TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.5, May 2014, pp. 3979 ~ 3986
DOI: http://dx.doi.org/10.11591/telkomnika.v12i5.5005

An Improved Method of SVM-BPSO Feature Selection Based on Cloud Model

Ji-zhen Li*[1], Xiang-ru Meng[1], Jing Wen[2], You Xu[1]
[1]School of Information and Navigation, Air Force Engineering University, Xi'an 710077, China;
[2]Army Officer Academy, Hefei 230031, China
*Corresponding author, e-mail: lijizhen1986@126.com, xrmeng@126.com, ellywen2008@163.com, youxu2013@126.com

Abstract

An improved method of SVM-BPSO feature selection based on cloud model is proposed to solve the local deadlock problem of the current SVM-BPSO methods. This method uses the Wrapper evaluation strategy, making SVM as the classifier, BPSO algorithm is adopted to conduct a whole search in the feature space to seek out the optimal feature subset from the classification results of SVM. The inertia weight of BPSO algorithm is adjusted by Cloud Model intelligently and self-adaptively, the whole and local searching capability of SVM-BPSO feature selection algorithm get balanced, and prevent it into the local deadlock effectively. Analysis on simulations shows that the improved method of SVM-BPSO feature selection based on Cloud Model can be able to dap out from the local optimum with a faster convergence rate, and shows good experimental effect.

Keywords: *feature selection, support vector machines (SVM), binary particle swarm optimization (BPSO), cloud model*

1. Introduction

Feature selection, which aims at the high-dimensional data objects, is able to search the smallest feature subset which could identify the target effectively from a set of features without losing the intrinsic value of the date and to remove the unrelated redundancy attribute, and which could increase the quality of the data and reduce the feature space dimension, and could solve the issue of the curse of dimensionality effectively and increase classification precision and efficiency. It is a hot issue in the fields of internet security, pattern recognition and data mining etc. The result of the feature selection directly impacts on the classification precision and generalization performance of the classifier, which nature is an issue of combination optimization shows as formula (1) [1],

$$P = \arg\max\{F(X) : X \subseteq Y\}, s.t. |X| \le n. \quad (1)$$

In this formula, "n"is the upper limit number of the selected features, "Y" is the combination of the features, "X" is the selected feature subset, "$F(X)$" is the evaluation criterion. It shows that the feature selection needs to solve two issues as follows: one is the selection of evaluation criteria "$F(X)$" and another is the selection of algorithm.

The method of the feature selection can be partition as the Filter type and the wrapper type based on the evaluated strategy. The difference between these two types is whether using the learning algorithm or not in the process of the machine construction to evaluate the optimal feature subset [2]. This paper adopts the Wrapper evaluation strategy, making SVM as the classifier, BPSO algorithm is adopted to conduct a whole search in the feature space to seek out the optimal feature subset from the classification results of SVM. An improved method of SVM-BPSO feature selection based on cloud model is proposed to solve the local deadlock problem of the current SVM-BPSO feature selection methods. The inertia weight ω of BPSO algorithm is adjusted by cloud model intelligently and self-adaptively, the whole and local searching capability of SVM-BPSO feature selection algorithm get balanced, and prevent it from the local deadlock effectively.

Received November 1, 2013; Revised December 21, 2013; Accepted January 10, 2014

2. Work Related

This section contains a brief summary of SVM and BPSO, shows the existing research work of SVM–BPSO feature selection algorithm.

2.1. Support Vector Machine Classification Algorithm

Support Vector Machine (SVM) is a data mining method based on Statistical Learning Theory and with the principle of Structural Risk Minimization [3], which has very strong learning ability and generalization performance, and it can deal with pattern recognition and regression problems effectively.

SVM is proposed on the basis of searching for the Optimal Hyperplane which meets the requirements of classification [4]. To the nonlinear problem, SVM maps the input vector to a high-dimensional feature space by adopting the appropriate kernel function

$$K(\mathbf{x}_i, \mathbf{x}_j) = \left(\phi(x_i) \bullet \phi(x_j)\right),$$

Thus converted the nonlinear into linear problem to seek the Optimal Hyperplane [5]. At present, the most frequent research of the kernel function mainly contains Polynomial Function (PF), Radial Basis Function (RBF) and the Sigmoid Function (SF) [6], etc.

