

Na SUN, Yajian ZHOU, Yixian YANG

Consistency of weighted feature set and polyspectral kernels in individual communication transmitter identification

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2010

Abstract This paper presents a method using support vector machine with polyspectral kernels for classification of individual transmitters. Then, the neighborhood-rough-set-based weighted feature set is proposed. The experiments of the algorithms mentioned above indicate that they have consistency, which raises a new weighted kernel. The experiment shows that better classification rate can be achieved.

Keywords polyspectral kernel, support vector machine (SVM), neighborhood rough set, weighted feature set, weighted kernel

1 Introduction

On the difficulty for finding an optimal features' set of individual communication transmitters because of the faint discrepancies, it is extremely hard to highly classify the transmitters especially when they are with the same model and manufacturing lot. The features of "turn-on" transient signals are usually used for classifying the transmitters [1–3] because they have unique characteristics. These have achieved good recognition rate, however, the features are raw and unstable to be widely applicable.

Then, researchers paid more attention to using stable characteristics of transmitters instead of transient features. There is a consensus that higher order spectra (polyspectra) of signal plays an important role in the analysis of subtle characteristics of transmitters. The major problem of

polyspectra is the computation complexity caused by the increase of order. Even the calculus of the lowest order polyspectra (bispectra) is computationally expensive. Researchers are working at transforming the two-dimensional bispectrum into a one-dimensional template [4–6]. In nature, these algorithms are for feature extraction, which is, including linear or nonlinear mapping technique, high-dimensional data are projected toward an output space with dimension equal to the intrinsic dimension. Via feature extraction, the original physical significance is destroyed, and it is disadvantageous to the classification. Feature selection is another method. It chooses a subset of input variables by eliminating features with little or no predictive information, therefore physical significance can remain. The major problem of feature selection is how to evaluate the remaining variables in subset.

In this paper, we construct polyspectral kernels which are applied in support vector machines (SVMs). Meanwhile, the neighborhood-rough-set-based weighted feature subset is proposed. We design a simulation for analyzing the consistency of weighted feature set and polyspectrum kernels.

The rest of this paper is organized as follows: We define polyspectral kernels in support vector machines in Sect. 2, and the performance of the polyspectral kernels is investigated. In Sect. 3, the method of evaluating the features using the parameter "significances" is proposed. Then in Sect. 4, the performance of polyspectral kernels is compared with the performance of neighborhood-rough-set-based weighted feature subset. In Sect. 5, a weighted kernel is constructed from the algorithms mentioned above. Finally, the conclusions are made in Sect. 6.

Received March 5, 2010; accepted April 2, 2010

Na SUN (✉), Yajian ZHOU, Yixian YANG
Information Security Center, State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
Key Laboratory of Network and Information Attack and Defense Technology of Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China
E-mail: sunna_07@163.com

2 Polyspectral kernels in SVMs

2.1 Definition of polyspectra

Let $x(t)$ be a real stationary signal. The k th order moment of $x(t)$ is defined as

$$\begin{aligned}
m_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1}) &= E\{x(t)x(t+\tau_1)\cdots x(t+\tau_{k-1})\} \\
&= \int_{-\infty}^{\infty} x(t)x(t+\tau_1)\cdots x(t+\tau_{k-1})dt. \quad (1)
\end{aligned}$$

$m_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1})$ is the k th order correlation function of $x(t)$ where $k \geq 2$.

The k th order moment spectra of $x(t)$ is the discrete multi-dimensional Fourier transform of the k th order moment. It is defined as

$$\begin{aligned}
M_{kx}(\omega_1, \omega_2, \dots, \omega_{k-1}) &= \sum_{\tau_1=-\infty}^{\infty} \cdots \sum_{\tau_{k-1}=-\infty}^{\infty} m_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1}) e^{-j \sum_{i=1}^{k-1} \omega_i \tau_i} \\
&= \sum_{\tau_1=-\infty}^{\infty} \cdots \sum_{\tau_{k-1}=-\infty}^{\infty} \int_{-\infty}^{\infty} x(t) \prod_{n=1}^{k-1} x(t+\tau_n) e^{-j \sum_{i=1}^{k-1} \omega_i \tau_i} dt. \quad (2)
\end{aligned}$$

The polyspectrum for $k=2$ is the power spectrum, for $k=3$ it is called bispectrum and for $k=4$ it is called trispectrum.

