Neural Comput & Applic (2013) 22:187–193
DOI 10.1007/s00521-011-0694-3

ORIGINAL ARTICLE

Research of assembling optimized classification algorithm by neural network based on Ordinary Least Squares (OLS)

Xinzheng Xu · Shifei Ding · Weikuan Jia · Gang Ma · Fengxiang Jin

Received: 17 April 2011 / Accepted: 4 July 2011 / Published online: 22 July 2011

Abstract A new optimized classification algorithm assembled by neural network based on Ordinary Least Squares (OLS) is established here. While recognizing complex high-dimensional data by neural network, the design of network is a challenge. Besides, single network model can hardly get satisfying recognition accuracy. Firstly, feature dimension reduction is carried on so that the design of network is more convenient. Take Elman neural network algorithm based on PCA as sub-classifier I. The recognition precision of this classifier is relatively high, but the convergence rate is not satisfying. Take RBF neural network algorithm based on factor analysis as sub-classifier II. The convergence rate of the classifier algorithm is fast, but the recognition precision is relatively low. In order to make up for the deficiency, by carrying on ensemble learning of the two sub-classifiers and determining optimal weights of each sub-classifier by OLS principle, assembled optimized classification algorithm is obtained, so to some extent, information loss caused by dimensionality reduction in data is made up. In the end, validation of the model can be tested by case analysis.

X. Xu · S. Ding (✉) · W. Jia · G. Ma
School of Computer Science and Technology,
China University of Mining and Technology,
Xuzhou 221116, People's Republic of China
e-mail: dingsf@cumt.edu.cn

S. Ding
Key Laboratory of Intelligent Information Processing,
Institute of Computing Technology, Chinese Academy
of Sciences, Beijing 100080, China

F. Jin
Geomatics College, Shandong University of Science
and Technology, Qingdao 266510, China

Keywords Assemble learning · Neural network · Feature dimensionality reduction · Classifier · Ordinary Least Squares (OLS)

1 Introduction

Neural network [1, 2] is a nonlinear and adaptive information processing system that based on the intelligent computation of the Computer Network simulates biological neural network, which processes and memories information by simulating cranial nerve, and consists of large interconnect processing unit. The nature of pattern recognition [3, 4] can be regarded as an input/output classification system, while neural network can approximate any nonlinear system with arbitrary accuracy, which just shows the superiority of processing such problems. However, with the development of science and technology and the leap of information age, the growth and updates of data sets are faster, data dimensions are higher, and unstructured is more outstanding; Similarly, the identification accuracy is put forward a higher request. Today, the problems we have to process are mostly large sample data even mass data, and the results also required a higher accuracy; these present new challenges to the way of processing problems for us.

When neural network is used to process this kind of problems, the redundant raw data will occupy large memory spaces and waste computing time, leading to longer time for network training, hamper the convergence of training network, at finally, affect identification accuracy. So, we should take a series of preprocessing, which aims at improving the network's operation efficiency. Currently, proposing about 100 kinds of network models, which play their respective advantages in processing different problems, but the above all models, themselves are not perfect. Such as common Radial Basis Function (RBF), neural network [5] is a feed forward network model, and its advantage lies in the quick convergence and high training efficiency; however, the identification accuracy is unsatisfying to us. Elman regression neural network [6] is a feedback network, with feedback effects, identification accuracy is higher, but there is no ideal convergence speed.

Using assembling algorithm to process pattern recognition problems has achieved a good result [7–9]. This paper puts forward a new approach to complex identify problems, which is establishing neural network ensemble algorithm. First, the raw data is pre-processed, i.e., the data dimension is reduced. In order to simplify the network structure, the network training speed, convergence, and generalization capability needs to be improved. In the recognition process, the network assembling of RBF network and Elman is used in order to improve the network recognition accuracy. The algorithm consists of two sub-classifiers, one is the classifier I combined by the principal component analysis and Elman network organically, and one is classifier II combined by factor analysis and RBF network; It is a key problem to confirm each sub-classifier's weights, when two sub-classifiers are assembled into classification algorithm. This study adopts common least-square algorithm to study system synthetically and calculates all sub-classifiers' weights by solving the regression equation in order to get the optimal assembling model. This algorithm improves the operation efficiency of the network by reducing dimension, increases network recognition accuracy by the network assembling, and makes up for information loss caused by data dimension reduction to some extent at the same time. The new algorithm is established in order to provide a higher efficiency and a higher identification accuracy model for pattern recognition. Finally, the validity of the proposed algorithm is verified through the case analysis.

