Resemblance Coefficient and a Quantum Genetic Algorithm for Feature Selection*

Gexiang Zhang[1,2], Laizhao Hu[1], and Weidong Jin[2]

[1] National EW Laboratory, Chengdu 610031 Sichuan, China
dylan7237@sina.com
[2] School of Electrical Engineering, Southwest Jiaotong University, Chengdu 610031 Sichuan, China

Abstract. Feature selection is always an important and difficult issue in pattern recognition, machine learning and data mining. In this paper, a novel approach called resemblance coefficient feature selection (RCFS) is proposed. Definition, properties of resemblance coefficient (RC) and the evaluation criterion of the optimal feature subset are given firstly. Feature selection algorithm using RC criterion and a quantum genetic algorithm is described in detail. RCFS can decide automatically the minimal dimension of good feature vector and can select the optimal feature subset reliably and effectively. Then the efficient classifiers are designed using neural network. Finally, to bring into comparison, 3 methods, including RCFS, sequential forward selection using distance criterion (SFSDC) and a new method of feature selection (NMFS) presented by Tiejun Lü are used respectively to select the optimal feature subset from original feature set (OFS) composed of 16 features of radar emitter signals. The feature subsets, obtained from RCFS, SFSDC and NMFS, and OFS are employed respectively to recognize 10 typical radar emitter signals in a wide range of signal-to-noise rate. Experiment results show that RCFS not only lowers the dimension of feature vector greatly and simplifies the classifier design, but also achieves higher accurate recognition rate than SFSDC, NMFS and OFS, respectively.

1 Introduction

Feature selection is an important problem in pattern recognition, data mining and machine learning. The main task of feature selection is to select the most discriminatory features from original feature set to lower the dimension of pattern space in terms of internal information of feature samples. [1-5] Although it is very good that error probability of a classifier is chosen as the criterion of feature selection, it is very complicated to compute the error probability of a classifier even if class-condition probability density function is known. Moreover, the function is usually unknown in

* This work was supported by the National Defence Foundation (No.51435030101ZS0502 No.00JSOS.2.1.ZS0501), by the National Natural Science Foundation (No.69574026), by the Doctoral Innovation Foundation of SWJTU and by the Main Teacher Sponsor Program of Education Department of China.

E. Suzuki and S. Arikawa (Eds.): DS 2004, LNAI 3245, pp. 155–168, 2004.

factual situation, which brings many difficulties to analyzing the validity of a feature using directly the classification standard based on error probability. So a good solution to solve this problem is that some practical criterions must be found to decide the separability of different classes. [1-2,5] Up to now, many researchers have done much work in feature selection and have presented multiple class separability criterions and algorithms for finding the best feature subset, such as distance criterion [5], information entropy criterion [2,5], feature importance criterion [1], linear programming [3], independent feature selection [4], sequential forward selection [5], feature selection based on genetic algorithm [6], unsupervised feature selection [7], scalable feature selection [8], correlation-based feature selection method [9], feature selection based on analysis of class regions [10], feature selection with genetic algorithms [6,11-13], mutual information feature selection and Taguchi feature selection method [14]. Although some feature selection methods have been presented, the methods cannot arrive at completely satisfactory results. Feature selection is still a hot topic and is still paid much attention to by lots of researchers. [1-14]

In radar emitter signal recognition, the most significant and valid feature is not found easily because of many uncertain reasons. Some empirical methods and heuristic approaches are often used to extract features from radar emitter signals. Thus, subjectivity and guess are usually brought into feature extraction process. What is more, radar emitter signals are always interfered with by plenty of noise in the process of transmission and processing in scouts. Signal-to-noise rate (SNR) of radar emitter signals received by radar scouts varies in a large range from several dB to several tens of dB. These factors result in out-of-order distribution of feature vector and much overlapping among the features of different radar emitter signals in feature space so as to lower accurate recognition rate greatly. To eliminate the subjectivity and to enhance accurate recognition rate, multiple features must be extracted from radar emitter signals using different methods and good feature selection approach based on good evaluation criterion and efficient search algorithm must be explored to select the optimal feature subset. [5,6]

In this paper, a novel approach called resemblance coefficient feature selection (RCFS) is proposed. The main ideas of RCFS are that resemblance coefficient is chosen as class separability criterion and quantum genetic algorithm (QGA) [15,16] with rapid convergence and good global search capability is used to search the optimal feature subset from original feature set (OFS) composed of 16 features. First of all, definition and properties of resemblance coefficient and evaluation criterion based on resemblance coefficient are given. The detailed feature selection algorithm based on resemblance coefficient criterion and QGA is described. Then efficient classifiers are designed using neural network. Finally, to bring into comparison, RCFS, sequential forward selection based on distance criterion (SFSDC) [5] and a new method of feature selection (NMFS) [6] are used to make the experiment of feature selection. Dispensing with designating the dimension of feature vector, RCFS can select the best feature subset reliably and effectively. Furthermore, the feature subset obtained by RCFS achieves much higher accurate recognition rate than that of SFSDC, NMFS and OFS.

