Multiple Kernel Fuzzy Support Vector Machine Based on Kernel-Target Alignment

Jiaoyang Zhang, Qiang He
School of Science, Beijing University of Civil Engineering and Architecture, Beijing, 100044, China

Abstract: Support vector machines (SVMs) are currently widely used machine learning techniques. SVM is used to construct an optimal hyper-plane that implies an extraordinary generalization capability and good performances. So far, SVMs have already been successfully applied to many real fields. In view of the difficulties in kernel selection and sensitivity to noise, we propose Kernel-Target alignment based multiple kernel fuzzy support vector machine in this paper. The Kernel-Target alignment based multiple kernel learning is introduced to multiple kernel fuzzy support vector machine. It not only avoids the problem of kernel selection, but also improves the robustness to noise. The experimental simulation also validates the feasibility and effectiveness of the method.

Keywords: Kernel-Target alignment; Multiple kernel learning; Support vector machine; Fuzzy

1. Introduction

Support Vector Machine (SVM), which is based on statistical learning theory and was proposed by Vapnik et al. in 1995 [1-2]. It can improve the generalization ability of the learning machine with the principle of minimizing structural risk [2-3] as well as achieve the minimization of empirical risk and confidence range. Thus, SVM can obtain a good statistical rule in the case of fewer statistical samples. Support vector machine has been applied in many fields, such as text classification, speech recognition, emotional analysis and regression analysis [4-6] in that it is a powerful tool for solving the problems of small sample, nonlinear, high dimension, etc.

With the continuous developing and applying of SVM, however, there are some limitations has gradually revealed, and research about SVM in many aspects has yet to be explored and improved. For example, SVM is sensitive to noise points and outliers. In order to solve this problem, Lin et al. advanced the concept of Fuzzy Support Vector Machine (FSVM) [7-8], which introduced fuzzy membership into support vector machine as weight of every sample. FSVM reduces the influence of noise points and outliers on the final decision function to a certain extent, and improves the classification accuracy. Now it is also applied to many areas such as risk prediction, fault diagnosis, and handwritten string recognition, etc [9-11] .

In addition, selecting the kernel function and its parameters is an important issue for SVM. However, there is no effective means to complete the work. Multiple kernel learning (MKL) [12-15], an important result of nuclear methods, which has become a research hotspot of many scholars in the field of machine learning in recent years. Different kernel functions correspond to different similarity representations. A single kernel function often cannot fully portray the similarity between data, especially the similarity between complex data. Therefore, the combination of multiple kernels can describe the similarity of the data more accurately and can avoid the problem of kernel selection.

In this paper, inspired by FSVM and MKL, we put forward a multiple kernel fuzzy support vector machine based on Kernel-Target alignment to solve problem of kernel selection and sensitivity to noise. Membership evaluated by fuzzy rough set is introduced into multiple kernel support vector machines based on Kernel-Target alignment.

Experiments verify that the proposed multiple kernel fuzzy SVM (MKFSVM) performs better than the classical SVM, FSVM and multiple kernel SVM (MKSVM).

The following contents of this paper are structured as follows. section 2 briefly presents a review of SVM, FSVM and MKL. section 3 describes in detail multiple kernel fuzzy support vector machine based on Kernel-Target alignment. The experimental simulation will be presented in section 4. Finally, Section 5 summarizes the whole paper.

2. Preliminaries

In this section, we will first review the basic contents of SVM, FSVM [16]. We will then recall the related contents of MKL [17-18].

2.1. Svm and fsvm

For a dataset $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_l,y_l)\}$, where $x_i\in R^n$ and $y_i\in\{-1,1\}$ for $i=1,2,\cdots,l$. The goal of SVM is to obtain the optimal separating hyperplane $f(x)=\omega^{\mathrm{T}}x+b$, which can not only classify the data into two categories correctly, but also ensure the maximum margin between the two types of data samples. Where ω is the weight vector and $b\in R$ is the threshold value. According to the principle of structural risk minimization, the process of finding the optimal separating hyperplane can be reduced to the following optimization problem

$$\begin{aligned}&\min\ \frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{l}\xi_i\\&s.t.\ y_i\left(\omega^{\mathrm{T}}\cdot x_i+b\right)\ge 1-\xi_i\\&\qquad \xi_i\ge 0, i=1,2,\cdots,l\end{aligned}\tag{1}$$

where ξ_i is the error term, C determines the trade-off between margin maximization and training error minimization.