In the paper, the k th order moment spectrum is used as polyspectrum, although the higher-order moment spectra and higher-order cumulants spectra constitute the higher-order spectra (polyspectra).

2.2 Polyspectral kernels

The n -order polyspectral characteristics reflect $n-1$ order frequencies components; therefore, they take more information of the transmitters. However, n -order polyspectra is much more computationally expensive than $n-1$ order polyspectra. Moreover, the computation of higher dimensionality is much more expensive. It is difficult to be utilized due to the computational complexity of polyspectra and high dimensionality of the feature space caused by the orders of the polyspectra growing higher. To address this problem we present an algorithm of using polyspectral kernels in SVMs. The computation of polyspectra can be avoided; the inner product of polyspectra will be used, which reduces the computational complexity.

SVM is a popular technique for pattern recognition and is considered as the state-of-the-art tool for linear and nonlinear classification [7]. Some conventional kernels are usually mentioned, such as the radial-basis function (RBF) kernel and polynomial kernel [8]. The kernel function is defined as

$$K(\mathbf{x}, \mathbf{y}) = \sum_i \Phi(\mathbf{x})_i \Phi(\mathbf{y})_i, \quad (3)$$

where Φ is a mapping function. Computation of $K(\mathbf{x}, \mathbf{y})$ is usually much less than evaluating Φ , thus it allows us to

construct hyperplanes more easily in very high dimensional spaces. In this paper, the new polyspectral kernel is defined as a simple mathematical form; moreover, the computational complexity is reduced. In the same experimental environment, experiments show that calculation speed increases at least one order of magnitude when the order of polyspectra is low. The higher the order grows, the more obvious the advantage of polyspectral kernel takes on.

By defining $\mathbf{\Omega}_{k-1} = (\omega_1^T \omega_2^T \cdots \omega_{k-1}^T)^T$, $\mathbf{T}_{k-1} = (\tau_1^T \tau_2^T \cdots \tau_{k-1}^T)^T$, the k th polyspectral kernel is defined as the following:

$$\begin{aligned}
K(\mathbf{x}, \mathbf{y}) &= \langle M_{kx}(\mathbf{\Omega}_{k-1}), M_{ky}(\mathbf{\Omega}_{k-1}) \rangle \\
&= \int \cdots \int M_{kx}(\mathbf{\Omega}_{k-1}) \times M_{ky}^*(\mathbf{\Omega}_{k-1}) d\omega_1 \cdots d\omega_{k-1} \\
&= \left\langle \sum_{\tau_1=-\infty}^{\infty} \cdots \sum_{\tau_{k-1}=-\infty}^{\infty} \int_{-\infty}^{\infty} x(t) \prod_{n=1}^{k-1} x(t+\tau_n) e^{-j\mathbf{\Omega}_{k-1}^T \mathbf{T}_{k-1}} dt, \right. \\
&\quad \left. \sum_{\tau_1=-\infty}^{\infty} \cdots \sum_{\tau_{k-1}=-\infty}^{\infty} \int_{-\infty}^{\infty} y(t) \prod_{n=1}^{k-1} y(t+\tau_n) e^{-j\mathbf{\Omega}_{k-1}^T \mathbf{T}_{k-1}} dt \right\rangle \\
&= \sum_{\omega_1=-\infty}^{\infty} \cdots \sum_{\omega_{k-1}=-\infty}^{\infty} \left[\sum_{\tau_1=-\infty}^{\infty} \cdots \sum_{\tau_{k-1}=-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t_1) y(t_2) \right. \\
&\quad \left. \cdot \prod_{n=1}^{k-1} x(t_1 + \tau_n) \prod_{n=1}^{k-1} y(t_2 + \tau_n) \cdot e^{-j\mathbf{\Omega}_{k-1}^T \mathbf{T}_{k-1}} dt_1 dt_2 \right] \\
&= (2\pi)^{2(k-1)} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x(t) y(t+\tau) dt \right]^k d\tau. \quad (4)
\end{aligned}$$

It is clear that the polyspectral kernel $K(\mathbf{x}, \mathbf{y})$ is a function of the correlation of \mathbf{x} and \mathbf{y} .