The rest of this paper is organized as follows. Section 2 describes feature selection algorithm in details. Section 3 discusses classifier design using neural network. Experiment is made in section 4 to demonstrate that RCFS is effective, which is then followed by the conclusions in Section 5.

2 Resemblance Coefficient Feature Selection Algorithm

2.1 Definition and Property of RC

Definition 1. Suppose that one-dimensional functions $f(x)$ and $g(x)$ are continuous, positive and real, i.e.

$$f(x) \geq 0, g(x) \geq 0 \tag{1}$$

Resemblance coefficient of function $f(x)$ and $g(x)$ is defined as

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

In equation (2), the integral domains of $f(x)$ and $g(x)$ are their definable domains of the variable x. Moreover, when x is within its definable domain, the value of function $f(x)$ or $g(x)$ cannot be always equal to 0.

Property 1 The value domain of resemblance coefficient C_r is

$$0 \leq C_r \leq 1 \tag{3}$$

Because $f(x)$ and $g(x)$ are positive functions, according to the famous *Cauchy Schwartz* inequality, we can obtain

$$0 \leq \int f(x)g(x)dx \leq \sqrt{\int f^2(x)dx} \cdot \sqrt{\int g^2(x)dx} \tag{4}$$

$$0 \leq \frac{\int f(x)g(x)dx}{\sqrt{\int f^2(x)dx} \cdot \sqrt{\int g^2(x)dx}} \leq 1 \tag{5}$$

Obviously, we can get $0 \leq C_r \leq 1$. According to the conditions of *Cauchy Schwartz* inequality, if $f(x)$ equals to $g(x)$, resemblance coefficient C_r of $f(x)$ and $g(x)$ gets the maximal value 1. In fact, if and only if the $f(x)$-to-$g(x)$ ratio in every point is constant, resemblance coefficient C_r equals to 1. If and only if the integral of product of $f(x)$ and $g(x)$is zero, i.e. for arbitrary x, $f(x)$=0 or $g(x)$=0, resemblance coefficient C_r equals to the minimal value 0.

From definition 1, computing resemblance coefficient of two functions corresponds to computing the correlation of the two functions. The value of resemblance coefficient mainly depends on the characteristics of two functions. If $f(x)$ is in proportion to $g(x)$, i.e. $f(x)=kg(x)$, k>0, the value of resemblance coefficient C_r equals to 1, which indicates function $f(x)$ resembles $g(x)$ completely. As the overlapping of the two functions decreases gradually, resemblance coefficient C_r will increase gradually, which indicates that $f(x)$ and $g(x)$ are resemblant partly. When $f(x)$ and $g(x)$ are completely separable, C_r gets to the minimal value 0, which implies $f(x)$ does not resemble $g(x)$ at all.

2.2 Class Separability Criterion

Class separability criterion based on probability distribution must satisfy three conditions [5]: (i) the criterion function value is not negative; (ii) if there is not overlapping part of distribution functions of two classes, the criterion function value gets to the maximal value; (iii) if distribution functions of two classes are identical, the criterion function value is 0.

Class separability criterion function based on resemblance coefficient is defined as

$$J = 1 - \frac{\int f(x)g(x)dx}{\sqrt{\int f^2(x)dx} \cdot \sqrt{\int g^2(x)dx}} = 1 - C_r \tag{6}$$

According to the definition and property of resemblance coefficient, the value of J is always equal to or more than zero. For any x, if $f(x) \neq 0$ and $g(x)=0$ or if $g(x) \neq 0$ and $f(x)=0$, J arrives at the maximal value. If $f(x)$ is the same as $g(x)$, $J=0$. So the criterion function J given in equation (6) satisfies the three class separability conditions and can be used as a standard to decide whether two classes are separable or not.