The Lagrange multiplier method is used to transform the above optimization problem into the following dual problem

$$\begin{aligned}&\max\ \sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j\left\langle x_i,x_j\right\rangle\\&s.t.\ \sum_{i=1}^{l}\alpha_i y_i=0,\ 0\le\alpha_i\le C,\ i,j=1,2,\cdots,l\end{aligned}\tag{2}$$

Where α_i is the Lagrange multiplier, $\langle\cdot,\cdot\rangle$ represents the inner product. It can be concluded that the decision function is $f(x)=\mathrm{sgn}\left(\sum_{i=1}^{l}\alpha_i y_i\left\langle x_i,x\right\rangle+b\right)$

For the nonlinear separable data, although the separating hyperplane can be obtained by introducing the error term ξ_i, it cannot classify all samples correctly. Hence SVM was extended to the feature space. To construct a separating hyperplane in a feature space one first has to transform the n-dimensional input space into a high-dimensional renewable Hilbert space through a choice of a nonlinear mapping $\varphi:R^n\to H$. Without any knowledge of the mapping, the optimal separating hyperplane is constructed by using the dot product function in the feature space. The dot function is usually called a kernel function. According to Hilbert-Schmidt theorem, there is a relationship between the input space and its feature space for the dot product of two samples. The relationship is

$$k(x_i,x_j)=\left\langle\varphi(x_i),\varphi(x_j)\right\rangle\tag{3}$$

$k(x_i,x_j)$ is conventionally defined as a kernel function satisfying the Mercer theorem [2]. Then, equation (2) can be converted to:

$$\begin{aligned}&\max\ \sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j k\left(x_i,x_j\right)\\&s.t.\ \sum_{i=1}^{l}\alpha_i y_i=0,\ 0\le\alpha_i\le C,\ i,j=1,2,\cdots,l\end{aligned}\tag{4}$$

By calculating (4) , the decision function is

$$f(x)=\mathrm{sign}(\textstyle\sum_{i=1}^{l}y_i\alpha_i k(x_i,x)+b)\tag{5}$$

Linear kernels, polynomial kernels, and Gaussian kernels are commonly used kernel functions.

It is often difficult for SVM to obtain satisfactory decision function because the data is easily interfered by various factors such as noise in uncertain environment. In order to effectively decrease the interference of outliers or noise, in 2002, Liu proposed fuzzy support vector machine (FSVM). Each data point in the training dataset is assigned a membership. A data point is assigned a low membership if it is considered as an outlier or noise, which weakens its impact on the classification hyperplane. The FSVM corresponds to the following optimization problems:

$$\begin{aligned}&\min\ \frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{l}s_i\xi_i\\&s.t.\ y_i\left(\omega^{\mathrm{T}}\cdot x_i+b\right)\ge 1-\xi_i\\&\qquad \xi_i\ge 0, i=1,2,\cdots,l\end{aligned}\tag{6}$$

Where s_i is the fuzzy membership, which is the attitude of the corresponding point x_i toward the class. Obviously, s_i plays a weighting role on the ξ_i in the objective function, which makes the noise point have little influence on the final hyperplane. This algorithm enhances the robustness of SVM.

The (6) can be transformed into its dual form by the Lagrange multiplier method:

$$\begin{aligned}&\max\ \sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j\left\langle x_i,x_j\right\rangle\\&s.t.\ \sum_{i=1}^{l}\alpha_i y_i=0,\ 0\le\alpha_i\le Cs_i,\ i,j=1,2,\cdots,l\end{aligned}\tag{7}$$

The resulting function is $f(x)=\mathrm{sgn}\left(\sum_{i=1}^{l}\alpha_i y_i k(x_i,x)+b\right)$.

Various kinds of fuzzy SVM to deal with different specific problems are put forwarded by other scholars based on the above FSVM algorithm.

These methods are proposed to deal with some uncertainties in practical problems, they are improvements and perfections of original SVM.