2.2. The Binary Particle Swarm Optimization Algorithm

Particle Swarm Optimization (PSO) [7] algorithm is a type of Evolutionary Computation technology which was invented in 1995 by Dr Kennedy and Dr Eberhart who were inspired by birds feeding behavior [8]. It is also a type of intelligent optimization algorithm that can solve the optimization problem of continuous function [9].In order to solve the problem of discrete, the Binary Particle Swarm Optimization (BPSO) algorithm was proposed in 1997 by Kennedy and Eberhart base on the standard PSO algorithm [10].

2.3. The Existing SVM - BPSO Algorithm Analysis

The SVM-BPSO algorithm is a kind of typical Wrapper feature selection algorithm. The literature [11] proposes a novel object detection method, namely the BPSO-SVM-based detection algorithm that combines BPSO and SVM techniques to cope with feature selection issue for object detection under complex scenarios, the feature dimensionality is reduced and classification performance of the SVM classifier is greatly enhanced [11]. The literature [12] proposed a BPSO-SVM wrapper mode feature selection algorithm, a number of feature subsets was randomly generated at first, and then BPSO algorithm searched the feature space guided by the result of SMVs'10-fold crossover validation, after numbers of iteration, the best fitness feature subset was selected out to train the predictor, the experimental results show the effectiveness of the proposed algorithm [12]. In order to increase the convergent speed and to improve the overall searching ability of the algorithm, literature [13] proposed a PSO method with adaptive inertia weight by the change of the number of iterations based on the analysis of inertia weight global best fitness of the PSO, the improved PSO increases the ability to avoid local optimum, and then a speaker recognition method using this improved algorithm to train SVM is presented, the experimental results show that the presented SVM method optimized by PSO for speaker recognition can achieve higher recognition accuracy and higher recognition speed [13].

In the above research, the parameters optimization setting of the SVM-BPSO algorithm becomes an important way to improve the performance of algorithm, and it also becomes a hot research of the SVM-BPSO algorithm. Among them, the parameter optimization methods of SVM mainly include Cross Validation method and Particle Swarm Optimization algorithm, etc. The optimization set of inertia weight ω is the most important thing for the BPSO algorithm, and the inertia weight ω is the most important parameter of the BPSO algorithm, the implementation effect of the algorithm largely determined by the selection of ω and its scope in [0, 1]. The setting of ω balance the performance of global search and local search of the BPSO algorithm, the value of ω need to be increased to improve the speed of particle and enhance the global search ability of the algorithm when the particle is far away from the global optimal solution, smaller ω will enhance the local search ability of the algorithm, so the value of ω need to be decreased to reduce the speed of particle for local meticulous search when it is close to the global optimal solution [14]. The common methods for changing weight include Linear Decreasing Weight, Adaptive Weight and Random Weight, etc.

The inertia weight ω with dynamic adjustment is quite important to the effect of the algorithm performance according to the above analysis of the existing SVM-BPSO algorithm. The standard PSO algorithm decreases the value of ω to adjust the performance of the algorithm, but it is easy for the "cross shock" phenomenon to happen near the global optimal solution, which searching for the optimal solution among the local space, making the algorithm into the local extremum and affecting the convergence efficiency, increasing the difficulty of global optimization algorithm [14]. In order to solve this problem, literature [13] and [14] improved the inertia weight ω respectively, made ω have the function of the dynamic adaptive adjustment, which improved the performance of the algorithm in a certain degree. However, the methods of above were lack of intelligent mechanism for improving ω , which couldn't meet the requirement of adjusting the ω intelligently, self-adaptively and dynamic, and the performance of the algorithm was restricted to improve further in a certain degree.