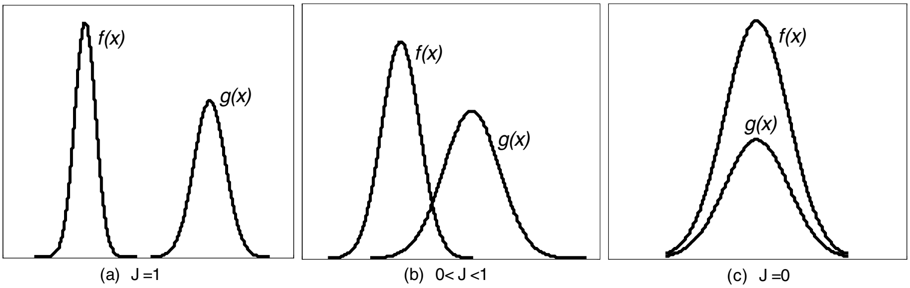

Fig. 1. Three separability cases of function $f(x)$ and $g(x)$.

When the two functions $f(x)$ and $g(x)$ in equation (6) are regarded respectively as probability distribution functions of feature samples of two classes A and B, several separability cases of A and B are shown in figure 1. For all x, if one of $f(x)$ and $g(x)$ is zero at least, which is shown in figure 1(a), A and B are completely separable and the criterion function J arrives at the maximal value 1. If there are some points of x that make $f(x)$ and $g(x)$ not equal to 0 simultaneously, which is shown in figure 1(b), A and B are partly separable and the criterion function J lies in the range between 0 and 1. For all x, if $f(x)=kg(x)$, which is shown in figure 1(c), $k=2$, A and B are not completely separable and the criterion function J arrives at the minimal value 0. Therefore, it is reasonable and feasible that the criterion function in equation (6) is used to compute separability of two classes. An additional explanation is that any function satisfied the conditions in definition 1 can be used as $f(x)$ or $g(x)$ in equation (6).

2.3 Feature Selection Algorithm

In feature extraction of radar emitter signals, all feature samples always vary in the neighboring area of expectation value because of plenty of noise and measurement errors. If occurrences of all samples are computed in statistical way, a feature probability distribution function can be obtained. The function can be considered approximately as a Gaussian distribution function with the parameters of expectation and variance of feature samples. Thus, according to the above criterion function, the feature selection algorithm of radar emitter signals is given as follows in detail.

Step 1 For a certain feature F_1, computing the class separability criterion function values of n radar emitter signals and constructing a matrix S called class separability matrix. S is

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix} \tag{7}$$

where $s_{ij}(i, j = 1, 2, \cdots, n)$ obtained according to equation (6) is separability of the ith radar emitter signal and the jth radar emitter signal. In the process of computing S, the feature samples of each radar emitter signal are considered as a Gaussian function with the parameters of expectation μ and variance σ of feature samples, where μ and σ are respectively

$$u = E(Z) \tag{8}$$

$$\sigma = D(Z) \tag{9}$$

Where Z is a vector and Z is

$$Z = [z_1, z_2, \cdots, z_M] \qquad z_p(p = 1, 2, \cdots, M) \tag{10}$$

Where z_p is the pth feature value and M is the number of feature values. According to definition and properties of resemblance coefficient, we can see easily that $s_{ii} = 1, (i = 1, 2, \cdots, n)$ and $s_{ij} = s_{ji}, (i, j = 1, 2, \cdots, n)$, i.e. class separability matrix S is a diagonal matrix. So it is enough to compute all elements above diagonal.

Step 2 Choosing a threshold value r of separable degree of two signals. If the element s_{ij} in matrix S is more than r, the ith radar emitter signal and the jth radar emitter signal are separable and we set its corresponding value to zero. Otherwise, If the element s_{ij} of matrix S is less than r, the ith radar emitter signal and the jth radar emitter signal are not separable and we set its corresponding value to one. Thus, we obtain another matrix P called class separability reduction matrix. Matrix P is composed of "0" and "1".

Step 3 Computing class separability matrix S^l of the lth feature and class separability reduction matrix $P^l\,(l = 1, 2, \cdots L)$ in the same way in step 1 and step 2, where L is the number of features. That is $P^l = \{p^l_{ij}\}, i, j = 1, 2, \cdots L$.