2 The feature dimension reduction

For neural network, the large number of features provides available information; meanwhile, they also increase the difficulty in processing problems. For the massive data, it is very important for analyzing, extracting useful information features, eliminating influences of the related or repeat factors, reducing the dimension of information features, while not impacting on solving problem. That is to say, it is very important to find a group of the most effective classification from the many features and so as to design the classifier effectively [10–12]. Then, reduced-dimension data as neural network's input make data reduced dimension with neural network organically. This paper adopts classical PCA and FA [13] as a reduced-dimension tool, they are all used to approach the approximation of covariance, estimate the correlations among the variables through the correlation coefficient matrix, and solve the eigenvalues and eigenvectors of the correlation coefficient matrix, and there are no correlations among the principal components and factors.

2.1 PCA reduced-dimension principle

PCA is a statistical analysis method that transforms multiple features indexes into a few composite indexes from the angle of features effectiveness, which find a few comprehensive factors to replace the original numerous variables, make these comprehensive factors as far as possible to reflect the information of original variables, and not related to each other, so as to achieve the simplified purpose. PCA is a linear combination that the principal component is expressed as variables, and the key is to explain the total variance of variables.

Let the samples contain p variables, $x_1, x_2, \ldots, x_p$ are analyzed by PCA, and then they are assembled into p comprehensive variables, namely:

$$\begin{cases} y_1 = c_{11}x_1 + c_{12}x_2 + \cdots + c_{1p}x_p \\ y_2 = c_{21}x_1 + c_{22}x_2 + \cdots + c_{2p}x_p \\ \cdots \\ y_p = c_{p1}x_1 + c_{p2}x_2 + \cdots + c_{pp}x_p \end{cases} \tag{1}$$

and:

$$c_{k1}^2 + c_{k2}^2 + \cdots + c_{kp}^2 = 1 \quad (k = 1, 2, \ldots, p) \tag{2}$$

y_1 is the maximum variance of mathematical expression (1) in all the linear combinations, y_2 next maximum, and so on, $y_\mathrm{p}$is the minimum. So the comprehensive factors $y_1, y_2, \ldots, y_p$ are, respectively, called the *first, second*, …, pth principal component of the original variables, their variances decline one by one. Choosing $m(m < p)$ principal components, the rate of the summation of the variances of first n principal components and the total variances are called Total Cumulative Variance Contribution Ratio (TCR) a.

$$a = \left(\sum_{i=1}^{m} \lambda_i\right) \Big/ \left(\sum_{i=1}^{p} \lambda_i\right) \tag{3}$$

In mathematical expression (1), a closes to 1, so the m principal components almost contain the information of the original variables $x_1, x_2, \ldots, x_p$, the number of variable reduce to m from p, and the m variables should contain most of the original information, which play a role of reducing dimension.

2.2 FA reduced-dimension principle

The FA's basic idea is to classify observation variables, the higher correlation ones are grouped together, and there is a low correlation for the variables in different class, so each class of variables actually represents a basic structure, namely a common factor. The problem is changed into trying to use the summation of the linear function of common factors and special factors to describe each variable of original observation. The FA's basic problem is the related coefficient between each variable to decide the factor loading. FA is the linear combination of each factor that expressed by variable and mainly explains covariance between variables.

