A Hybrid Classifier Based on Rough Set Theory and Support Vector Machines*

Gexiang Zhang[1], Zhexin Cao[2], and Yajun Gu[3]

[1] School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031, Sichuan, China
gxzhang@ieee.org
[2] College of Profession and Technology, Jinhua 321000 Zhejiang, China
[3] School of Computer Science, Southwest University of Science and Technology, Mianyang 621002 Sichuan, China

Abstract. Rough set theory (RST) can mine useful information from a large number of data and generate decision rules without prior knowledge. Support vector machines (SVMs) have good classification performances and good capabilities of fault-tolerance and generalization. To inherit the merits of both RST and SVMs, a hybrid classifier called rough set support vector machines (RS-SVMs) is proposed to recognize radar emitter signals in this paper. RST is used as preprocessing step to improve the performances of SVMs. A large number of experimental results show that RS-SVMs achieve lower recognition error rates than SVMs and RS-SVMs have stronger capabilities of classification and generalization than SVMs, especially when the number of training samples is small. RS-SVMs are superior to SVMs greatly.

1 Introduction

For many practical problems, including pattern matching and classification [1,2] function approximation [3], data clustering [4] and forecasting [5], Support Vector Machines (SVMs) have drawn much attention and been applied successfully in recent years. The subject of SVM covers emerging techniques that have been proven successful in many traditionally neural network-dominated applications [6]. An interesting property of SVM is that it is an approximate implementation of the structure risk minimization induction principle that aims at minimizing a bound on the generation error of a model, rather than minimizing the mean square error over the data set [6]. SVM is considered as a good learning method that can overcome the internal drawbacks of neural networks [7,8,9]. Although SVMs have strong capability of recognizing patterns and good capabilities of fault-tolerance and generalization, SVMs cannot reduce the input data and select the most important information.

Rough Set Theory (RST) can supplement the deficiency of SVMs effectively. RST, introduced by Zdaislaw Pawlak [10] in his seminal paper of 1982, is a new mathematical approach to uncertain and vague data analysis and is also a new

* This work was supported by the National EW Laboratory Foundation (No.NEWL51435QT22 0401).

L. Wang and Y. Jin (Eds.): FSKD 2005, LNAI 3613, pp. 1287 – 1296, 2005.

fundamental theory of soft computing [11]. In recent years, RST becomes an attractive and promising research issue. Because RST can mine useful information from a large number of data and generate decision rules without prior knowledge [12,13], it is used generally in many fields [10-16], such as knowledge discover, machine learning, pattern recognition and data mining. RST has strong capabilities of qualitative analysis and generating rules, so it is introduced to preprocess the input data of SVMs so as to extract the key elements to be the inputs of SVMs.

In our prior work, an Interval-Valued Attribute Discretization approach (IVAD) was presented to process the continuously interval-valued features of radar emitter signals [17]. RST was combined with neural networks to design rough neural networks and experimental results verify that rough neural networks are superior to neural networks [18]. Unfortunately, neural networks have some unsolved problems, such as over-learning, local minimums and network structure decision, especially many difficulties for determining the neural nodes of hidden layers [7,8,9]. So this paper incorporates SVMs with RST to design a hybrid classifier called Rough Set Support Vector Machines (RS-SVMs). The new classifier inherits the merits of both RST and SVMs. Experimental results show that the introduction of RST not only enhances recognition rates and recognition efficiencies of SVMs, but also strengthens classification and generalization capabilities of SVMs.

This paper is organized as follows. Section 2 gives feature selection method using RST. Section 3 presents a hybrid classifier based on RST and SVMs. Simulation experimental results are analyzed in section 4. Conclusions are drawn in Section 5.

2 Feature Selection Method

RST can only deal with discrete attributes. In engineering applications, especially in pattern recognition and machine learning, the features obtained using some feature extraction approaches usually vary in a certain range (interval values) instead of fixed values because of some reasons such as plenty of noise. The existing discretization methods based on cut-splitting cannot deal with the information system that contains some interval attribute values effectively, while IVAD can discretize well the interval-valued continuous features. So the IVAD is firstly used to process the features.