After step 1, 2 and 3 are finished, there are l class separability reduction matrices because every feature has a class separability reduction matrix. Thus, a problem appears subsequently that which separability reduction matrices should be chosen to form the best feature subset and how the matrices are chosen. Obviously, the problem is a combinatorial problem with T_c combinations and T_c is

$$T_c = C_l^1 + C_l^2 + \cdots + C_l^l \tag{11}$$

Genetic algorithm (GA) is a global optimization method and GA has strong robustness and general applicability. Because GA cannot be restricted by the nature of optimization problems and GA can deal with very complicated problems that cannot be solved by using traditional optimization methods, GA has become an attractive optimization approach and has been used generally in many fields [6,11-13,17-19]. Especially, quantum genetic algorithm (QGA), a new probability optimization method, is paid much attention to in recent years [15-16,20-22]. QGA is based on the concepts and principles of quantum computing. QGA uses a novel quantum bit (qubit) chromosome representation instead of binary, numeric, or symbol representation. The characteristic of the representation is that any linear superposition of solutions can be represented. Different from conventional GAs in which crossover and mutation operations are used to maintain the diversity of population, the evolutionary operation in QGA is implemented by updating the probability amplitudes of basic quantum states using quantum logic gates so as to maintain the diversity of population. QGA has good characteristics of rapid convergence, good global search capability, simultaneous exploration and exploitation, small population instead of degrading the performances of algorithm. [15,16, 20-22] So in this paper, QGA is used to solve the described combinatorial optimization problem.

Step 4 QGA [15,16] is used to search automatically the optimal feature subset from original feature set. The detailed steps using QGA are as follows.

(1) Initialization of QGA includes the process of choosing population size h and choosing the number m of qubits. The population containing h individuals is represented as $P=\{p_1,p_2,\cdots,p_h\}$, where p_j (j=1,2,$\cdots$,h) is the jth individual and p_j is

$$p_j = \left[\begin{array}{c|c|c|c} \alpha_1 & \alpha_2 & \cdots & \alpha_m \\ \beta_1 & \beta_2 & \cdots & \beta_m \end{array}\right] \tag{12}$$

In equation (12), $\alpha_i, \beta_i (i = 1, 2, \cdots, m)$ are respectively probability amplitudes of quantum basic states $|1\rangle$ and $|0\rangle$ of the ith qubit. In the beginning of QGA, all α_i , β_i

(i=1,2,$\cdots$,m) =$1/\sqrt{2}$, which indicates that all quantum basic states are superposed by the same probability. Evolutionary generation *gen* is set to 0.

(2) According to the probability amplitudes of all individuals in population P, observation state R of quantum superposition is constructed. R={a_1, a_2,⋯, a_h}, where a_j (j=1, 2,⋯,h) is observation state of each individual, i.e. a binary string composed of "0" and "1".

(3) Fitness function is used to evaluate all individuals in population P. If the dimension of feature vector of an individual in P is d and the binary string $b_1 b_2 \cdots b_K$ (b_k='0' or '1', $k = 1,2,\cdots,m$) is quantum observation state in QGA, the fitness function is defined as

$$f = d + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} q_{bij}$$
$$q_{bij} = (b_1 q_{ij}^1) \,\&\, (b_2 q_{ij}^2) \cdots \&\, (b_m q_{ij}^m) \tag{13}$$

where the symbol '&' stands for 'AND' operation in Boolean algebra. q_{bij} ($i, j = 1,2,\cdots,n$) is the element of the ith row and the jth column in matrix Q_b that is class separability reduction matrix of an individual in population. Obviously, the smaller the function f is, the better the feature subset obtained is.

(4) Maintaining the optimal individual in population P and judging terminal condition of QGA. If terminal condition is satisfied, the algorithm ends. Otherwise, the algorithm continues.

(5) Quantum rotation angles of quantum rotation gates are obtained using the method in reference [16]. Quantum rotation gates obtained operate on probability amplitudes of all individuals, i.e. the probability amplitudes of all individuals in population P are updated.

(6) Evolutionary generation *gen* increases 1. The algorithm goes to (2) and continues.

3 Classifier Design

Feature selection can be considered as a transformation that transforms the radar emitter signal feature from high dimensional feature space into low dimensional feature space and extracts the most discriminatory information and removes the redundant information. Although feature selection is an important process, it is not the final step in radar emitter signal recognition. The recognition task is to be finished only by the classifier. So classifier design is also an important process subsequent to feature extraction and feature selection in radar emitter signal recognition.

The recent vast research activities in neural classification have established that neural networks are a promising alternative to various conventional classification methods. Neural networks have become an important tool for classification because neural networks have the following advantages in theoretical aspects. [23,24] First, neural networks are data driven self-adaptive methods in that they can adjust themselves to the data without any explicit specification of functional or distributional form for the underlying model. Second, they are universal functional approximators in that neural networks can approximate any function with arbitrary accuracy. Third, neural networks are nonlinear models, which makes them flexible in modeling real world complex relationships. Finally, neural networks are able to estimate the posterior probabilities, which provide the basis for establishing classification rule and performing statistical analysis. So neural network classifers are used generally in signal recognition.