Experiment 1

In order to illustrate the performance of polyspectral kernel, four sets of data are provided, amplitude modulation (AM), frequency modulation (FM), frequency shift keying (FSK), and frequency-hopping spread spectrum (FHSS) samples included. Figure 1 shows that the recognition rates vary with the order of polyspectra for two classes of patterns. It seems that the performance for

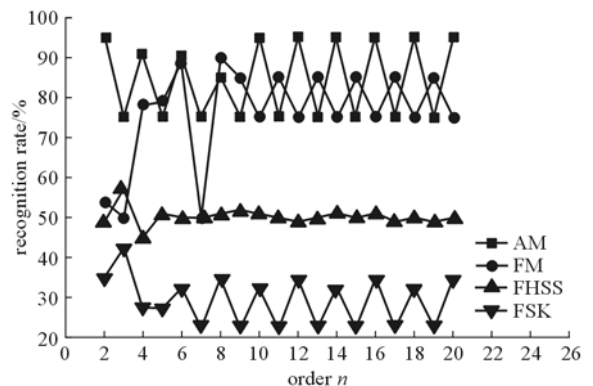


Fig. 1 Recognition rates of polyspectral kernel

classification of AM and FM signals are much better than that of FSK and FHSS signals. The result also represents that higher order of polyspectra cannot always bring about higher classification effect. Certainly the recognition rates will decline as the number of classes increases in terms of multi-classification.

3 Neighborhood-rough-set-based weighted feature set

3.1 Neighborhood rough set

Neighborhood rough set (RST) is an approach proposed in this paper for not only feature selection, but also advanced feature evaluation. In RST, a set of all similar objects is called an elementary set, which makes a fundamental atom of knowledge. Any union of elementary sets is called a crisp set and other sets are referred to as rough set [9]. Reference [10] introduced neighborhood relation into rough set methodology.

Let an information system be defined as a function $f : U \times A \rightarrow V$, where U is the set of all objects, A is the set of all attributes, V is the set of all attribute values. A neighborhood decision system is denoted as a 6-tuple $NDS = \{U, A, V, f, \delta, \theta\}$, where δ is the neighborhood measure function and θ is the threshold. The lower and upper approximations of the decision D with respect to attributes B are denoted as [10]

$$\underline{N}_B D = \cup_{i=1}^N \underline{N}_B X_i, \quad \overline{N}_B D = \cup_{i=1}^N \overline{N}_B X_i, \quad (5)$$

where

$$\begin{aligned} \underline{N}_B X &= \{x_i | \delta_B(x_i) \subseteq X, x_i \in U\}, \\ \overline{N}_B X &= \{x_i | \delta_B(x_i) \cap X \neq \emptyset, x_i \in U\}. \end{aligned} \quad (6)$$

Then the decision boundary region of D with respect to attribute B defined as [10]

$$BN(D) = \overline{N}_B X - \underline{N}_B X. \quad (7)$$

The dependency degree of D to B [10] is defined as the ratio of consistent objects:

$$\gamma_B(D) = \frac{|\text{POS}_B(D)|}{|U|}, \quad (8)$$

in which $\text{POS}_B(D)$ is called positive region of decision [10].

Given a decision system $\langle U, C, D \rangle$, $B \subseteq C$, $a \in B$, the significance of an attribute is defined as [10]

$$\text{SIG}(a, B, D) = \gamma_B(D) - \gamma_{B-a}(D). \quad (9)$$

Given a decision system $\langle U, C, D \rangle$, $B \subseteq C$, $a \notin B$, the significance of an attribute is defined as [10]

$$\text{SIG}(a, B, D) = \gamma_{B \cup a}(D) - \gamma_B(D). \quad (10)$$

In Ref. [11], a forward greedy search algorithm for attribute reduction is proposed, and the minimal feature subset is achieved. More details of how to get the minimal feature subset can be found in Ref. [11].