The key problem of IVAD is to choose a good class-sepability criterion function. When an attribute value varies in a certain range, in general, the attribute value always orders a certain law. Without loss of generality, suppose the law is approximate Gaussian distribution. This paper uses the below class-separability criterion function in feature discretization.

$$J = 1 - \frac{\int f(x)g(x)dx}{\sqrt{\int f^2(x)dx} \cdot \sqrt{\int g^2(x)dx}} \tag{1}$$

In (1), $f(x)$ and $g(x)$ represent the probability distribution functions of attribute values of two objects in universe U in a decision system, respectively. Using the criterion function and discretization algorithm in [17], the interval-valued features can be discretized effectively. After discretizing continuous features, some methods in

RST can be used to select the most discriminatory feature subset from the original feature set composed of a large number of features. This paper introduces attribute reduction method based on discernibility matrix and logic operation [12,13] to reduce discretized decision table. The detailed reduction algorithm is as follows.

Step 1 Computing discernibility matrix C_D of decision table.

Step 2 For the elements c_{ij} $(c_{ij} \neq 0, c_{ij} \neq \phi)$ of all nonempty set in discernibility matrix C_D, construct corresponding disjunction logic normal form.

$$L_{ij} = \underset{a_i \in c_{ij}}{\vee} a_i \tag{2}$$

Step 3 Conjunction operation is performed using all disjunction logic normal form L_{ij} and a conjunction normal form L is obtained finally, i.e.

$$L = \underset{c_{ij} \neq 0, c_{ij} \neq \phi}{\wedge} L_{ij} \tag{3}$$

Step 4 Transforming the conjunction normal form L into disjunction normal form L' and achieve $L' = \vee L$.

Step 5 The results of attribute reduction is achieved. In disjunction normal form L', each conjunction item corresponds a result of attribute reduction of decision table and the attributes contained in the conjunction item constitute a set of condition attribute after reduction.

Although RST finds all the reducts of the information system, the multi-solution problem brings many difficulties to decide the input features of classifiers of automatic recognition. So this paper introduces the complexity of feature extraction to solve the problem. The complexity is measured using consuming time of feature extraction. In all reducts of the information system, the feature subset with the lowest complexity is considered as the final feature subset.

3 Rough Set Support Vector Machines

SVMs have good classification, fault-tolerance and generalization capabilities. Though, SVMs cannot select and reduce the input data. If the dimensionality of input vector is very high, the training time and testing time of SVMs will be very long. Moreover, high-dimensional input feature vector usually has some redundant data, which may lower the classification and generalization performances of SVMs. So it is very necessary to use some methods to preprocess the input data of SVMs. Fortunately, RST can mine useful information from a large number of features and eliminate the redundant features, without any prior knowledge. To introduce strong capabilities of qualitative analysis and generating rules into SVMs, this section uses RST as preprocessing step of SVMs to design a hybrid classifier called RS-SVMs. The structure of RS-SVMs is shown in Fig. 1. The steps of designing RS-SVM are as follows.

Step 1 Training samples are used to construct a decision table in which all attributes are represented with interval-valued continuous features.

Step 2 IVAD is employed to discretize the decision table and discretized decision table is obtained.

Step 3 Attribute reduction methods are applied to deal with the discrete attribute table. Using attribute reduction method, multiple solutions are usually obtained simultaneously. So the complexity of feature extraction is introduced to select the final feature subset with the lowest cost from the multiple reduction results. After selection again, the final decision rule can be achieved.

Step 4 According to the final decision rule obtained, Naïve Scaler algorithm [12] is used to discretize the attribute table discretized by using IVAD and decide the number and position of cutting points. Thus, all cutting-point values are computed in terms of the attribute table before discretization using IVAD and the discretization rule, i.e. the preprocessing rule of SVMs, is generated.

Step 5 The training samples are processed using the preprocessing rule and then are used to be the inputs to train SVMs.

Step 6 When SVM classifiers are tested using testing samples or SVM classifiers are used in practical applications, the input data are firstly dealt with using preprocessing rule and then are applied to be inputs of trained SVMs.