3. The Improved SVM - BPSO Feature Selection Methods Based on Cloud Model

In order to solve the above problems, this section proposes the improved method of SVM-BPSO feature selection based on cloud model, and also enhances the performance of SVM-BPSO feature selection algorithm further. In this improved method, the inertia weight ω of BPSO algorithm is processed fuzzily based on cloud model according to the relationship between the fitness of the current particle and the average fitness of all the current particles, so as to meet the requirement of adjusting the ω intelligently, self-adaptively and dynamic, which prevent the algorithm into local deadlock effectively, and improve the overall performance of the SVM-BPSO feature selection algorithm. In this section, the theory of cloud model is briefed first, then describes the improved algorithm in detail, and demonstrates the feasibility with theoretical analysis, the related parameters of the improved algorithm is analyzed and set.

3.1. Cloud Model Theory

Cloud Mode is used to solve the uncertainty conversion between the qualitative concept and the quantified value based on fuzzy logic and probability, which is proposed by Academician Li Deyi, which provides a new research method for uncertain artificial intelligence [15]. Cloud Mode has been widely used and got better effect in the aspects such as intelligent control, data mining and fuzzy evaluation.

On the assumption that the collection $U = \{x\}$ is a quantitative domain which represented by numerical, C_n is the qualitative concept of U , if the quantitative value $x \in U$ is a random implementation of the qualitative concept C_n, the confirmable degree $\mu \in [0,1]$ from x to C_n is a random number with stable tendency, which shows as $\mu : U \to [0,1], \forall x \in U \; x \to \mu$, and the distribution of the x in the domain U is called cloud, every (x, μ) is called a cloud droplet [16]. There are three numerical characteristics $N^3(Ex, En, He)$ that can describe the Normal Cloud Model.

3.2. The Description of the Improved Method

The improved method of SVM-BPSO feature selection based on the Cloud Model use the SVM to train the training sample firstly for getting the final SV sets, and then find the final optimal feature sets by the BPSO algorithm from the selected SV sets, the final classified result which tested by putting the test sample into the optimal features will be got. The frame of the improved method is shown as Figure 1.

The innovation of this method is to adjust the inertia weight ω intelligently and self-adaptively by the BPSO algorithm which based on the theory of the cloud model, and prevent the improved method from getting into the local deadlock by adjusting the ω intelligently, and increase the whole performance of the SVM-BPSO feature selection algorithm. The improved mechanism of the inertia weight ω is as formula (2) shows.

$$\omega = \begin{cases} \omega_{\max} * \mathrm{FCG}\left(\dfrac{f_{avg}}{f},\ \mathrm{Ex},\ \mathrm{En},\ \mathrm{He}\right), & f_{avg} \le f \\ \omega_{\max}, \quad f_{avg} > f \end{cases} \tag{2}$$

In the formula, $\omega_{\max}$ is the maximum of the inertia weight, f is the fitness value of the current particle, f_{avg} is the average fitness value of all the current particles, $\mathrm{FCG}(f_{avg}/f,\ \mathrm{Ex},\ \mathrm{En},\ \mathrm{He})$ is the coefficient of the inertia weight which is treated fuzzily by the Cloud Model. It can be seen from the formula (2) that the inertia weight will be diminished with:

$$\omega = \omega_{\max} * \mathrm{FCG}(f_{avg}/f,\ \mathrm{Ex},\ \mathrm{En},\ \mathrm{He})$$

in order to protect the particle when the function value of the current objective is better than the average objective $f_{avg} \le f$, and the inertia weight will be increased with $\omega = \omega_{\max}$ in order to make the particle closer to the better search zone When the function value of the current objective is worse than the average objective $f_{avg} > f$. In concluded, this mechanism can not only increase the efficiency of the algorithm for searching, but also balance the ability of the BPSO algorithm for the whole search and the local search. It also can prevent the BPSO algorithm from getting into the local deadlock and increase the performance of the SVM-BPSO feature selection algorithm.