FA model, let the observable random vector $X_i = x_1, x_2, \ldots, x_n$, unobservable vector $F_j = F_1, F_2, \ldots, F_m$,

$$X_i = \sum_{j=1}^{m} a_{ij} F_j + c_i \varepsilon_i \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m) \tag{4}$$

where $m < n$, a_{ij} is factor loading, it means the related coefficient between the ith variable and the jth factor and reflects the importance of the ith variable in the jth factor, F is called the common factor, they are jointly emergent factors in the all original observation variables expression, independent and unobservable theory variables. c_j means the loading of unique factor, and ε_i is the unique factor affecting X_i.

Let $A = (a_{ij})$, and A is called factor loading matrix, when A's structure is not easy to explain main factor that should do a series of rotating on A in order to make A approach 'The most concise structure standard', so that make the projection of each test vector as far as possible polarize into 1 and 0 in the new coordinate. So the optimal subsets of different variables are found from numerous factors, and the multivariable systems results and the influence of various factors on the system can be described by the information in the subsets so as to approach the purpose of reduced dimension.

3 Neural network ensemble

For the complex recognition problems, the single classifier is often difficult to achieve ideal recognition accuracy, and it also has deficiencies. The main idea of ensemble learning (EL) is to use multiple classifiers to solve the same question, and the purpose is to improve the recognition and generalization ability of learning algorithm more efficiently [16, 17]. The multiple assembling classifiers are adopted in order to improve the recognition accuracy, and it is the key problem how to confirm the weight of each classifier.

3.1 Sub-classifier design

Designing neural network sub-classifier, this study adopts neural network algorithm based on feature dimension reduction, and the low dimension data of reduced dimension are used to train the network as a network's input, then getting the improved neural network model, namely the sub-classifier, Fig. 1 is an algorithm flow chart of sub-classifier.

3.1.1 Elman neural network algorithm based on PCA

Elman regression neural network is based on the structure of back propagation (BP) neural network [14], the output of the hidden layer automatically connects to the input of hidden layer by the delay and memory of undertake layer, the auto-connection way makes it become sensitive to history state data, and the join of internal feedback network increases the ability that network processes dynamic information by itself. Through storing, the internal state makes it to have the function of mapping dynamic nature. So that classification instruments have time-varying ability. Firstly, the feature dimension of the raw data is reduced by PCA in order to decrease the network's input, thereby the Elman neural network classifier based on the PCA is built, noting as the sub-classifier I. The algorithm flow chart of sub-classifier I is shown in Fig. 2 [15].

The sub-classifier I increases the training speed and convergence by improving Elman neural network structure to some degrees and enhances the efficiency of the network to process problems. The algorithm has relatively high recognition accuracy; however, the convergence is not acceptable.

3.1.2 RBF neural network algorithm based on FA

RBF network is a kind of neural network model with simple structure, training forthright, wide application and has the features that the structure is self-adaptive determination and the output is not associated with initial value. The advantage is completion of the work of nonlinear learning algorithm by using the linear learning algorithm, while keeping the high precision of nonlinear algorithm, and having the best approximation and the global optimum, etc. There are two steps about RBF network training process, the first step: determining the link weights without teachers' learning, the second step: determining weights with teachers' learning. It

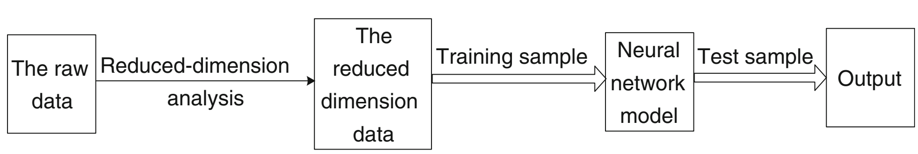

Fig. 1 Topology structure of neural network algorithm based on feature dimension reduction