2.2. Multiple kernel learning

For SVM, selection of kernel function and its parameters plays an important role. Generally, the kernel function and its parameters with higher accuracy are selected. The calculation cost is relatively high. MKL combining multiple kernel functions to replace the single kernel, can characterize the similarity between samples more accurately, then SVM can achieve better classification performance with low computational cost. The combination mode is shown in the figure below:

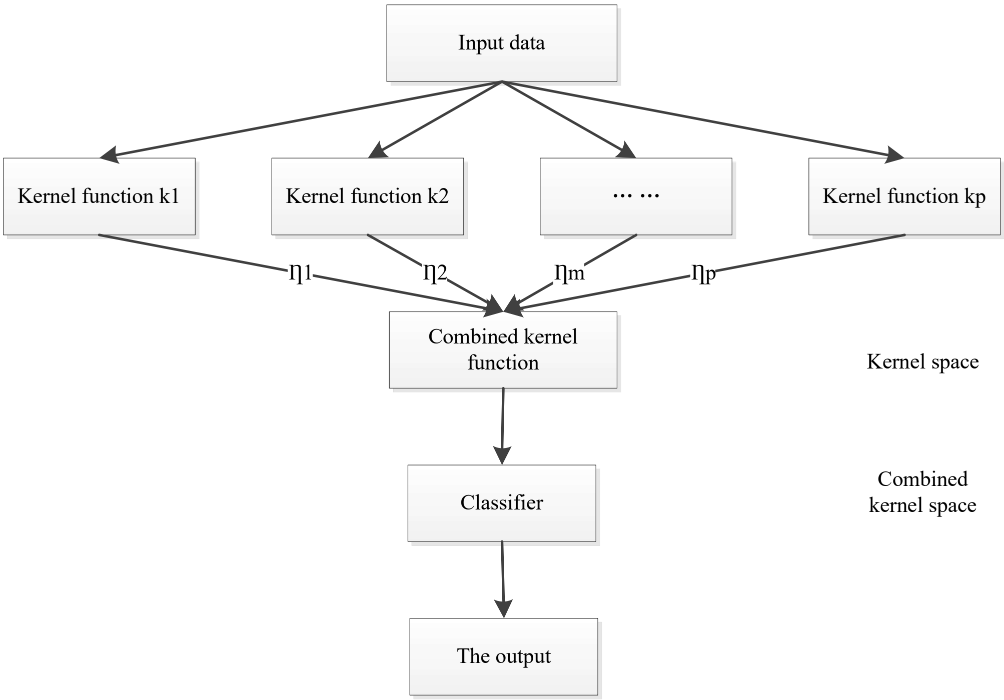

Figure 1. Sketch map of combination of multiple kernel functions

Where $\eta_m, m=1,2,\cdots,p$ is weight coefficient for each kernel function respectively.

In Multiple kernel support vector machine (MKSVM), combination of kernel and the calculation of weight is mainly considered. At present, the combination of kernel functions can be divided into the following three types:

Linear combination method [19] is the most widely used, such as direct summation kernel and weighted summation kernel. The formula is as follows

$$k_\eta\left(x_i,x_j\right)=f_\eta\left(\left\{k_m\left(x_i^m,x_j^m\right)\right\}_{m=1}^{P} \mid \eta\right)=\sum_{m=1}^{P}\eta_m k_m\left(x_i^m,x_j^m\right) \tag{8}$$

Where η is the kernel weight, p is the number of kernel functions. And η is set to 1 in the direct summation kernel.

Nonlinear combination method [20-22], such as multiplication, power, and exponentiation, the formula of the multiplication is as follows:

$$k_\eta\left(x_i,x_j\right)=\prod_{m=1}^{P}k_m\left(x_i^m,x_j^m\right) \tag{9}$$

Data-dependent combination method assigns a specific weight to each data sample. It can get the local distribution of data, and then the appropriate combination rule can be learned for each region according to the distribution.

The calculation of multiple kernel weight coefficients is also a key issue in the research of MKL. At present, there are five kinds of calculation methods of relevant weight coefficients: fixed rules, heuristic approaches [23], optimization approaches [24], Bayesian approaches and boosting approaches [25].