3.2 Neighborhood-rough-set-based weighted feature set

Generally, the method of feature selection aims at how to evaluate the subsets and then grade a subset of features to remove most redundant features from the data. The disadvantage is that after the process of reducing, $\text{SIG}(a, B, D)$ will usually be abandoned. We think that it could be useful for constructing a more effective subset, that is, it could be utilized to construct a “weighted” feature set which will be used in SVM.

The process will be described as follows:

Step 1 Let the raw feature dataset be a matrix $X^T = \{\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_N^T\}$. X is matrix of size $N \times M_1$, where \mathbf{x}_i is a feature vector and N is the number of samples.

Step 2 Using the forward greedy search algorithm defined in Ref. [11], the matrix X is transformed into a new matrix, which is denoted as $R_X^T = \{\mathbf{r}_1^T, \mathbf{r}_2^T, \dots, \mathbf{r}_N^T\}$. R_X is a matrix of size $N \times M_2$, in which $M_2 \leq M_1$, means that the feature set has been reduced after being transformed by the algorithm.

Step 3 In Step 2, the significance of an attribute has been calculated, and $r_{i1}, r_{i2}, \dots, r_{iM_2}$, $i = 1, 2, \dots, N$, are descending orderly sequences ordered by their significance. Take the record of the significances of every attribute which is selected into the feature subset. Denote the significances as $\gamma_1, \gamma_2, \dots, \gamma_{M_2}$, $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_{M_2}$.

Step 4 Normalize the reduced dataset R_X . Let the normalization of R_X be

$$z_{ij} = \frac{r_{ij} - \bar{r}_j}{s_j}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, M_2, \quad (11)$$

where

$$\bar{r}_j = \frac{1}{N} \sum_{i=1}^N r_{ij}, \quad s_j^2 = \frac{1}{N} \sum_{i=1}^N (r_{ij} - \bar{r}_j)^2. \quad (12)$$

Step 5 In standard SVMs [8], the decision function is obtained as

$$f(\mathbf{x}) = \text{sgn} \left[\sum_{i=1}^l y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \right], \quad (13)$$

where K is the kernel function.

Define $A = \text{diag}(\gamma_1, \gamma_2, \dots, \gamma_{M_2})$, then the decision function is rewritten as

$$f(\mathbf{x}) = \text{sgn} \left[\sum_{i=1}^l y_i \alpha_i K(A \mathbf{x}_i, A \mathbf{x}) + b \right]. \quad (14)$$

4 Experiments for consistency of significances and polyspectral kernels

Experiment 2

The data used in Experiment 1 are also applied to the experiment of weighted feature set in this section. The polyspectra with different order of the data are chosen as the attributes and Gaussian kernel is selected for SVM in simulation. A series of neighborhood measure function δ is adopted for comparison. Take AM samples for example, given that the order of polyspectra are selected from 2 to 20, Table 1 shows the significance of every attribute when different δ is selected.

Table 1 significances of attributes

n	significances of attributes				
	$\delta=0.1$	$\delta=0.2$	$\delta=0.4$	$\delta=0.6$	$\delta=0.8$
2	0.08	0.07	0.08	0.06	0.09
3	0.04	0.03	0.02	0.04	0.04
4	0.06	0.06	0.07	0.06	0.05
5	0.05	0.04	0.04	0.04	0.02
6	0.04	0.06	0.06	0.09	0.05
7	0.04	0.03	0.03	0.02	0.03
8	0.06	0.07	0.08	0.08	0.08
9	0.05	0.04	0.04	0.03	0.03
10	0.07	0.08	0.09	0.07	0.06
11	0.04	0.03	0.03	0.04	0.02
12	0.06	0.07	0.07	0.08	0.09
13	0.04	0.03	0.02	0.02	0.02
14	0.05	0.06	0.07	0.05	0.08
15	0.03	0.02	0.02	0.04	0.03
16	0.07	0.08	0.07	0.08	0.08
17	0.03	0.03	0.02	0.03	0.04
18	0.07	0.08	0.08	0.09	0.08
19	0.04	0.04	0.03	0.02	0.02
20	0.08	0.08	0.08	0.08	0.09

Table 1 seems complex and no rules to follow. So we observe the classification rate.