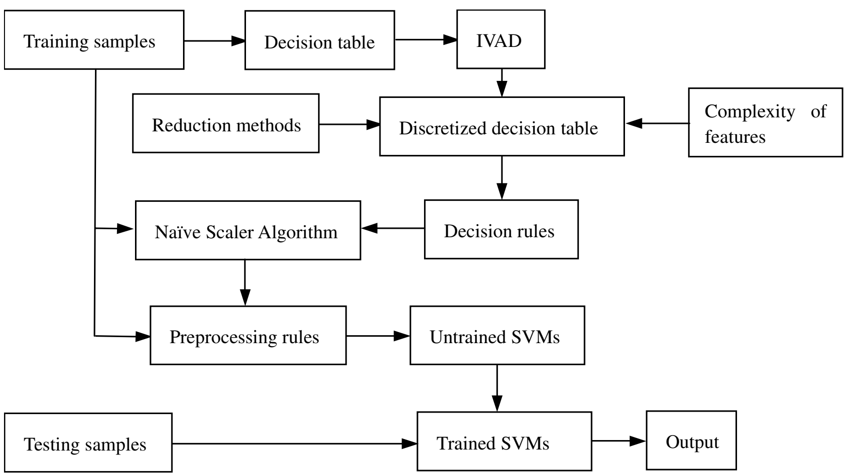

Fig. 1. Structure of RS-SVMs

SVMs were originally designed for binary classification. How to effectively extend it for multiclass classification is still an ongoing research issue [19]. Currently there are two types of approaches for multiclass SVM. One is by constructing and combining several binary classifiers while the other is by directly considering all data in one optimization formulation [19]. The former approach including mainly three methods: one-versus-rest (OVR) [19,20], one-versus-one (OVO) [19,21] and support vector machines with binary tree architecture (BTA) [22]. Some experimental results [1-9,19-22] show that the combinatorial classifier of several binary classifiers is a valid and practical way for solving muticlass classification problem.

4 Simulations

We choose 10 radar emitter signals to make simulation experiments. The 10 signals are represented with $x_1, x_2, \cdots, x_{10}$, respectively. In our prior work, 16 features have been extracted from the 10 radar emitter signals [23,24,25]. The 16 features are represented with $a_1, a_2, \cdots, a_{16}$, respectively. After discretization and reduction, the final result a_5, a_{10} is obtained. To bring into comparison, several feature selection approaches including Resemblance Coefficient method (RC) [25], Class-Separability method (CS) [26], Satisfactory Criterion method (SC) [27], Sequential Forward Search using distance criterion (SFS) [28], Sequential Floating Forward Search using distance criterion (SFFS) [28] and New Method of Feature Selection (NMFS) [29]. The results obtained using RC, CS, SC, SFS, SFFS and NMFS are a_2a_7, a_4a_{15}, a_5a_{12}, a_1a_4, a_4a_5 and a_6a_7, respectively. To test the classification performance of the results obtained by the 7 feature selection methods, BTA is used to construct SVM classifiers to recognize 10 radar emitter signals. Average accurate recognition rates obtained by using the 7 feature selection methods are shown in Table 1.

Table 1. Comparison of recognition rates (RR) obtained by 7 methods

Methods	RC	CS	SC	SFS	SFFS	NMFS	Proposed
Features	a_2a_7	a_4a_{15}	a_5a_{12}	a_1a_4	a_4a_5	a_6a_7	a_5a_{10}
RR (%)	93.89	87.69	95.12	63.27	84.73	77.68	95.32

Table 1 shows that the average recognition rate of the proposed method is higher than other 6 methods, which indicates that the feature selection method using RST is superior to other 6 methods. Simultaneously, the experimental results show that the introduced discretization method is feasible and valid.

The classification and generalization capabilities of RS-SVMs are compared with those of SVMs using the following experiments. 6 classifiers including OVR-SVM, OVO-SVM, BTA-SVM, OVR-RS-SVM, OVO-RS-SVM and BTA-RS-SVM are employed to recognize the 10 radar emitter signals. The inputs of the 6 classifiers uses the selected feature subset obtained by the proposed method, i.e. two features a_5 and a_{10}. Performance criterions including recognition error rate and recognition efficiency are used to evaluate the several classifiers. Recognition efficiency includes training time (Trt) and testing time (Tet). The samples in training group are used to train 6 classifiers and then the samples in testing group are applied to test the trained classifiers. Statistical results of 100 experiments using the 6 classifiers are shown in Table 2, respectively.

To compare the training time and the capabilities of classification and generalization of SVMs with those of RS-SVMs, different samples including 10, 20, 30, 40 and 50 are respectively applied to train OVR-SVM, OVO-SVM, BTA-SVM and OVR-RS-SVM, OVO-RS-SVM, BTA-RS-SVM. Also, testing samples of 5 dB, 10 dB, 15 dB and 20 dB are respectively used to test trained SVM and RS-SVM. After 100 experiments, the changing curves of average recognition rates (ARR) obtained using OVR-SVM and OVR-RS-SVM, OVO-SVM and OVO-RS-SVM, BTA-SVM and BTA-RS-SVM, are shown in Fig.2, Fig.3, Fig.4, respectively. The average training time (ATT) spent by OVR-SVM and OVR-RS-SVM, OVO-SVM and OVO-RS-SVM, BTA-SVM and BTA-RS-SVM, are shown in Fig.5, Fig.6, Fig.7, respectively. All experiments are made using a personal computer (P-IV, CPU: 2.0GHz, EMS memory: 256Mb).