In this paper, heuristic learning is adopted to obtain the kernel weight, and the multiple kernel weight coefficient is calculated based on kernel alignment [26-28]. Through this method, the weight coefficients are determined by calculating the similarity between the corresponding kernel matrices of the kernel functions. Based on the training set T, kernel alignment between kernel matrix K_1 and kernel matrix K_2 is defined as:

$$A\left(T,K_1,K_2\right)=\frac{\langle K_1,K_2\rangle_F}{\sqrt{\langle K_1,K_1\rangle_F\langle K_2,K_2\rangle_F}} \tag{10}$$

in which $\left\langle K_1, K_2 \right\rangle_F = \sum_{i=1}^{l}\sum_{j=1}^{l} K_1\left(x_i, x_j\right) K_2\left(x_i, x_j\right)$. The optimization problem of MKSVM is transformed from (1) to:

$$\max \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j k_\eta\left\langle x_i, x_j\right\rangle$$
$$s.t.\ \sum_{i=1}^{l}\alpha_i y_i = 0,\ 0 \le \alpha_i \le C,\ i,j = 1,2,\cdots,l \tag{11}$$

Where k_η is synthesis kernel. The classification decision function becomes $f\left(x\right) = \mathrm{sgn}\left(\sum_{i=1}^{l}\alpha_i y_i k_\eta\left(x_i, x\right) + b\right)$.

3. Kernel-Target Alignment Based Multiple Kernel Fuzzy Support Vector Machine

Due to limitations in kernel selection and sensibility to noise data, this paper introduces kernel alignment weighted summation based multi-kernel learning into SVM, and uses the fuzzy rough set method to obtain the sample membership. A multiple kernel fuzzy support vector machine (MFSVM) based on kernel alignment is proposed.

Rough set (RS) is a mathematical theory of uncertainty proposed by Polish scholar Pawlak [29] in the 1980s. It can analyze data and find hidden knowledge from data. As a promotion of RS, Fuzzy Rough Set (FRS), it can deal with real-valued data. The FRS theory was proposed firstly by Dubois and Prade [30-31], and has been rapidly developed [32-37]. The existing FRS method has been successfully applied to many practical problems [38-40]. In this paper, we use the low approximation operator of FRS based on Gaussian kernel as the membership degree of the case [41]. For $x \in A$, $s = \inf_{x' \notin A}\sqrt{1-\left(k\left(x, x'\right)\right)^2}$. Where s is the membership that x belongs to A, and $k(x, x')$ is Gaussian Kernel.

In this paper, a weighted sum of multiple kernels based on kernel alignment is adopted. The corresponding formula is as follows:

$$K_\eta\left(x_i, x_j\right) = \sum_{m=1}^{p}\eta_m K_m\left(x_i, x_j\right) \tag{12}$$

in which η_m, $m = 1, 2, \cdots, P$ is the kernel weight, K_m, $m = 1, 2, \cdots, P$ is a set of basic kernel matrix. Kernel weights can be decided by maximizing the similarity between the combined kernel and the ideal kernel. We can get the following formula [16,42]

$$\max A\left(\mathrm{T}, K_\eta, yy^T\right) = \frac{\left\langle \sum_{m=1}^{P}\eta_m K_m, yy^T\right\rangle_F}{\sqrt{\left\langle \sum_{m=1}^{P}\eta_m K_m, \sum_{m=1}^{P}\eta_m K_m\right\rangle_F \left\langle yy^T, yy^T\right\rangle_F}} = \frac{\sum_{m=1}^{P}\eta_m\left\langle K_m, yy^T\right\rangle_F}{l\sqrt{\sum_{m,j=1}^{P}\eta_m\eta_j\left\langle K_m, K_j\right\rangle_F}} \tag{13}$$

where $y = \left(y_1, y_2, \cdots, y_l\right)^T$, yy^T is an ideal kernel for binary tasks based on training set T. The above formula is equivalent to:

$$\max_\eta \sum_{m=1}^{p}\eta_m\left\langle K_m, yy^T\right\rangle_F$$
$$s.t.\ \sum_{m,j=1}^{p}\eta_m\eta_j\left\langle K_m, K_j\right\rangle_F = C, \eta_m \ge 0 \tag{14}$$