The structure of neural network classifier is shown in Fig.2. In Fig.2, L_1 is the input layer that has L neurons corresponding to radar emitter signal features. L_2 is the hidden layer and '*tansig*' is chosen as the transfer functions. L_3 is output layer that has the same number of neurons as radar emitter signals to be recognized. Transfer function in output layer is '*logsig*'. We choose RPROP algorithm [25] as the training algorithm of the neural network. The ideal outputs of neural network are "1". Output tolerance is 0.05 and output error is 0.001.

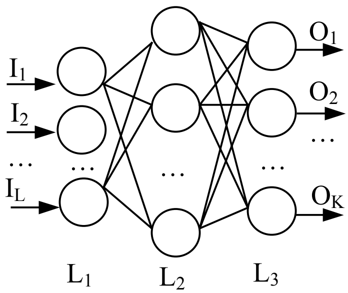

Fig. 2. The structure of neural network classifier.

4 Experimental Results

To demonstrate the feasibility and effectiveness of the proposed approach, 10 typical radar emitter signals are chosen to make the simulation experiment. They are CW, BPSK, QPSK, MPSK, LFM, NLFM, FD, FSK, IPFE and CSF, respectively. In our prior work [26-30], 16 features of 10 radar emitter signals were studied. The original feature set is composed of 16 features that are labeled as 1, 2, $\cdots$, 16. The features are respectively fractal dimensions including information dimension [26], box dimension [26] and correlation dimension [27], two resemblance coefficient features [28], Lempel-Ziv complexity [27], approximate entropy [29], wavelet entropy and eight energy distribution features based on wavelet packet decomposition [30]. In the experiment, for every radar emitter signal, 150 feature samples are generated in each SNR point of 5dB, 10dB, 15dB and 20dB. Thus, 600 samples of each radar emitter signal in total are generated when SNR varies from 5dB to 20dB. The samples are classified into two groups: training group and testing group. Training group, one third of the total samples generated, is applied to make the simulation experiment of feature selection and neural network classifer training. Testing group, two thirds of the total samples generated, is used to test trained neural network classifers.

In resemblance coefficient feature selection (RCFS), threshold value r is set to 0.99, and population size in QGA is set to 20 and the number of qubits is set to 16. To bring into comparison, sequential forward selection using distance criterion (SFSDC) [5] and a new method of feature selection (NMFS) [6] are also used to select the best feature subset.

Distance criterion function in SFSDC is

$$G_q = \sum_i \sum_j \frac{(m_i - m_j)^2}{\sigma_i^2 + \sigma_j^2}, \quad q = 1, 2, \cdots, n \tag{14}$$

where m_i and m_j are the mean values of all samples of the ith signal and the jth signal, respectively. σ_i and σ_j are the variance values of all samples of the ith signal and the jth signal, respectively. n is the number of features.

The criterion function for evaluating the best feature subset in NMFS is described as follows. Suppose the distance d_{ij} between class ω_i and class ω_j is

$$d_{ij} = \left(\sum_n^N \omega_{ij}^n \cdot \left| m_i^n - m_j^n \right|^p \right)^{\frac{1}{p}}, \qquad p \geq 1 \tag{15}$$

In equation (15), m_i and m_j are respectively average feature vectors of class ω_i and class ω_j and they are respectively

$$m_i = (m_i^1, m_i^1, \cdots, m_i^N) \tag{16}$$

$$m_j = (m_j^1, m_j^1, \cdots, m_j^N) \tag{17}$$

In equation (15), ω_{ij}^n is the weighted value of class ω_i and class ω_j and is defined as

$$\omega_{ij}^n = e^{-a(\sigma_i^n + \sigma_j^n)}, \quad a > 0 \tag{18}$$

In equation (18), σ_i^n and σ_j^n are respectively variances of feature vectors of class ω_i and class ω_j and they are respectively

$$\sigma_i = (\sigma_i^1, \sigma_i^1, \cdots, \sigma_i^N) \tag{19}$$

$$\sigma_j = (\sigma_j^1, \sigma_j^1, \cdots, \sigma_j^N) \tag{20}$$

In NMFS, genetic algorithm is used to search the best feature subset from the original feature set. If there are R classes , the fitness function is

$$f = \sum_{i=1}^{R-1} \sum_{j>1}^{R} d_{ij} \tag{21}$$

Because SFSDC and NMFS cannot decide automatically the dimension of the best feature subset selected, the dimension of feature subset obtained by RCFS is chosen as that of SFSDC and NMFS so as to draw a comparison of recognition results of three methods.