Experiment 3

When weighted features are used, for multi-classifier (class = 5), it is shown in Fig. 2 that the AM classification rate can be better and stable if $0.2 \leq \delta \leq 0.5$. The result is what we would like, that is, an appropriate value of δ can make stable and satisfactory performance, rather than best performance. Experiments of FM and FHSS signals are similar, whereas performance of FSK is unsatisfactory just as Experiment 1.

Figure 3 shows the significances of attributes which are the n th polyspectra of AM sample as $\delta = 0.4$. Compared to Fig. 1, it can be found that the fluctuations of significances

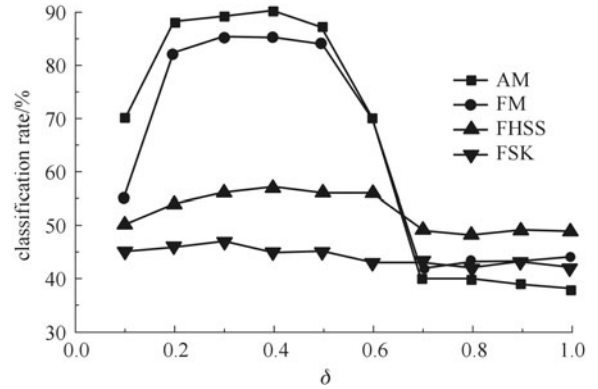


Fig. 2 Classification rate with δ

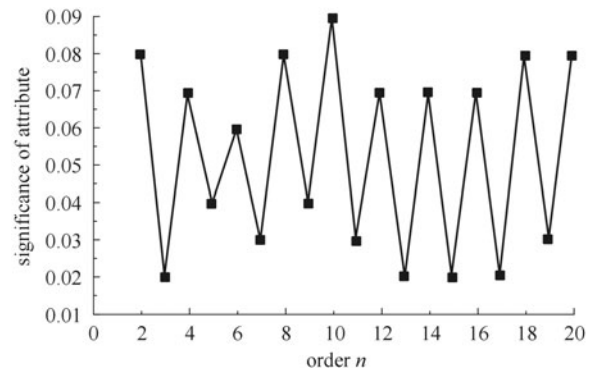


Fig. 3 Significances of attributes ($\delta = 0.4$)

are, in general, consistent with the curve in Fig. 1. The result reveals two points:

1) The method of neighborhood-rough-set-based weighted feature set and the method of polyspectral kernels are both effective to some degree. As Experiment 1 shows, higher order does not always bring about higher performance of recognition.

2) In case the two methods are combined, it will be easier to choose the parameter of polyspectra and neighborhood measure function δ . At least the performance of recognition would not too bad.

To address the issue, an improved algorithm, which can be called weighted kernel, will be proposed in the next section.

5 Weighted polyspectral kernel

Combined kernel function has been proposed in Ref. [12], which means the combination of different kernel functions by addition and multiplication operators. In this section, we make some improvements on the algorithm of neighborhood-rough-set-based weighted feature set, using combined kernel function, called weighted kernel, which can be described as follows:

1) Assuming that the n th polyspectra are chosen to be features, $2 \leq n \leq N$, denote corresponding significances as $\lambda_1, \lambda_2, \dots, \lambda_N$.

2) The combined kernel function is represented as [13]

$$K_{\text{comb}} = \alpha_1 K_1 \circ \alpha_2 K_2 \circ \dots \circ \alpha_m K_m, \quad (15)$$

where K_i , $i = 1, 2, \dots, m$, denotes the kernel function, \circ represents the operator between the two kernel functions, which can be addition and multiplication operators. α_i is the weighted parameter of the kernel function K_i , $\sum_{i=1}^N \alpha_i = 1$.

3) It can be seen that $\sum_{i=1}^N \lambda_i = 1$; moreover, the significances have been evaluated. Thus, we would like to construct a weighted kernel as

$$K_{w1} = \lambda_1 K_1 + \lambda_2 K_2 + \dots + \lambda_N K_N. \quad (16)$$

Experiment 4

The dataset used in Experiment 1 are also applied. Figure 4 shows the classification rates for AM, FM, FHSS and FSK signals when weighted kernel is utilized. It is implicated that the performances are stable, and the result indicates that the polyspectra feature is inapplicable for recognition of FHSS and FSK signals, especially FSK signals. Besides, in terms of training time, the speed of using weighted polyspectral kernels is much higher than conventional kernels, although the weighted polyspectral kernels look more complex.