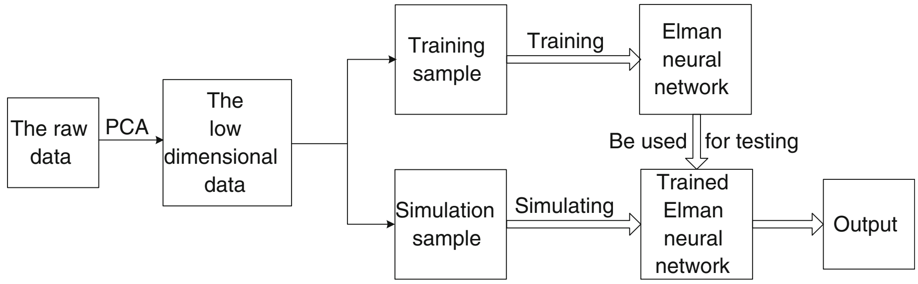

Fig. 2 Topology structure of sub-classifier I

is a key problem to confirm the quantity of hidden units, usually start training from 0, the hidden layer automatically increases the number of neuron by examining the error, repeats this process until it reaches an error requirements or maximum hidden layer cell number.

According to the own property of RBF, if RBF has overmuch network input, it is a difficult problem to confirm the number of neuron and that will affect network training. The raw data are reduced dimension by FA, then the reduced-dimension data are used for the input of network to build neural network classifier based on FA–RBF, noting for sub-classifier II, so that it can be easily design hidden layer, improve training speed, and save the operation time. The algorithm flow chart of sub-classifier II is shown in Fig. 3.

The sub-classifier II reduces network input, and it is useful to design network and simplify network structure, so that network convergence has faster speed and higher efficiency. This algorithm has made a lot of achievements in the control field, but made a little success in recognition and prediction field relatively.

3.2 Assembling optimization algorithm

The main task of neural network assembling optimization algorithm is to seek all sub-classifiers' weights and reduce the generalization error of network [16, 17].

3.2.1 The model of assembling optimization algorithm

The recognition accuracy of the sub-classifier based on Elman neural network is relatively ideal, but the training speed is not ideal; the training speed of the sub-classifier based on RBF is ascendant, but the recognition accuracy is not ideal. Therefore, the two sub-classifiers are assembled to learn, which can get a result of complement each other's advantages. In order to cope with the large sample, the raw data were reduced dimension when establishing sub-classifier, which results in a little loss to some degrees, while the assembling learning makes up for the impact of information loss in a sense. The Fig. 4 is a model of assembling learning. In this model, the two sub-classifiers are assembled to learn, the working principle is to input processing problem into two sub-classifiers, respectively, the each sub-classifier will get different results, these will be weighted by assembled processor, finally, the model outputs accurate recognition results.

Let Q is a processing problem, it is recognized by two sub-classifiers, the sub-classifier I gets a recognition result A, the sub-classifier II gets a recognition result B, and the relation of A and B is mutual independence. C^t, the output value of the problem, is made up of the weights of A and B, and let their weights be w_1, w_2, respectively, as follows:

$$C^t = Aw_1 + Bw_2. \qquad (5)$$

3.2.2 The optimizing of assembling optimization algorithm

Now, the key problem is how to determine the weights w_1 and w_2, of the two sub-classifiers for the above model (5), we can use Ordinary Least Squares (OLS) principle to evaluate the optimal estimation of the weights w_1, w_2.

For the assembling learning of the samples, the input training samples are recognized by two sub-classifiers, which can be gotten different results. For the ith training sample, let a_i means the recognition results of the