First of all, the original feature set (OFS) is used to train neural network classifiers (NNC) whose structure is 16-25-10. The samples in testing group are employed to test the trained neural network classifiers in a wide range of SNR. After 20 experiments, the recognition results are shown in table 1. The average recognition error rate is 4.83% in table 1. Then, RCFS is applied to select the best feature subset from the original feature set. The feature subset selected is composed of feature 5 and 10. The result is identical in 30 experiments. After neural network classifiers that have the structure of 2-15-10 are trained, testing results are shown in table 2, in which the average recognition error rate is only 1.34%. Finally, the dimension of feature subset in SFSDC and NMFS is designated as 2 and the two methods are respectively used to make the experiment of feature selection. The feature subset obtained by SFSDC is composed of feature 4 and 5. Feature 6 and 7 constitute the optimal feature subset in NMFS. Obviously, the structures in SFSDC and NMFS are the same as that in RCFS.

Table 1. Average recognition error rates obtained using original feature set.

Types	5 dB	10 dB	15 dB	20 dB	Average
BPSK	0.00	0.00	0.00	0.00	0.00
QPSK	66.67%	33.33%	0.00	0.00	24.75%
MPSK	4.00%	0.20%	0.00	0.00	1.41%
LFM	0.00	0.00	0.00	0.00	0.00
NLFM	0.00	0.00	0.00	0.00	0.00
CW	0.00	0.00	0.00	0.00	0.00
FD	0.10%	0.00	0.00	0.00	0.03%
FSK	0.00	0.00	0.00	0.00	0.00
IPFE	0.00	0.00	0.00	0.00	0.00
CSF	23.00	33.33%	0.00	0.00	14.08%

Table 2. Average recognition error rates using the feature subset obtained by RCFS.

Types	5 dB	10 dB	15 dB	20 dB	Average
BPSK	0.00	5.68%	0.00	0.00	1.42%
QPSK	4.57%	42.47%	0.82%	0.00	11.96%
MPSK	0.00	0.00	0.00	0.00	0.00
LFM	0.00	0.00	0.00	0.00	0.00
NLFM	0.00	0.00	0.00	0.00	0.00
CW	0.00	0.00	0.00	0.00	0.00
FD	0.00	0.00	0.00	0.00	0.00
FSK	0.00	0.00	0.00	0.00	0.00
IPFE	0.00	0.00	0.00	0.00	0.00
CSF	0.00	0.00	0.00	0.00	0.00

When SNR varies from 5dB to 20dB, recognition results of SFSDC and NMFS are shown in table 3 and table 4, respectively. The results in table 1, 2, 3 and 4 are statistical results of 20 experiments and all values are recognition error rates. Table 5 shows comparison results of OFS, RCFS, SFSDC and NMFS. In table 5, ATG an ARR are the abbreviations of average training generation of NNC and accurate recognition rate, respectively.

From table 1 to table 5, several conclusions can be drawn: (i) in comparison with OFS, RCFS not only lowers the dimension of feature vector and the cost of feature extraction greatly, but also simplifies classifier design and enhances recognition efficiency and accurate recognition rate; (ii) in comparison with SFSDC and GADC, RCFS selects better features and achieves higher accurate recognition rate because

Table 3. Recognition results using the feature subset obtained by SFSDC.

Types	5 dB	10 dB	15 dB	20 dB	Average
BPSK	9.20%	8.00%	13.53%	10.90%	10.41%
QPSK	43.00%	34.87%	45.02%	42.15%	41.0%
MPSK	24.87%	18.45%	17.22%	26.53%	21.77%
LFM	15.90%	0.00	0.00	0.00	3.97%
NLFM	0.00	0.00	0.00	0.00	0.00
CW	0.00	0.00	0.00	0.00	0.00
FD	0.00	0.00	0.00	0.00	0.00
FSK	0.00	0.00	0.00	0.00	0.00
IPFE	2.75%	0.57%	8.22%	0.00	2.88%
CSF	0.75%	0.00	1.75%	1.97%	1.12%

Table 4. Recognition results using the feature subset obtained by NMFS.