Table 2. Experimental result comparison of 6 classifiers (%)

Signals	SVMs			RS-SVMs		
	OVR	OVO	BTA	OVR	OVO	BTA
x_1	0	0	20.36	0.40	0	0
x_2	13.86	0	26.00	0	0	0
x_3	0	0	0	0	0	0
x_4	0	0	0	0	0	0
x_5	0	0	0	0	0	0
x_6	0	0	0	0	0	0
x_7	0	0	0	0	0	0
x_8	73.20	0	0	0	0	0
x_9	0.33	0	0	0	0	0
x_{10}	0	0	0.39	0	0	0
Error rate	8.74	0	4.68	0.04	0	0
Trt (sec.)	754.74	51.96	101.14	878.11	52.58	108.53
Tet (sec.)	233.58	6.56	4.22	232.38	27.64	25.02

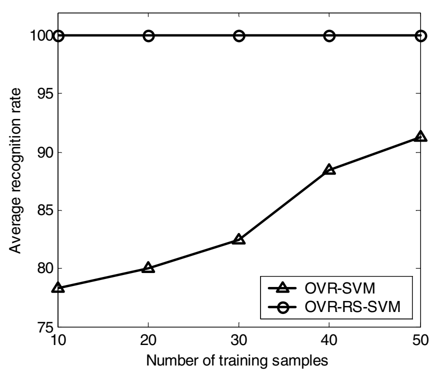

Fig. 2. ARR of OVR-SVM and OVR-RS-SVM

From Table 2 and Fig. 2 to Fig.7, several conclusions can be drawn as follows.

(1) Table 2 shows that recognition rates of 3 RS-SVM classifiers including OVR-RS-SVM, OVO-RS-SVM and BTA-RS-SVM are higher than or not less than those of 3 SVM classifiers including OVR-SVM, OVO-SVM and BTA-SVM. When the number of training samples is 50, three RS-SVM classifiers including OVR-RS-SVM,

OVO-RS-SVM, BTA-RS-SVM and OVO-SVM classifier are good classifiers to recognize the 10 radar emitter signals using the selected features.