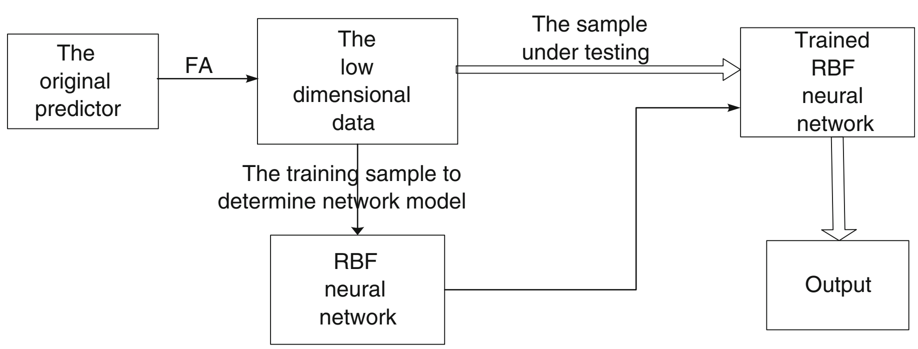

Fig. 3 Topology structure of sub-classifier II

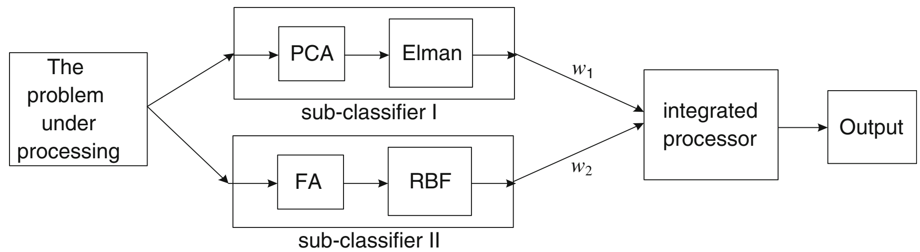

Fig. 4 Optimized classification algorithm assembled with two sub-classifiers

sub-classifier I, b_i means the recognition results of the sub-classifier II, c_i^t means the output of the assembling algorithm, and the actual value of sample be c_i^0, $i = 1, 2, \ldots, n$.

The actual value of problem is C^0, so the binary linear regression equation can be written:

$$C^0 = \gamma + A\alpha + B\beta \tag{6}$$

and

$$\min E = \sum_{i=1}^{n} \left(c_i^0 - E\left(c_i^0\right)\right)^2 = \sum_{i=1}^{n} \left(c_i^0 - \gamma - a_i\alpha - b_i\beta\right)^2 \tag{7}$$

In order to make E to approach the minimum value, α, β, γ should be satisfied:

$$\begin{cases} \frac{\partial E}{\partial \alpha} = \sum_{i=1}^{n} \left(C_i^0 - a_i\alpha + b_i\gamma - \gamma\right)A_i = 0 \\ \frac{\partial E}{\partial \beta} = \sum_{i=1}^{n} \left(C_i^0 - a_i\alpha + b_i\beta - \gamma\right)B_i = 0 \\ \frac{\partial E}{\partial \gamma} = \sum_{i=1}^{n} \left(C_i^0 - a_i\alpha + b_i\beta - \gamma\right) = 0 \end{cases} \tag{8}$$

Let

$$X = \begin{bmatrix} 1 & a_1 & b_1 \\ 1 & a_2 & b_2 \\ \vdots & \vdots & \vdots \\ 1 & a_n & b_n \end{bmatrix}_{n\times 3} \qquad \eta = \begin{bmatrix} \gamma \\ \alpha \\ \beta \end{bmatrix}_{3\times 1}$$

And the equation (8) can be simplified as matrix form:

$$X^T X\eta = X^T C^0 \tag{9}$$

According to OLS principle, η's estimated value of least squares can be solved:

$$\eta = \left(X^T X\right)^{-1} X^T C^0 \tag{10}$$

The regression coefficient α, β can be solved by η; furthermore, the weights w_1, w_2 can be solved also:

$$w_1 = \frac{\alpha}{\alpha + \beta}, \quad w_1 = \frac{\beta}{\alpha + \beta} \tag{11}$$

When the output value C^t closes to the actual value C^0 unlimitedly, the error E approaches minimum:

$$\min E = \sum_{i=1}^{n} \left(c_i^t - c_i^0\right)^2 = \sum_{i=1}^{n} \left(a_i w_1 + b_i w_2 - c_i^0\right)^2 \tag{12}$$

What this model shows obviously is if $w_1 = 1$, $w_2 = 0$, which represents that classifier I plays a role, and if $w_1 = 0$, $w_2 = 1$, which represents that classifier II plays a role. The trained assembling algorithm can produce the optimal weights, and the simulation samples can be classified by the built assembling classification algorithm.

