input objective function and constraint conditions are defined as antigens, and the memory function is used to determine whether such problems have been solved.

Step 2: To produce the initial antibody. Compare with step 1 to test whether the corresponding memory cell information exists. If there is, generate the corresponding initial antibody based on the memory cell. Otherwise, generate the initial antibody at random. In the feasible region, the initial antibody deficiency individuals are randomly generated by computer to generate antibodies.

Step 3: To calculate affinity and concentration. The fitness of genetic algorithm is reflected by the affinity between antibody and antigen, and the concentration of antibody is represented by the affinity of antibody. Compared with the control, the antibody with high affinity is selected into the storage unit of immune genetic algorithm.

Step 4: To inhibit and promote antibodies. In order to suppress and promote the antibody, the antibody with high affinity to the antigen was adopted, which has a high probability of being promoted. In order to prevent the antibody from losing its diversity, the weight coefficient is calculated by the fuzzy neural network to construct the excitation coefficient, which ensures the diversity of the antibody.

Step 5: To produce new antibodies. New antibodies are obtained by genetic operators such as selection operator, crossover operator and mutation operator.

Step 6: To adjust the weight coefficient adaptively. By using the Sugeno fuzzy neural network system to adaptively adjust the weight of the antibody as the coefficient of the concentration between antibodies. In the process of antibody renewal, the individuals with high antibody concentration will be inhibited and the individuals with low concentration will be promoted. It not only ensures the convergence of the algorithm, but also takes into account the diversity of antibodies. The fuzzy neural network structure [12] is shown in figure 2.

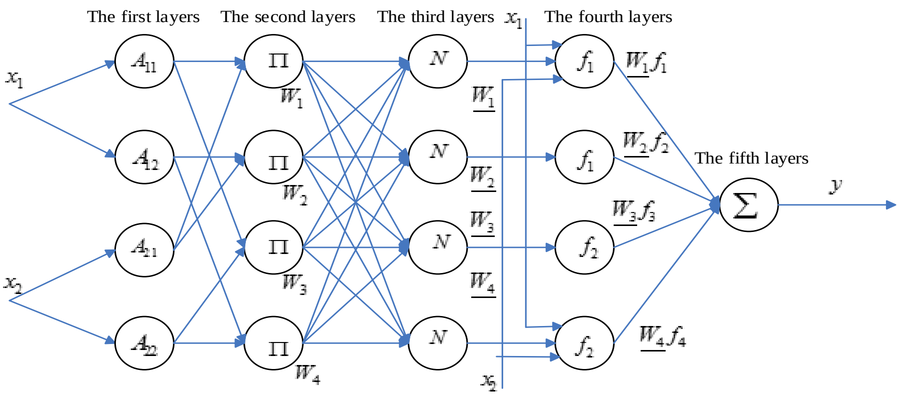

Figure 2 Fuzzy Neural Network Structure (Two input single output)

(1) The first layer is a fuzzy subset layer: each unit represents a fuzzy segmented subset whose membership function is an activation function. $A_{ij}=(x_i)(i=1,2;\ j=1,2)$

(2) The second layer is the combinatorial layer: each unit is marked with the symbol "$\prod$", which combines different fuzzy subsets of different input variables, and the arithmetic product of all input signals represents the output of each unit.

(3) The third layer: each unit is marked with the symbol "*N*", whose function is to normalize the excitation intensity of each rule, and calculate the normalized incentive intensity of rule *i* by using the following formula, where W_1 denotes the incentive intensity of the rule.

$$\underline{W_i}=\frac{W_i}{W_1+W_2+W_3+W_4}, \quad i=1,2,3,4$$

(4) The fourth layer: each unit is marked with the symbol "f", and the linear function f_i is an activation function. The input is the product of the activation function (x_1 and x_2) and the normalized excitation intensity $\underline{W_i}$, which is transmitted from the input of the network and the third layer to the fourth layer, and the output is the product of the activation function and the normalized excitation intensity:

$\underline{W_i} f_i=\underline{W_i}(p_i x_1+q_i x_2+r_i)$, $i=1,2,3,4$。 $\{p_i, q_i, r_i\}$ are the conclusion parameters for each unit of this layer.

(5) The fifth layer is the output layer, marked with the symbol "$\sum$", the total output of which may be expressed as:

$$y=\sum_i \underline{W_i} f_i=\frac{\sum_i W_i f_i}{\sum_i W_i}$$

Step 7: Genetic manipulation. This step can work in parallel with crossover and mutation operation. According to the excitation coefficient $\Delta\omega$ calculated by fuzzy neural network, the excellent antibodies can be preserved for crossover and mutation, and then new antibodies can be obtained. The antibody with good affinity is guaranteed.