By Lagrange multiplier method, it is converted into its dual problem:

$$\max_\eta \sum_{m=1}^{p}\eta_m\left\langle K_m, yy^T\right\rangle_F - \lambda\left(\sum_{m,j=1}^{p}\eta_m\eta_j\left\langle K_m, K_j\right\rangle_F - C\right)$$
$$s.t.\ \eta_m \ge 0 \tag{15}$$

In (15), λ is changing along with C. According to the definition of kernel alignment, the value of the kernel alignment remains unchanged when the kernel weight η changes linearly. In that, $\lambda = 1$ can be considered, then the above optimization problem (15) becomes

$$\max_\eta \sum_{m=1}^{p}\eta_m\left\langle K_m, yy^T\right\rangle_F - \sum_{m,j=1}^{p}\eta_m\eta_j\left\langle K_m, K_j\right\rangle_F$$
$$s.t.\ \eta_m \ge 0 \tag{16}$$

In order to avoid kernel overfitting its alignment in the training set, we add the constraint of η in the optimization problem, namely, it becomes the following formula:

$$\max_\eta \sum_{m=1}^{p}\eta_m\left\langle K_m, yy^T\right\rangle_F - \sum_{m,j=1}^{p}\eta_m\eta_j\left\langle K_m, K_j\right\rangle_F - \delta\sum_{m=1}^{p}\eta_m^2$$
$$s.t.\ \eta_m \ge 0 \tag{17}$$

Where $\delta > 0$ .Defining a matrix $O_{mj} = \begin{cases}1 & m = j\\ 0 & m \ne j\end{cases}$, the above formula can be changed to

$$\max_\eta \sum_{m=1}^{p}\eta_m\left\langle K_m, yy^T\right\rangle_F - \sum_{m,j=1}^{p}\eta_m\eta_j\left(\left\langle K_m, K_j\right\rangle_F + \delta O_{mj}\right)$$
$$s.t.\ \eta_m \ge 0 \tag{18}$$

Since (18) is a convex quadratic programming problem, the unique solution of η can be obtained. After substituting into equation (11), the synthesis kernel is obtained. Then, it is introduced into the fuzzy support vector machine based on membership degree of FRS. The optimization problem becomes as follows:

$$\max \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j k_\eta \left\langle x_i, x_j \right\rangle$$
$$s.t. \sum_{i=1}^{l}\alpha_i y_i = 0,\ 0 \le \alpha_i \le Cs_i,\ i,j = 1,2,\cdots,l \qquad (19)$$

The decision function $f(x) = \mathrm{sgn}\left(\sum_{i=1}^{l}\alpha_i y_i k_\eta (x_i, x) + b\right)$ is obtained.

4. Experimental Simulation

UCI database proposed by the University of California at Irvine is a database for machine learning. It is a frequently used standard test data set. To verify the feasibility and validity of the proposed method, nine data sets are selected from UCI database in this work. The relevant information is shown in Table 1. Classes 2 and 3 are considered as one class for wine datasets since only two classifications are considered in this paper.

Table 1. Data information

Number	Dataset name	Number of samples	Number of attributes	Number of categories
1	Horse	368	23	2
2	Hepatitis	155	20	2
3	Heart	270	14	2
4	CT	221	37	2
5	Breastcancer	683	10	2
6	Ionosphere	351	34	2
7	Lymphagraphy	148	19	2
8	Wine	178	14	3
9	Wpbc	198	34	2

The experiment is carried out on a PC (CPU: 2.60GHz Memory 4.00GB). The operating system is Windows 8.1 and the experimental tool is MATLAB R2014b. In the experiment, all data is normalized, and the Gaussian kernels with different kernel parameters are used in the kernel function. The alignment optimization problem is solved with the given the penalty coefficient $C = 100$ and $\delta = 10\,\hat{}\,2$. we choose seven Gaussian kernel $k\left(x_i, x_j\right) = \exp\left(-\left\|x_i - x_j\right\|^2 / 2\sigma^2\right)$ with different parameters σ as the basic kernels. The parameters of the seven base kernels $k_i, i = 1,\cdots,7$ for each data set are listed in Table 2.