Types	5 dB	10 dB	15 dB	20 dB	Average
BPSK	32.30%	0.00	0.00	0.00	8.07%
QPSK	7.20%	0.50%	0.00	0.00	1.92%
MPSK	58.10%	0.00	0.00	0.00	14.53%
LFM	0.00	0.00	0.00	0.00	0.00
NLFM	55.54%	10.50%	0.00	0.00	16.47%
CW	0.00	0.00	0.00	0.00	0.00
FD	0.00	0.00	0.00	0.00	0.00
FSK	0.00	0.00	0.00	0.00	0.00
IPFE	23.70%	0.00	0.00	0.00	5.92%
CSF	19.30%	0.00	0.00	0.00	4.82%

Table 5. Comparison results of 4 feature sets.

Methods	Feature set	Structure of NNC	ATG of NNC	Average ARR
RCFS	5,10	2-15-10	248.10	98.66%
OFS	1~16	2-25-10	324.67	95.17%
SFSDC	4,5	2-15-10	4971.80	91.88%
NMFS	6,7	2-15-10	1146.50	94.83%

the training generation of NNC using the feature subset selected by RCFS is much less than that of SFSDC and GADC, and the average recognition error rate of RCFS is only 1.34% which is less 6.78%, 3.83% and 3.49% than that of SFSDC, GADC and OFS, respectively.

5 Concluding Remarks

This paper proposes a novel feature selection called resemblance coefficient feature selection approach. The main points of the introduced method are as follows:

(1) An effective class separability criterion function is defined with resemblance coefficient. The definition of resemblance coefficient is given and the properties of resemblance coefficient are analyzed. Using resemblance coefficient, a novel class separability criterion function is presented. The criterion function satisfies the three conditions [5] that any class separability criterion based on probability distribution must satisfy. Only using the internal information of feature samples, the presented evaluation criterion can decide the dimension of the best feature subset automatically. Therefore, the class separability criterion can overcome the problems that most existing feature selection methods need choose the dimension of the feature subset before feature selection is made and multiple tries of different dimensions must be done.

(2) An efficient optimization algorithm called quantum genetic algorithm is introduced to select the best feature subset from the original feature set composed of a large number of features. In QGA, a novel chromosome representation called qubit representation is used to represent more individuals than conventional genetic algorithm with the same population size and a novel evolutionary operation called quantum rotation gate update procedure is applied to generate the next population. Thus, QGA can maintain the population diversity in the process of searching the optimal solution and avoid the problem of selection pressure of conventional genetic algorithm. Also, there are little relations between the individuals in the child population in QGA. So QGA has good characteristics of rapid convergence, good global search capability, simultaneous exploration and exploitation, small population instead of degrading the performances of algorithm.

In order to bring into comparison, 3 methods including RCFS, SFSDC and NMFS are used respectively to select the optimal feature subset from original feature set (OFS) composed of 16 features of radar emitter signals in this paper. In the simulation experiments, the feature subsets, obtained from RCFS, SFSDC and NMFS, and OFS are employed respectively to recognize 10 typical radar emitter signals in a wide range of signal-to-noise rate. Experimental results show that RCFS not only lowers the dimension of feature vector greatly and simplifies the classifier design, but also achieves 98.66% accurate recognition rate, which is higher 6.78%, 3.83% and 3.49% than that of SFSDC, NMFS and OFS, respectively.

Although RCFS is used only in radar emitter signal feature selection in this paper, in fact, from the above analysis, RCFS has generality and can be applied in many other fields, such as data mining and machine learning. Moreover, the distribution function of feature samples in computing resemblance coefficients is not limited to Gaussian distribution.