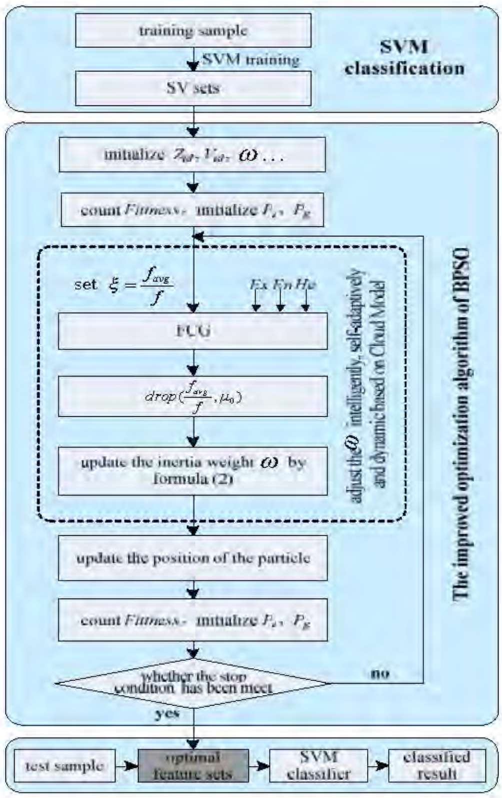

Figure 1. Frame of the Improved Method of SVM-BPSO Feature Selection Based on Cloud Model

To sum up, the table 1 gives the description of the improved method of the SVM-BPSO feature selection based on Cloud Model.

Table 1. Description of the Improved Method of SVM-BPSO Feature Selection Based on Cloud Model

Input: Set the original sample as $S=\{(\mathbf{x}_i,y_i)|\mathbf{x}_i\in R^d, y_i\in\{1,2,\cdots p\}, i=1,\ldots,l_s\}$ (p is the number of classes, l_s is the number of supportive vectors), and set the population scale as n, the maximum quantity of iteration as t, the fitness function as *Fittness*, and the threshold value as *Th*.

Output: The sample after feature selected $S'=\{(\mathbf{x}_i,y_i)|\mathbf{x}_i\in R^{d'}, y_i\in\{1,2,\cdots p\}, i=1,\ldots,l_s\}$.

Step1 Initialize the position, speed and $\omega_{\max}$ etc. of the every disperse particle from the population.

Step2 Solve the optimal position of the particle that can be researched so far and the optimal position of all the particle swarm that can be researched so far by the fitness function *Fittness*.

Step3 Set $\xi=\frac{f_{avg}}{f}$ and input it to the FCG, then resulted as $drop(\frac{f_{avg}}{f},\mu_0)$.

Step4 Update the inertia weight by formula (2).

Step5 Put the inertia weight ω updated into the particle updating formulas of BPSO algorithm in order to update the velocity and position of the particle.

Step6 Compare the particle fitness value to the ever best position it has been, then update the optimal position of the particle that can be researched so far and the optimal position of all the particle swarm that can be researched so far.

Step7 If the stop condition has been meet, stop searching and output the result S', otherwise go back to the Step3 and keep searching.

3.3. The Setting and Analysis of the Parameters

(1) The coding mode for the particle

The substance of the feature selection is to create a feature subset with the number M (M $\le$ N) which selected from the feature set with sample number as N. Code the feature set firstly and define every feature as a particle binary variable, the dimensionality of the particle space is depend on the dimensionality of the original feature set. If the ith bit is 1, well then the ith feature is selected, otherwise it is not selected. Each particle represents a different feature subset, the position component values of each particle corresponds the status of each feature subset in the feature combination. For instance, the particle p = (0100010) means the second and the sixth feature are selected, that is to say the feature subset is {2, 6}.