Table 2 Parameters setting of base kernel

Number	Dataset	Parameters of basic kernel $k_i, i = 1,\cdots,7$
1	Horse	$2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6$
2	Hepatitis	$2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$
3	Heart	$2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$
4	CT	$2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$
5	Breastcancer	$2^1, 2^2, 2^3, 2^4, 2^5, 2^6, 2^7$
6	Ionosphere	$2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$
7	Lymphagraphy	$2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3$
8	Wine	$2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6$
9	Wpbc	$2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0$

In this work, the experiment compares the classification performance of the MFSVM, classical SVM, FSVM and MKSVM. Moreover, the experiment adopts 10-fold-cross-validation method.

The experimental results are shown in Table 3. This Table shows that MFSVM method proposed in this paper has high classification accuracy on most databases. The effectiveness of the method is also verified.

To further verify the performance of MFSVM in noisy environment, in all 9 data sets, we randomly select some samples which are close to the classification surface, then change the class labels. The comparing results of the test accuracy for the four algorithms are shown in Table 4 (the noise ratio is 10%) and Table 5 (the noise ratio is 20%).

When 10% noise is added, the experimental results show that the classification accuracy of MFSVM in all 9 data sets is higher than that of the other three methods, and it has the highest classification accuracy. In the case of 20% noise, although the classification accuracy of the four methods is relatively lower than that of Table 4, MFSVM has the highest classification accuracy in all databases. The results further verify that the MFSVM method proposed in this paper not only avoids the difficult problem of kernel selection, but also has strong anti-noise ability. Therefore, MFSVM is not only feasible, but also has a wider range of applications.

Table 3. Comparison of classification accuracy without noise

Number	Dataset	Test accuracy(%)			
		Svm	Fsvm	Mksvm	Mfsvm
1	Horse	78.00± 3.62	81.24± 3.50	81.78± 3.87	83.15± 3.99
2	Hepatitis	70.06±11.01	69.77± 8.62	70.40± 4.01	71.65± 5.21
3	Heart	78.25± 9.04	80.37± 6.31	82.96± 8.04	83.70± 8.04
4	CT	90.51± 5.10	90.95± 5.60	90.95± 7.73	91.88± 3.49
5	Hreastcancer	96.92± 2.14	96.92± 2.24	96.93± 1.44	97.36± 2.28
6	Ionosphere	94.02± 4.94	94.87± 4.43	95.43± 3.61	96.00± 4.30
7	Lymphography	68.15±12.39	69.59±10.02	70.82±14.69	70.82±14.69
8	Wine	96.15± 4.05	97.69± 3.71	98.46± 3.24	98.46± 3.24
9	Wpbc	77.22± 7.70	80.17± 7.83	81.28± 8.90	82.89± 7.42

Table 4. Comparison of classification accuracy with 10% noise

Number	Dataset	Test accuracy(%)			
		Svm	Fsvm	Mksvm	Mfsvm
1	Horse	80.69± 2.91	81.77± 3.42	81.23± 2.94	82.31± 3.99
2	Hepatitis	67.56±10.21	69.43±10.05	71.65±10.26	72.22±13.09
3	Heart	75.93± 8.05	79.26± 5.84	79.63± 6.36	81.11± 4.76
4	CT	87.77± 5.30	89.60± 3.69	90.95± 5.67	91.86± 5.98
5	Breastcancer	91.96± 3.76	95.33± 2.51	93.69± 3.04	96.78± 1.52
6	Ionosphere	89.17± 4.01	94.02± 3.92	91.44± 4.27	94.58± 3.68
7	Lymphography	63.79±16.69	64.46±16.35	68.36±15.51	70.05±11.36
8	Wine	89.23± 8.27	93.85± 4.86	96.15± 4.05	98.46± 3.24
9	Wpbc	76.39± 9.85	78.39± 8.01	81.94±12.00	82.28±8.01