3.2.3 The procedure of assembling optimization algorithm

When facing with the complex processing problem, neural network based on feature dimension reduction is used to design the sub-classifier, although the dimension reduction would lose some information, the corresponding recognition accuracy do not become lower. On the network operation efficiency, the sub-classifier reduces network input, which makes the network easy to be designed, simplifies the network structure, reduces the convergence steps, and increases the training speed of the network. As to there are some imperfection when the each sub-classifier process problem, let the each sub-classifier be assembled to study, and compute the weights of each sub-classifier by OLS algorithm, eventually, get a neural network classification algorithm based on OLS. The basic process of this assembling classification algorithm is as follows:

Step 1: standardizing the raw data, computing R, the mutual relation matrix of sample;
Step 2: computing eigenvalues and eigenvectors of R by the Jacobian method, computing CR, the contribution rate of variance, and TCR, the cumulative contribution rate of variance;
Step 3: doing orthogonal transformation about the standardized data, extracting principal component according to TCR, then go to Step 6;
Step 4: computing U,the orthogonal matrix of R, extracting the number of principal component according to TCR, decomposing matrix U, and getting U_0, the Component Matrix;
Step 5: Further computing the Component Matrix U_0, according to practical problems, rotating U_0 and calculating factor score, then go to Step7;

Step 6: designing Elman neural network structure according to the extracted principal component and practical problems, taking the low-dimensional data of Step 3 as the input of the network, the output of the network is A after training;
Step 7: designing input/output nodes of the RBF neural network according to the number of the extracted principal factors and practical problems, taking the low-dimensional data of Step 5 as the input of the network, the output of the network is B after training;
Step 8: getting equation (8) according to OLS algorithm and the formula (6);
Step 9: solving the coefficient of regression according to the formula (10), and further getting the weights A and B according to the formula (11), finally, obtaining the optimal assembling model.

4 The example analysis

The data set of the simulation experiment is the authority test data set in current machine learning field [18], and the data set is divided into training set and simulation set according to actual need. In order to evaluate the performance of assembling classifier expediently, we respectively compare the alone output of the each sub-classifier, namely the data set is recognized by Elman neural network classifier based on the PCA, RBF neural network classifier based on FA, and the assembling classifier of this paper.

Firstly, let $w_1 = 1$, $w_2 = 0$, namely the problem is recognized by the classifier I. The raw data are processed by PCA reduced-dimension analysis, extracting the opportune number of principal components, and reducing the input of network. Then, the low-dimensional training samples are trained by Elman network, and the low-dimensional simulation samples are used to test. The testing results are listed in the Table 1. Secondly, let $w_1 = 0$, $w_2 = 1$, then the classifier II is used to recognize the problems, and the recognition results are listed in the Table 1.

Finally, in accordance with the established algorithm process in this study, do a simulation experiment about the test data sets, and the final calculation results are listed in Table 1.

The conclusion can be seen from the comparison results of two sub-classifiers processing problems in the Table 1, the PCA–Elman is more than the FA–RBF in recognition accuracy, and the PCA–Elman has less error sum squares compared with FA–RBF, but the sub-classifier I has more convergence steps than the sub-classifier II in operating efficiency. However, the recognition accuracy of the assembling algorithm of the two sub-classifiers is obviously higher than the any one, and the error sum square is obviously less than the any one. The results show that the recognition accuracy of assembling algorithm has achieved ideal requirements; the established novel algorithm is effective and feasible.