(2) The setting of the parameters

The main parameters of SVM: The parameter of the SVM only contains the Kernel Function and Penalty Factor. The type of Kernel Function decides the mapping way from the input space to the feature space. The Penalty Factory is used to balance the training error and the complexity of the model. The Radial Basis Function (RBF) has a strong learning ability which only needs to optimize one parameter called Kernel Bandwidths σ, so the RBF is chosen as the Kernel Function for the SVM, the mathematical expression of the Kernel Function is

$$K(\mathbf{x}_i,\mathbf{x}_j)=\exp(-\|\mathbf{x}_i-\mathbf{x}_j\|^2/2\sigma^2).$$

Kernel bandwidth σ is the average distance between the heterogeneous support vector of the support vector set which shows as $\overline{d}=(\sum_{i=1}^{l_{sv+}}\sum_{j=1}^{l_{sv-}}\|x_i-x_j\|)/l_{sv+}l_{sv-}$, l_{sv+} and l_{sv-} are the quantity of the support vector $\mathbf{x}_{sv+}$ and $\mathbf{x}_{sv-}$ which belongs to the two kinds of samples. The Penalty Factory is assumed as 100. Parameters optimize of the SV set training is managed by the 5 fold cross. The detailed settings of the main parameters are showed as Table 2.

The main parameters of BPSO: The most important for BPSO algorithm is to set the fitness function *Fittness* and the inertia weight ω. The fitness function of this method is confirmed overall consideration from three factors such as the classified precision rate *accuracy*, the original feature dimensionality *feature_dim* and the selected feature subset dimensionality *feature_new*. The fitness function is shown as formula (3).

$$Fitness=accuracy/(3*100)+1/(3*feature\_dim)+1/(3*feature\_new) \quad (3)$$

The adjustment mechanism of the inertia weight ω for the BPSO algorithm has been described detailed before as formula (2). In addition, the setting of other main parameter of BPSO algorithm is shown as Table 2.

The main parameters of CG: Because the Normal Cloud Model is described by numerical characteristics $N^3(Ex, En, He)$, so the main settings of the parameters for Cloud Model are the values of the Expectation Ex, the Entropy En and the Excess Entropy He. The Half Drop Cloud Mode is chosen for this improved method because of the inertia weight ω is monotone decreased between [0, 1], and the Expectation $Ex = 0$, the Entropy $En = 1/3$ (cloud width $\omega = 1$), the Excess Entropy $He = 0.1$ which is decided according to the experience. The main parameters of the CG are shown as Table 2.

Table 2. Main Settings for the Parameters of the Algorithm

The main parameters of SVM		**The main parameters of BPSO**		**The main parameters of CG**	
Parameter name	The values	Parameter name	The values	Parameter name	The values
Type of Kernel Function	RBF	Fitness function	F =(9)	Type of CG	Half Drop Cloud Mode
Kernel bandwidth σ	$\sigma=\bar{d}$	Inertia weight ω	ω =(8)	Ex	Ex =0
Penalty Factor C	C=100	Learn Factor	$C_1=C_2=2$	En	En =1/3
L- fold cross	L=5	Maximum quantity of t	t=100	He	He=0.1

4. The Experimental Analysis

4.1. The Pretreatment of the Data

The pretreatment mainly contains the selection of the data source and the format normalization of the data. This paper uses the DARPA evaluation data sources, which mainly contains four kinds of internet attack such as DoS, Probe, R2L and U2R, and simulate the internet security event with above attacks. This paper normalizes the data sample by using the following standard format $x_{new\,i} = (x_i - min\{x_i\})/(max\{x_i\} - min\{x_i\})$ in order to balance the availability of every feature in the sample, and mapping the values of the training sample and test sample into the interval $[0,1]$, $max\{x_i\}$ and $min\{x_i\}$ are the maximum and minimum of the ith feature x_i. The data sample is selected from the original sample by the uniformly-spaced collection method to ensure the universality of the data, the experimental data sample include one training sample and one test sample, every record has forty one features. The details of each data sample show as the Table 3.

Table 3. Structure of Experimental Data Sample

Type of data	Training sample	Test sample
Normal	500	1800
DoS	300	1000
R2L	100	700
Probe	270	1400
U2R	30	100
Total	**1200**	**5000**

4.2. The Results and Analysis of the Experiment

After the pretreatment of the date, do the feature selection by using the algorithms which were mentioned in the literatures [11-13], the standard SVM-BPSO algorithm unimproved by the Cloud Model and the improved method of SVM-BPSO feature selection based on Cloud Model separately, then classify the test sample and compare the feature dimensionality, classified precision and the time ratio (the ratio of other algorithm's time cost and the improved algorithm's). Assume the scale of the particle population as 10, and other parameters were set as chapters 3.3. The results of the compare for these algorithms were shown as Table 4.