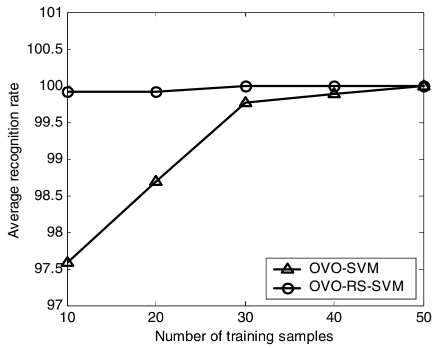

Fig. 3. ARR of OVO-SVM and OVO-RS-SVM

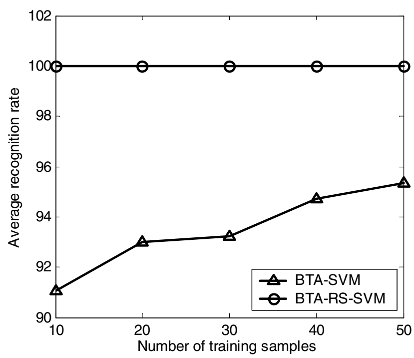

Fig. 4. ARR of BTA-SVM and BTA-RS-SVM

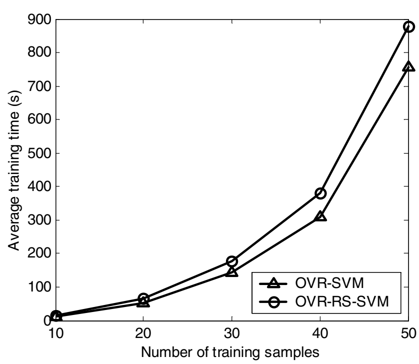

Fig. 5. ATT of OVR-SVM and OVR-RS-SVM

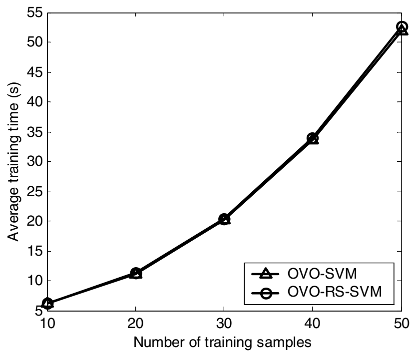

Fig. 6. ATT of OVO-SVM and OVO-RS-SVM

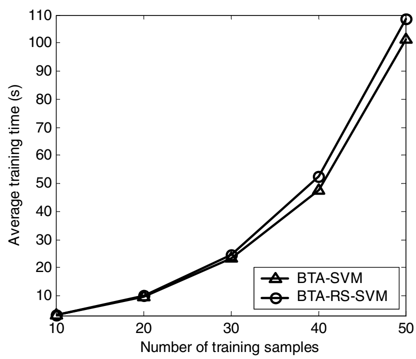

Fig. 7. ATT of BTA-SVM and BTA-RS-SVM

(2) Table 2 and Fig. 5 to Fig.7 show that RS-SVM classifiers need more time than that of SVM classifiers because the discretization procedure in RS-SVM consumes a little time.

(3) From Fig.2 to Fig.4, recognition error rates of 3 RS-SVM classifiers are lower than those of 3 SVM classifiers when the number of training samples varies from 10 to 50. The experimental results indicate that classification and generalization capabilities of RS-SVM classifiers are much stronger than those of SVM classifiers, especially when the number of training samples is small. Fig.2 to Fig.4 also show that classification and generalization capabilities of RS-SVM classifiers when the number of training samples is 10 correspond with those of SVM classifiers when the number of training samples is 50. That is to say, RS-SVM classifiers with 10 training samples are superior to SVM classifiers with 50 training samples because the former have much lower recognition error rates and much shorter training and testing time than the latter. In 3 RS-SVM classifiers, the OVO-RS-SVM classifier is the best from recognition rate and recognition efficiency.

(4) When the number of training samples is 50, OVO-SVM seems to be the best classifier from the evaluation criterions of 6 classifiers. However, Fig.3 and Fig.6 indicate that OVO-RS-SVM is superior to SVM when the number of training samples decreases.

(5) If the same values of evaluation criterions of classifiers are obtained, RS-SVM classifiers need much shorter training time and testing time than that of SVM classifiers because RS-SVM classifiers need smaller training samples.

Therefore, the above analysis indicates that the introduction of rough set theory decreases recognition error rates and enhances recognition efficiencies of SVM classifiers, and strengthens classification and generalization capabilities of SVM classifiers.

5 Conclusions

This paper combines RST with SVMs to design a hybrid classifier. RST is used to preprocess the input data of SVMs both in training procedure and in testing procedure. Because RST selects the most discriminatory features from a large number of features and eliminates the redundant features, the preprocessing step enhances the efficiency of SVMs in training and testing phases and strengthens classification and generation capabilities of SVMs. Experimental results verify that RS-SVMs are much superior to SVMs in recognition capability and in recognition efficiency. The proposed hybrid classifier is promising in other applications, such as image recognition, speech recognition and machine learning.

References

1. Osareh, A., Mirmehdi1, M., Thomas, B., and Markham, R.: Comparative Exudate Classification Using Support Vector Machines and Neural Networks. Lecture Notes in Computer Science, Vol.2489. (2002) 413-420
2. Foody, G.M., Mathur, A.: A Relative Evaluation of Multiclass Image Classification by Support Vector Machines. IEEE Transactions on Geoscience and Remote Sensing, Vol.42, No. 6. (2004) 1335-1343
3. Ma, J.S., Theiler, J., and Perkins, S.: Accurate On-line Support Vector Regression. Neural Computation, Vol.15, No.11. (2003) 2683-2703
4. Ben-Hur, A., Horn, D., Siegelmann, H.T., and Vapnik, V.: Support Vector Clustering. Journal of Machine Learning Research, Vol.2, No.2. (2001) 125-137
5. Kim, K.J.: Financial Time Series Forecasting Using Support Vector Machines. Neurocomputing, Vol.55, No.1. (2003) 307-319
6. Dibike, Y.B., Velickov, S., and Solomatine, D.: Support Vector Machines: Review and Applications in Civil Engineering. Proc. of the 2nd Joint Workshop on Application of AI in Civil Engineering, (2000) 215-218
7. Kecman, V.: Learning and Soft Computing, Support Vector Machines, Neural Networks and Fuzzy Logic Models. The MIT Press, Cambridge, MA (2001)
8. Wang, L.P. (Ed.): Support Vector Machines: Theory and Application. Springer-Verlag, Berlin Heidelberg New York (2005)