Table 5. Comparison of classification accuracy with 30% noise

Number	Dataset	Test accuracy(%)			
		Svm	Fsvm	Mksvm	Mfsvm
1	Horse	73.69± 5.76	75.05± 6.01	76.37± 6.34	78.83± 5.01
2	Hepatitis	64.83±10.46	65.74± 9.85	70.11± 8.64	70.40± 7.71
3	Heart	78.15± 6.86	79.63± 8.42	82.22± 7.96	84.07± 9.08
4	CT	83.81± 9.38	88.74± 8.22	89.64± 6.94	91.46± 6.71
5	Breastcancer	85.82± 3.82	88.73± 3.9	93.71± 2.15	95.18±3.00
6	Ionosphere	84.60± 6.08	89.46± 5.22	88.88± 4.57	91.17± 4.94
7	Lymphography	64.26±12.74	67.13± 9.22	68.46±12.67	70.82± 8.81
8	Wine	87.69± 9.73	90.00±11.49	93.85± 4.86	94.62± 7.30
9	Wpbc	73.28± 6.90	79.83± 3.09	82.11±10.86	83.89± 9.30

5. Conclusion

Considering the FSVM and MKL, we Introduced MKL based on kernel alignment into FSVM. Then a multiple kernel fuzzy support vector machine model depend on Kernel-Target alignment is proposed. The simulation results show that this method combines the advantages of FSVM and MKL, and improves the classification performance to a certain extent especially in noisy samples. This can prove the feasibility and validity of this method in dealing with the problems of difficult selection for kernel function and sensitivity to noise data. Furthermore, we will study the better method of kernel combination and how to improve the calculation of sample membership degree since the method in this paper involves the calculation of multiple kernel combination and membership degree.

6. Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grants 61473111, and Research Foundation of Beijing University of Civil Engineering and Architecture (KYJJ2017017).

References

[1] Vapnik V.N. The nature of statistical learning theory. New York. Springer-Verlag. 1999, 69-110.

[2] Vapnik V.N. Statistical learning theory. New York. A Wiley-Interscience Publication. 1998.

[3] Deng N.Y., Tian Y.J. A New Method in data mining Support Vector Machine. Beijing. Science Press. 2004.

[4] Mahdi G.V., Mojtaba T., Puneet M., Zahra I. Temperature modulation of electronic nose combined with multi Class support vector machine classification for identifying export caraway cultivars. Postharvest Biology and Technology. 2018, 138, 134-139.

[5] Chen W., Zhang J.F., Gao M.Y., Shen G.X. Performance improvement of 64-QAM coherent optical communication system by optimizing symbol decision boundary based on support vector machine. Optics Communications. 2018, 410, 1-7.

[6] Bissan G., Joe N.S. High dimensional data classification and feature selection using support vector machines. European Journal of Operational Research. 2018, 265, 993-1004.

[7] Lin C.F., Wang S.D. Fuzzy support vector machines. IEEE Transactions on Neural Networks. 2002, 13, 464-471.

[8] Huang H.P., Liu Y.H. Fuzzy support vector machines for pattern recognition and data mining. Intl.j.of Fuzzy Systems. 2002, 4, 826-835.

[9] Fan Q., Wang Z., Li D.D., Gao D.Q., Zha H.Y. Entropy based fuzzy support vector machine for imbalanced datasets. Knowledge-Based Systems. 2017, 115, 87-99.

[10] Hang J., Zhang J.Z., Cheng M. Application of multi class fuzzy support vector machine classifier for fault diagnosis of wind turbine. Fuzzy Sets and Systems. 2016, 297, 128-140.

[11] Liu Y., Zhang B.L., Chen B., Yang Y.D. Robust solutions to fuzzy one-Class support vector machine. Pattern Recognition Letters. 2016, 71, 73-77.

[12] Bach F.R., Lanckriet G.R.G., Jordan M.I. Multiple kernel learning, conic duality, and the SMO algorithm. Proc. International Conference on Machine Learning. 2004, 6.

[13] Zhao Y.G., Song Z., Feng Z., Shao L. Learning a multiple kernel similarity metric for kinship verification. Information Sciences. 2018, 430–431, 247-260.

[14] Sajjad K.S., Poorya Z.M., Hamid B. Detection of evolving concepts in non-Stationary data streams, A multiple kernel learning approach. Expert Systems with Applications. 2018, 91, 187-197.

[15] Mehmet G., Ethem A. Multiple kernel learning algorithms. Journal of Machine Learning Research. 2011, 12, 2211-2268.

[16] Cristianini N., Shawetaylor J., Elisseeff A., et al. On kernel-target alignment. International Conference on Neural Information Processing Systems. Natural & Synthetic. MIT Press. 2001.