5 Conclusions and discussion

This paper designs the RBF neural network classifier based on FA and the Elman neural network classifier based on PCA, the two sub-classifiers are all based on the neural network algorithm of the feature dimension reduction, reduce the feature dimension of the processing data, which can make the designing of network structure is more convenient, and improve the network operation efficiency. For the two sub-classifiers, the convergence speed of the RBF neural network based on FA is faster, but there are some shortcomings about the recognition accuracy, while the recognition accuracy of the Elman neural network based on PCA is relatively higher, but there are some weaknesses about the training speed. The two sub-classifiers are assembled to learn, which can realize complement each other's advantages, get a higher recognition accuracy, and make up for the information loss caused by data reduced dimension. The key problem about assembling learning is how to determine the weights of all sub-classifiers, and we use the OLS principle to build a regression model and determine the optimal weights value of all sub-classifiers by solving the regression model. The simulation experiment shows that there is higher recognition accuracy about the neural network sub-classifier based on feature dimension reduction and the neural network assembling algorithm of regulating the weights of the sub-classifier based on OLS.

This study is about the two sub-classifiers are assembled to learn, which can be easy to get the optimal solution of

Table 1 Comparison of every classifier's performance

Recognition algorithm	Weights	Training steps	Identify accurate rate (%)	Error sum squares
Sub-classifier I	0.7194	262	86	12.8243
Sub-classifier II	0.2851	25	72	27.3361
Assembling algorithm	–	–	96	0.8766

the weights, and get the optimal assembling classification algorithm. When faced with more complex problems, and that require multiple sub-classifiers assembling, how to determine the number of adopted sub-classifier, the less number is hard to satisfy the expected effect, and the large number could increase the complexity of system in designing and calculating; how to design the sub-classifier and which network should be used into the assembling, so as to how to calculate the weights of sub-classifiers, etc. These are all the challenging problems in assembling learning and the focus of the next work too.

Acknowledgments This work is supported by the Basic Research Program (Natural Science Foundation) of Jiangsu Province of China (No.BK2009093), the National Natural Science Foundation of China (No.60975039, and No.41074003), and the Opening Foundation of Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (No.IIP2010-1).

References

1. Mccllochw S, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biol 10(5):115–133
2. Ding SF, Jia WK, Su CY et al (2011) Research of neural network algorithm based on factor analysis and cluster analysis. Neural Comput Appl 20(2):297–302
3. Sun JX (2002) Modern pattern recognition. National University of Defence Technology Press, Changsha
4. Bian ZQ, Zhang XG (2000) Pattern recognition. Tsinghua University Press, Beijing
5. Moody J, Dkaren CJ (1989) Fast learning in networks locally-tuned processing units. Neural Comput 1(2):281–294
6. Elman JL (1990) Finding structure in time. Cogn Sci 14(2):179–211
7. Tang CS, Jin YH (2003) A multiple classifiers integration method based on full information matrix. J Softw 14(6):1103–1109
8. Sun L, Han CZ, Shen JJ et al (2008) Generalized rough set method for ensemble feature selection and multiple classifier fusion. Acta Automatica Sinica 34(3):298–304
9. Gu Y, Xu ZB, Sun J et al (2006) An intrusion detection ensemble system based on the features extracted by PCA and ICA. J Comput Res Develop 43(4):633–638
10. Ding SF, Jia WK, Su CY et al (2008) Research of pattern feature extraction and selection. Proc seventh Int Conf Mach Learn Cybernetics 1:466–471
11. Ding SF, Jia WK, Su CY et al (2008) A survey on statistical pattern feature extraction. Lect Notes Artif Intell 5227:701–708
12. Foman G (2003) An exnetsive empirical study of feater selection metrics for text classification. J Mach Learn Res 3:1289–1305
13. Johnson RA, Wichern DW (2007) Applied multivariate statistical analysis, 6th edn. Prentice Hall, Englewood Cliffs
14. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representation by back-propagating errors. Nature 3(6):533–536
15. Ding SF, Jia WK, Su CY et al (2008) PCA-based Elman neural network algorithm. Lect Notes Comput Sci 5370:315–321
16. Zhou ZH, Chen SF (2002) Neural network ensemble. Chinese J Comput 25(1):1–8
17. Liu Y, Yao X (1999) Ensemble learning via negative correlation. Neural Netw 12(10):1399–1404
18. http://archive.ics.uci.edu/ml/machine-learning-databases/parkinsons/[OL].2009.3