9. Samanta, B.: Gear Fault Detection Using Artificial Neural Networks and Support Vector Machines with Genetic Algorithms. Mechanical Systems and Signal Processing, Vol.18, No.3. (2004) 625-644
10. Pawlak, Z.: Rough Sets. Informational Journal of Information and Computer Science, Vol.11, No.5. (1982) 341-356
11. Lin, T.Y.: Introduction to the Special Issue on Rough Sets. International Journal of Approximate Reasoning, Vol.15, No.4. (1996) 287-289
12. Wang, G.Y.: Rough Set Theory and Knowledge Acquisition. Xi'an Jiaotong University Press, Xi'an (2001)
13. Walczak, B., Massart, D.L.: Rough Sets Theory. Chemometrics and Intelligent Laboratory Systems, Vol.47, No.1. (1999) 1-16
14. Dai, J.H., Li, Y.X.: Study on Discretization Based on Rough Set Theory. Proc. of the first Int. Conf. on Machine Learning and Cybernetics, (2002) 1371-1373
15. Roy, A., Pal, S.K.: Fuzzy Discretization of Feature Space for a Rough Set Classifier. Pattern Recognition Letter, Vol.24, No.6. (2003) 895-902
16. Mitatha, S., Dejhan, K., Cheevasuvit, F., and Kasemsiri, W.: Some Experimental Results of Using Rough Sets for Printed Thai Characters Recognition. International Journal of Computational Cognition. Vol.1, No.4. (2003) 109–121
17. Zhang, G.X., Hu, L.Z., and Jin, W.D.: Discretization of Continuous Attributes in Rough Set Theory and Its Application. Lecture Notes in Computer Science, Vol.3314. (2004) 1020-1026
18. Zhang, G.X., Hu, L.Z., and Jin, W.D.: Radar Emitter Signal Recognition Based on Feature Selection Algorithm. Lecture Notes in Artificial Intelligence, Vol.3339. (2004) 1108-1114
19. Hsu, C.W., Lin, C.J.: A Comparison of Methods for Multiclass Support Vector Machines. IEEE Transaction on Neural Networks. Vol.13, No.2. (2002) 415-425
20. Rifkin, R., Klautau, A.: In Defence of One-Vs-All Classification. Journal of Machine Learning Research, Vol.5, No.1. (2004) 101-141
21. Platt, J.C., Cristianini, N., and Shawe-Taylor, J.: Large Margin DAG's for Multiclass Classification. Advances in Neural Information Processing Systems, Vol.12. (2000) 547-553
22. Cheong, S.M., Oh, S.H., and Lee, S.Y.: Support Vector Machines with Binary Tree Architecture for Multi-class Classification. Neural Information Processing-Letters and Reviews, Vol.2, No.3. (2004) 47-51
23. Zhang, G.X., Hu, L.Z., and Jin, W.D.: Intra-pulse Feature Analysis of Radar Emitter Signals. Journal of Infrared and Millimeter Waves, Vol.23, No.6. (2004) 477-480
24. Zhang, G.X., Rong, H.N., Jin, W.D., and Hu, L.Z.: Radar Emitter Signal Recognition Based on Resemblance Coefficient Features. Lecture Notes in Artificial Intelligence, Vol.3066. (2004) 665-670
25. Zhang, G.X., Jin, W.D., and Hu, L.Z.: Resemblance Coefficient and a Quantum Genetic Algorithm for Feature Selection. Lecture Notes in Artificial Intelligence. Vol.3245. (2004) 155-168
26. Zhang, G.X., Hu, L.Z., and Jin, W.D.: Quantum Computing Based Machine Learning Method and Its Application in Radar Emitter Signal Recognition. Lecture Notes in Artificial Intelligence, Vol.3131. (2004) 92-103
27. Zhang, G.X., Jin, W.D., and Hu, L.Z.: A Novel Feature Selection Approach and Its Application. Lecture Notes in Computer Science, Vol.3314. (2004) 665-671
28. Mitra, P., Murthy, C.A., and Pal, S.K.: Unsupervised Feature Selection Using Feature Similarity. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.3. (2002) 301-312
29. Lü, T.J., Wang, H., and Xiao, X.C.: Recognition of Modulation Signal Based on a New Method of Feature Selection. Journal of Electronics and Information Technology, Vol.24, No.5. (2002) 661-666