Table 4. Experimental Results of Algorithm in this Paper Compared with others

Type of algorithm	Feature dimensionality	Classified precision	Time ratio
Original sample	41	97.39%	—
Algorithm in literature [11]	14	**99.36%**	465.56
Algorithm in literature [12]	21	96.91%	22.63
Algorithm in literature [13]	23	97.85%	1.13
Standard SVM-BPSO	17	98.90%	0.74
Algorithm in this paper	**11**	98.25%	**1**

As the conclusion of the Table 4, the algorithm in literature [11] trained the classifier by using the method of 5-fold cross validation and picked the average of 5 times SVM classified results as the classified precision rate, that lead to the time cost too long but with lower feature dimensionality and the highest classified precision. The time cost of the algorithm in literature [12] is less than the algorithm in literature [11], because it does not consider the optimal settings for the parameters, but it increased the feature dimensionality and reduced the classified precision ratio in a certain extent. The literature [13] proposed a PSO method with self-adaption the inertia weight to train the SVM, although it increased the classified precision and the efficiency, the feature dimensionality is worse than other algorithms. The experimental curve results of the improved method compared with others showed as Figure 2.

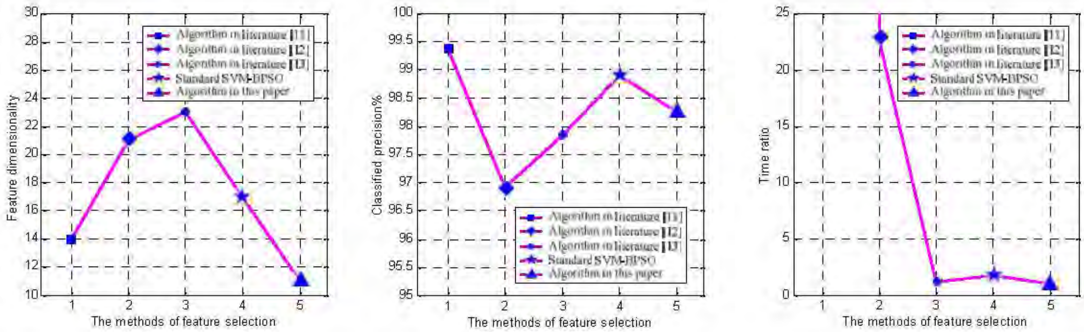

Figure 2. Experimental Curve Results of the Improved Method Compared with others

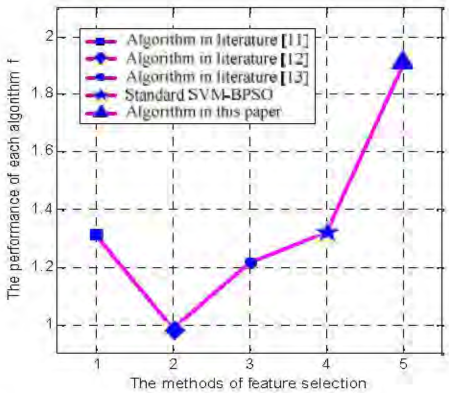

Figure 3. Performance Curve of the Improved Method Compared with others

Both the Table 4 and the Figure 2 shows a conclusion that the improved method of SVM-BPSO feature selection based on Cloud Model can reduce the feature dimensionality effectively while keeping the classified precision and the response time of the algorithm, it could solve the local deadlock problem effectively. Formulize the performance of the algorithm as formula (4).

$$f = accuracy/(3*100) + 41/(3*feature\_new) + 1/(3*time\_ratio) \quad (4)$$

accuracy is the ratio of classified precision, *feature_new* is the original feature dimensionality, *time_ratio* is the cost time ratio of the algorithm. The performance curves of each algorithm as Figure 3 showed, so the improved method of SVM-BPSO feature selection based on Cloud Model is better than others.