References

1. Zhao J., Wang G.Y., Wu Z.F., etc.: The Study on Technologies for Feature Selection. Proc. of 1th Int. Conf. on Machine Learning and Cybernetics. (2002) 689-693
2. Molina L.C., Belanche L., Nebot A.: Feature Selection Algorithms: A Survey and Experimental Evaluation. Proc. of Int. Conf. on Data Mining. (2002) 306-313
3. Guo G.D., Dyer C.R.. Simultaneous Selection and Classifier Training Via Linear Programming: A Case Study for Face Expression Recognition. Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. (2003) 346-352
4. Bressan M., Vitria` J.: On the Selection and Classification of Independent Features. IEEE Trans. on Pattern Analysis and Machine Intelligence. Vol.25, No.10, (2003) 1312-1317
5. Bian Z.Q., Zhang X.G.: Pattern recognition (Second edition). Beijing: Tsinghua University Press, 2000
6. Lü T.J., Wang H., 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
7. Rhee F.C.H. and Lee Y.J.: Unsupervised feature selection using a fuzzy genetic algorithm. Proceedings of 1999 IEEE International Fuzzy Systems Conference. (1999) Vol.3, 1266-1269
8. Chakrabarti S., Dom B., Agrawal R and Raghavan P.: Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies. The VLDB Jounal. Springer-Verlag. (1998) Vol.7, 163-178
9. Liu H.Q., Li J.Y., Wong L S.: A comparative study on feature selection and classification methods using gene expression profiles. Genome Informatics. (2002) Vol.13, 51-60
10. Thawonmas R. and Abe S.: A novel approach to feature selection based on analysis of class regions. IEEE Transactions on Systems, Man, and Cybernetics----Part B: Cybernetics.(1997) Vol.27, No.2, 196-207
11. Garrett D., Peterson D.A., Anderson C W and Thaut M H.: Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering. (2003) Vol.11, No.2, 141-144
12. Haydar A., Demirekler and Yurtseven.: Speaker identification through use of features selected using genetic algorithm. Electronics Letters. (1998) Vol.34, No.1, 39-40
13. Jack L.B. and Nandi A.K.: Genetic algorithms for feature selection in machine condition monitoring with vibration signals. IEE Proceedings on Vision Image Signal Processing. (2000) Vol.147, No.3, 205-212
14. Kwak N. and Choi C.H.: Input feature selection for classification problems. IEEE Transactions on Neural Networks. (2002) Vol.13, No.1, 143-159
15. Zhang G.X., Jin W.D., Li N.: An improved quantum genetic algorithm and its application, Lecture Notes in Artificial Intelligence. (2003) Vol.2639, 449-452
16. Zhang G.X., Hu L.Z., Jin W.D.: Quantum Computing Based Machine Learning Method and Its Application in Radar Emitter Signal Recognition. Lecture Notes in Computer Science (LNCS), 2004.8 (to appear)
17. Youssef H., Sait S.M., Adiche H.: Evolutionary algorithms, simulated annealing and tabu search: a comparative study. Engineering Application of Artificial Intellegence. (2001) Vol.14, 167-181
18. Srinivas M., Patnaik L.M.: Genetic algorithms: a survey. Computer. (1994) Vol.27, No.6, 17-26

19. Koza J.R.: Survey of genetic algorithm and genetic programming. Proceedings of the 1995 Microelectronics Communications Technology Producing Quality Products Mobile and Portable Power Emerging Technologies. (1995) 589-594
20. Han K.H., Kim J.H.: Genetic quantum algorithm and its application to combinatorial optimization problems. Proceedings of the 2000 IEEE Conference on Evolutionary Computation. (2000) 1354-1360
21. Yang J.A., Li B., Zhuang Z.Q.: Research of quantum genetic algorithm and its application in blind source separation. Journal of Electronics. (2003) Vol.20, No.1, 62-68
22. Li Y., Jiao L.C.: An effective method of image edge detection based on parallel quantum evolutionary algorithm. Signal Processing. (2003) Vol.19, No.1, 69-74
23. Zhang G.P.: Neural Networks for Classification: a Survey. IEEE Trans. on System, Man, and Cybernetics-Part C: Application and Reviews. Vol.30, No.4, (2000) 451-462
24. Kavalov D., Kalinin V.: Neural Network Surface Acoustic Wave RF Signal Processor for Digital Modulation Recognition. IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control. Vol.49, No.9, (2002) 1280-1290
25. Riedmiller M., Braun H.: A Direct Adaptive Method for Faster Back Propagation Learning: The RPROP Algorithm. Proc. of IEEE Int. Conf. on Neural Networks. (1993) 586-591
26. Zhang G.X., Jin W.D., Hu L.Z.: Fractal Feature Extraction of Radar Emitter Signals. Proc. of the third Asia-Pacific conf. on Environmental Electromagnetics. (2003) 161-164
27. Zhang G.X., Hu L.Z., Jin W.D.: Complexity Feature Extraction of Radar Emitter Signals. Proc. of the third Asia-Pacific Conf. on Environmental Electromagnetics. (2003) 495-498
28. Zhang G.X., Rong H.N., Jin W.D., Hu L.Z.: Radar emitter signal recognition based on resemblance coefficient features, Lecture Notes in Computer Science (LNCS). (2004) Vol.3066, 665-670
29. Zhang G.X., Rong H.N., Hu L.Z., Jin W.D.: Entropy Feature Extraction Approach of Radar Emitter Signals. Proceedings of International Conference on Intelligent Mechatronics and Automation. 2004.8 (to appear)
30. Zhang G.X., Jin W.D., Hu L.Z.: Application of Wavelet Packet Transform to Signal Recognition. Proceedings of International Conference on Intelligent Mechatronics and Automation. 2004.8 (to appear)