(1) Crossover operator. Adaptive crossover probability P_c based on hamming distance can be used in the crossover process.

$$P_m = \begin{cases} \Delta\omega\left(\dfrac{f_{\max} - f}{f_{\max} - f_{avg}}\right)^{\frac{1}{GH(i,j)}}, & f \geq f_{avg} \\ k, & f < f_{avg} \end{cases}$$

, f is the value of higher fitness between two individuals; f_{avg} and $f_{\max}$ are the average fitness and maximum fitness of the current population respectively. $GH(i, j)$ is the hamming distance between two individuals, reflecting the degree of similarity between individuals.

(2)Mutation operator, $P_m = \begin{cases} \Delta\omega\left(\dfrac{f_{\max} - f}{f_{\max} - f_{avg}}\right), & f \geq f_{avg} \\ k, & f < f_{avg} \end{cases}$. Using adaptive non-uniform mutation operator, the mutation probability varies dynamically with excitation coefficient ($\Delta\omega$), which ensures the diversity of population.

Step 8: To output results. According to the affinity of the antigen and the concentration of the antibody, whether the antibody is stable or not, if the antibody is stable, the optimal result of the problem can be obtained, otherwise, the third step will be turned to.

4 Data Processing

According to the standard data collected by the above algorithm steps, the following processing is carried out: under the condition that the original feature space is not changed, a related feature subset is selected from the feature set. A new low-dimensional feature space is constructed with its more important features. Then, the useful information from the original feature item is extracted, the new feature item is established, and the new feature space is mapped to the low-dimensional feature space. The low dimensional space should be able to reflect the correlation information and semantic information of the high-dimensional feature items. After the above processing, the spatial dimension of the data can be reduced to less than one dimension and one hundred dimensions, and the immune genetic algorithm is used to optimize the solution.

The classification of standard categories is varied, and different classification standards correspond to different subsets of standards. For different subsets of standards, it is now the common practice in the world to divide formal standards and factual standards according to their formative mechanisms [13]. The data comes from 500 standard sets in the full text database system of national military standards. The expert survey method (Delphi method) is used. Firstly, the selection of experts in related fields is determined, the number of experts is 15-25, and the expert consultation table is designed. To highlight the relative importance of pairwise indicators, according to the evaluation feedback to constantly modify and improve. According to the result, the judgment moments of each index layer are constructed. Matrix, and calculate the relative weight of each index. The sampling data is normalized and the relevant standard eigenvector is transformed into multidimensional space. The immune genetic algorithm is used to optimize the solution to form an efficient service standard library. In this paper, the technical standard is taken as an example, the above problem can be solved by a hypothetical N-dimensional stereo coordinate system. Suppose $p(x_1, x_2, \ldots x_n)$ is a standard eigenvector, then the distance between the standard eigenvector and the objective function (antigen) F is $D(x, y) = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}$ .The standard library is a collection of all standard feature vectors, such as the sum of six secondary indicators, technical form, technical specification, standard update, standard service, data interface, etc. The characteristic vectors of each subset in the multidimensional space are set in the form of: basic standards, product standards, information identification, design standards, process standards, installation standards, inspection and test standards, packaging standards, handling standards, storage standards, delivery standards, maintenance standards, service standards, equipment and process equipment standards, safety standards infrastructure, energy standards, medical and health and occupational health standards, as well as environmental standards and other standards, according to the sequence of corresponding numbering A_i. Let the adaptive weight coefficient Δw be 1.8 (usually 1-2.5 with reference to the experience of text clustering experts).

Table 1 Technical standard data

Feature vector	Standard form （0.2115）	Technical manual （0.1780）	Technical order （0.2491）	Standard renewal degree （0.1165）	Standard service （0.1685）	Data interface （0. 0764）
A1	0.56	0.75	0.86	0.48	0.57	0.64
A2	0.81	0.85	0.69	0.63	0.61	0.45
A3	0.67	0.78	0.83	0.81	0.65	0.72
A4	0.92	0.56	0.74	0.62	0.71	0.51
A5	0.83	0.61	0.84	0.75	0.81	0.45
A6	0.78	0.66	0.82	0.91	0.87	0.28
A7	0.64	0.53	0.84	0.77	0.81	0.49
A8	0.81	0.38	0.91	0.59	0.56	0.70
A9	0.63	0.56	0.48	0.42	0.61	0.54
A10	0.34	0.59	0.71	0.49	0.29	0.47
A11	0.65	0.78	0.56	0.67	0.42	0.66
A12	0.75	0.45	0.12	0.64	0.58	0.12
A13	0.58	0.78	0.94	0.76	0.75	0.65
A14	0.85	0.84	0.87	0.58	0.84	0.37
A15	0.23	0.65	0.45	0.65	0.64	0.67
A16	0.56	0.43	0.85	0.45	0.87	0.38
A17	0.75	0.87	0.74	0.18	0.64	0.67
A18	0.64	0.86	0.68	0.58	0.89	0.45

5 Simulation experiment

MATLAB software [14] is used to verify the above technical standard data. The platform is a desktop Lenovo computer, which is mainly configured with Intel i5 8 GB memory.