5. Conclusion

This paper proposed an improved method of SVM-BPSO feature selection based on the Cloud Model, which still made the SVM as the classifier, and selected the optimal feature subset by the BPSO algorithm. The inertia weight ω of BPSO algorithm is adjusted by Cloud Model intelligently, self-adaptively and dynamic which aims at the issue of the local deadlock with the BPSO algorithm, and the whole and local searching capability of SVM-BPSO feature selection algorithm get balanced. Compared the experimental results of the improved algorithm with the current SVM-BPSO feature selection algorithms, it can be seen that the improved method of SVM-BPSO feature selection based on the Cloud Model could solve the issue of the local deadlock effectively and the whole performance get better. The improved method of SVM-BPSO feature selection based on the Cloud Model also provide a new model for the feature selection research of the hot fields such as the Model Indentify, Internet Security and Data Mining etc.

References

[1] Zhu Shuxin, Hu Bin. Hybrid feature selection based on improved genetic algorithm. *Telkomnika*. 2013; 11(4): 1725-1730.
[2] Shuzlina Abdul-Rahmana, Azuraliza Abu Bakara, Zeti-Azura Mohamed-Hussein. An intelligent data pre-processing of complex datasets. *Intelligent Data Analysis*. 2012; 16: 305-325.
[3] Ding Shi-fei, Qi Bing-juan, Tan Hong-yan. An Overview on Theory and Algorithm of Support Vector Machines. *Journal of University of Electronic Science and Technology of China*. 2011; 40(1): 2-10.
[4] V Vapnik. Statistical Learning Theory. New York: Wiley. 1998.
[5] Ding Junping, Cai Wandong. An Improved Support Vector Machine Technology for Meta-Information Classification. *Journal of Xi'an Jiaotong University*. 2011; 45(8): 37-42.
[6] Jun Tan, Xingshu Chen, Min Du. An Internet Traffic Identification Approach Based on GA and PSO-SVM. *Journal of Computers*. 2012; 7(1): 19-29.
[7] Kennedy J, Eberhart RC. *Particle swarm optimization*. Proc of the First IEEE International Conference on Neural Networks. Perth, Australia: IEEE Press, 1995; 1942-1948.
[8] Kennedy J, Eberhart RC. A discrete binary version of the particle swarm algorithm [A]. IEEE International Conference on Systems, Man, and Cybernetics. Orlando, FL: IEEE Press. 1997; (5): 4104-4109.
[9] Li-Yeh Chuang, Cheng-Hong Yang, Sheng-Wei Tsaic. Complementary distribution BPSO for feature selection. *Intelligent Data Analysis*. 2012; 16: 183-198.
[10] He Jie, Guo Hui. A modified particle swarm optimization algorithm. *Telkomnika*. 2013; 11(10): 6209-6215.
[11] Pan Hong, Li Xiao-bing, Jin Li-zuo. A Binary Particle Swarm Optimization and Support Vector Machine-based Algorithm for Object Detection. *Journal of Electronics & Information Technology*. 2011; 33(1): 11-121.
[12] Qiao Li-yan, Peng X i-yuan, Peng Yu. BPSO-SVM Wrapper for Feature Subset Selection. *Acta Electronica Sinica*. 2006; 34(3): 496-498.
[13] Li Ming, Zhang Yong, Li Jun-quan. Application of Improved PSO-SVM Approach in Speaker Recognition. *Journal of University of Electronic Science and Technology*. 2007; 36(6): 1345-1349.
[14] Wu Ding-hai, Zhang Pei-lin, Li Sheng. Adaptive double particle swarms optimization algorithm based on chaotic mutation. *Control and Decision*. 2011; 26(7):1083-1086.
[15] Li Hailin, Guo Chonghui, Qiu Wangren. Similarity Measurement between Normal Cloud Models. *Acta Electronica Sinica*. 2011; 39(11): 2561-2567.
[16] Wang Shang-Guang, Sun Qi-Bo, Zhang Guang-Wei. Uncertain QoS-Aware Skyline Service Selection Based on Cloud Model. *Journal of Software*. 2012; 23(6): 1397-1412.