[17] He Q., Wu C. Membership evaluation and feature selection for fuzzy support vector machine based on fuzzy rough sets. Soft Computing. 2011, 15, 1105-1114.

[18] Lin C.F., Wang S.D. Training algorithms for fuzzy support vector machines with noisy data. Pattern Recognition Letters. 2004, 25, 1647-1656.

[19] Christoudias C.M., Urtasun R., Darrell T. Bayesian localized multiple kernel learning. Univ California Berkeley. 2009, 7, 1531-1565.

[20] Ben-Hur A., Noble W.S. Kernel methods for predicting protein-protein interactions. Bioinformatics. 2005, 1, 38-46.

[21] Cristianini N., Shawe-Taylor J. An introduction to support vector machines and other kernel-Based learning methods. Printed in the United Kingdom at the University Press. 2000.

[22] Diego I.M.D., Munoz A., Moguerza J.M. Methods for the combination of kernel matrices within A support vector framework. Machine Learning. 2010, 78, 137-174.

[23] Qiu S., Lane T. A framework for multiple kernel support vector regression and its applications to sirna efficacy prediction. Transactions on Computational Biology & Bioinformatics. 2009, 6, 190.

[24] Varma M., Ray D. Learning the discriminative power-Invariance trade-off, International Conference on Computer Vision. 2007, 1-8.

[25] Du H.Y. Research of reduced multiple kernel support vector machine. Beijing Jiaotong University. 2015.

[26] Wang T., Tian S., Zhao D . An overview of kernel alignment and its applications. Artificial Intelligence Review. 2015, 43, 179-192.

[27] Zhong S., Chen D., Xu Q., et al. Optimizing the gaussian kernel function with the formulated kernel target alignment criterion for two-class pattern classification. Pattern Recognition. 2013, 46, 2045-2054.

[28] Liu X.D., Luo B., Chen Z.Q. Optimal model selection for support vector machines. Journal of Computer Research and Development. 2005, 42, 576-581.

[29] Pawlak Z. Rough sets. Kluwer Academic Publishers. 1991.

[30] Dubois D., Prade H. Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems. 1990, 17, 191-209.

[31] Dubois D., Prade H. Putting rough sets and fuzzy sets together intelligent decision support. Springer Netherlands. 1992, 203-232.

[32] Bhatt R.B., Gopal M. On fuzzy rough sets approach to feature selection. Pattern Recognition Letters. 2005, 26, 965-975.

[33] Chen D.G., Tsang E.C.C., Zhao S.Y. An approach of attributes reduction based on fuzzy T L Rough Sets. International Conference on Systems, Man and Cybernetics. 2007, 486-491.

[34] Chen D., Wang X., Zhao S. Attribute reduction based on fuzzy rough sets.Rough Sets and Intelligent Systems Paradigms. 2007, 381-390.

[35] Chen D., Tsang E.C.C., Yeung D.S., et al. The parameterization reduction of soft sets and its applications. Computers Mathematics with Applications. 2005, 49, 757-763.

[36] Hu Q., Yu D., Xie Z. Information preserving hybrid data reduction based on fuzzy rough techniques. Pattern Recognition Letters. 2006, 27, 414-423.

[37] Tsang E.C.C., Chen D., Yeung D.S. et al. Attributes reduction using fuzzy rough sets. Transactions on Fuzzy Systems. 2008, 16, 1130-1141.

[38] Hassanien A.E. Fuzzy rough sets hybrid scheme for breast cancer detection. Image Vision Computing. 2007, 25, 172-183.

[39] Liu M., Chen D., Wu C. et al. Fuzzy reasoning based on new fuzzy rough set and its application to scheduling problems. Computers Mathematics with Applications. 2006, 51, 1507-1518.

[40] Sarkar M. Ruggedness measures of medical time series using fuzzy-rough sets and fractals. Pattern Recognition Letters. 2006, 27, 447-454.

[41] Chen D.G., Yang Y.Q., Wang H. Granular computing based on fuzzy similarity relations. Soft Computing. 2011, 15, 1161-1172.

[42] Chen L., Chen D., Wang H. Alignment based kernel selection for multi label learning. Neural Processing Letters. 2018, 1-21.