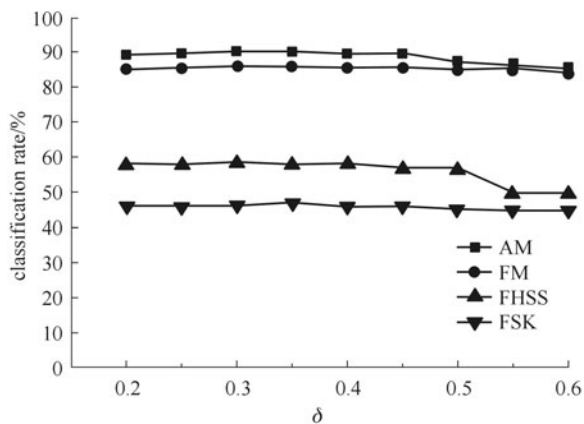


Fig. 4 Classification rate with δ (weighted kernel)

6 Conclusions

In this paper, we construct polyspectral kernels applied in SVMs and discuss the performances of this algorithm for different signals. To address the problem of fluctuating classification rates, we propose a novel method of neighborhood-rough-set-based weighted feature set. The experiment shows how to choose a suitable neighborhood measure function δ , and it also gives a suggestion of significance of every feature. However, the calculation of

conventional kernel in SVMs is harder than polyspectral kernels. Thus, the weighted polyspectral kernel is developed. Although the combined kernel function is more computationally expensive than single polyspectral kernel, it takes better and stable performance of recognition. The work is imperfect with respect to that this method seems inapplicable for FHSS and FSK signals.

Acknowledgements This work was supported by the National High Technology Research and Development Program of China (Grant No. 2009AA01Z430), the Natural Science Foundation of Beijing (No. 9092009), and the National Science and Technology Major Program (2009ZX03004-003-03).

References

- Toonstra J, Kinsner W. Transient analysis and genetic algorithms for classification. In: Proceedings of IEEE WESCANEX 95: Communications, Power, and Computing. 1995, 2: 432–437
- Shaw D, Kinsner W. Multifractal modelling of radio transmitter transients for classification. In: Proceedings of IEEE WESCANEX 97: Communications, Power and Computing. 1997, 306–312
- Sun L, Kinsner W, Serinken N. Characterization and feature extraction of transient signals using multifractal measures. In: Proceedings of 1999 IEEE Canadian Conference on Electrical and Computer Engineering. 1999, 2: 781–785
- Tugnait J K. Detection of non-Gaussian signals using integrated polyspectrum. IEEE Transactions on Signal Processing, 1994, 42 (11) : 3137–3149
- Liao X J, Bao Z. Circularly integrated bispectra: novel shift invariant features for high resolution radar target recognition. Electronics Letters, 1998, 34 (19): 1879–1880
- Xu S H, Huang B X, Huang Y C, Xu Z G. Identification of individual radio transmitters based on selected surrounding-line integral bispectra. In: Proceedings of the 9th International Conference on Advanced Communication Technology. 2007, 2: 1147–1150
- Burges C J C. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 1998, 2(2): 121–167
- Vapnik V. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995
- Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data. Boston: Kluwer Academic Publishers, 1991
- Hu Q H, Yu D R, Xie Z X. Neighborhood classifiers. Expert Systems with Applications, 2008, 34(2): 866–876
- Hu Q H, Yu D R, Liu J F, Wu C X. Neighborhood rough set based heterogeneous feature subset selection. Information Sciences, 2008, 178(18): 3577–3594
- Nguyen H N, Ohn S Y, Park J, Park K S. Combined kernel function approach in SVM for diagnosis of cancer. Lecture Notes in Computer Science, 2005, 3610: 1017–1026
- Tan Y, Wang J. A support vector machine with a hybrid kernel and minimal Vapnik-Chervonenkis dimension. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(4): 385–395