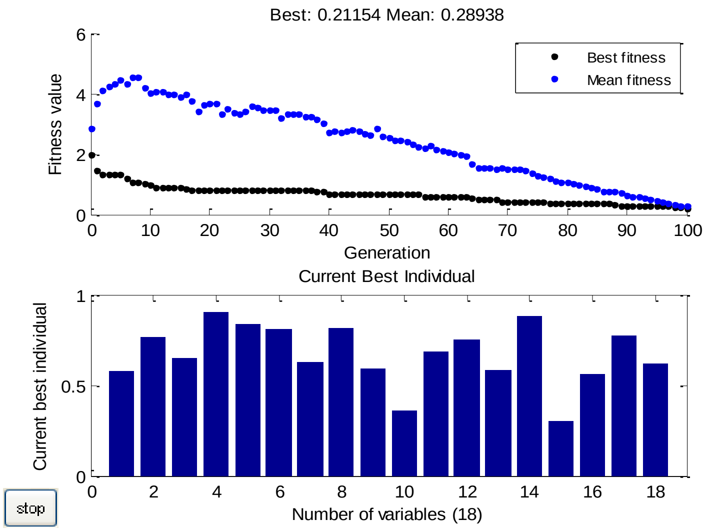

Figure 3 Standard form simulation data

The result of simulation is 0.2115. It can be seen that there are more standards of technical specification, and the data interface between the standards of standard knowledge base is only 0.0764, which should be considered in the next step of the construction of standard library. Through data simulation, all the standards can be grasped, which provides convenience for all kinds of work according to the standards, and provides data support for the subsequent updating of the standards and the standard association analysis.

6 Conclusion

With the advent of big data and industry 4.0, the standard system and standard resources in various industries need to be optimized and integrated. Based on the analysis of the status of standard resources, a fuzzy immune genetic algorithm (FIA) clustering model for optimal solution is proposed, and a service-oriented standard library is constructed through the optimization of the standard subset. Taking the technical standard as an example, the model is verified and analyzed by simulation. The results show that the standard eigenvector can be transformed into solving the optimization problem of multidimensional space, and the timeliness of standard clustering analysis is improved. It shows that the algorithm has good robustness, credibility and accuracy, and provides a solution to the standard resource clustering and integration of the standard system.

References

[1] Xi Hayun. About Drafting of Enterprise Products Standards [J]Modern Industrial Economy and Informationization 2015，3:99—100.

[2] Zhao Quanren. History of Standardization [M]. Beijing: China Standard Press，1993

[3]China Institute of Standardization. Study on the Construction of National Standard system [M]. Beijing: China Standard Press.

[4]Zhou Zhihua, Chen Shifu. Neural Network Ensemble [J]. Chinese J. Computers，2002, 25 (1)1-8.

[5]Li Guoyou, Zhou Qiaoling. Self-adaptive Neural Fuzzy-PID Control for Secondary Regulation Speed System [J]. Machinetool＆Hydraulics，2010, 38(21):95-98

[6]Sun Quanying, Wang Yilin, Du Xuwei. Application and Research of Genetic Algorithm in Mechanical Optimization [J].Journal of Ha R Bin University Of Science And Technology, 2015, 20(4):46-50.

[7]Cai Liangwei, Li Xia. Improvement on Crossover Operation of Genetic Algorithms [J].Systems Engineering and Electronics, 2006, 28 (6): 925-928.

[8]Pan Wei, Diao Huazong, Jing Yuanwei. An Improved Real-value Adaptive Genetic Algorithm [J].Control and Decision 2006,21(7): 792-795.

[9] Wei Shaoyuan, Wang Dongmei. Research in Vehicle Routing Problem Based on Hybrid Genetic Algorithm[J].Journal of Transportation Systems Engineering and Information Technology, 2006, 6(6): 50-52.

[10]LIAN Wanxiang, ZHU Canshi, HU Jianghua. Application of an Improved Genetic Algorithm to Transportation of Emergency Relief Materials [J]. China Safety Science Journal, 2013, 23 (5):172-175.

[11]Wu Yonggang , Xu Xiangxin , Liu Mei. Research on City Garbage Removal Routes Based on Genetic Algorithm [J]. Computer Simulation, 2012, 29(4): 259-262.

[12]Liang Yu. The Establishment of Non-linear Ship Model and the Design of ANFIS Ship Course Keeping Controller [D]. Dalian Maritime University, 2008.

[13]Ping Yang. The Reform of equipment purchase system and the Development of military Standards [M]. Beijing: National Defense Industry Press, 2015.

[14]Bao Ziyang, Yu Jizhou. Intelligent Optimization algorithm and its MATLAB example [M]. Beijing: Electronic Industry Press, 2016.