New Mixed Kernel Functions of SVM Used in Pattern Recognition

Hao Huanrui

Harbin University of Science and Technology, China
Harbin Finance University, China

Abstract: *The pattern analysis technology based on kernel methods is a new technology, which combines good performance and strict theory. With support vector machine, pattern analysis is easy and fast. But the existing kernel function fits the requirement. In the paper, we explore the new mixed kernel functions which are mixed with Gaussian and Wavelet function, Gaussian and Polynomial kernel function. With the new mixed kernel functions, we check different parameters. The results shows that the new mixed kernel functions have good time efficiency and accuracy. In image recognition we used SVM with two mixed kernel functions, the mixed kernel function of Gaussian and Wavelet function are suitable for more states.*

Keywords: *Support vector machine, kernel functions, pattern recognition, Wavelet function.*

1. Introduction

Along with the ceaseless development of social productivity, human life, production and scientific research activities are becoming more and more complex. Recognition is the most basic and important activities in human life, production and scientific research. It is the demand of the development of human productivity to use a machine, which helps realize the automatic pattern recognition. Developed on the basis of statistical learning theory [1], Support Vector Machine (SVM) is one of the advanced machine learning algorithm in the area of pattern identification [2, 3]. Due to using the principle of structural risk minimization instead of empirical risk minimization, SVM can handle the problem of small sample size better. As a result of using the idea of kernel function, it can turn the problem from nonlinear space to linear space in order to reduce the complex rate of the problem [4].

The classification of nonlinear problem has been resolved effectively by SVM with kernel functions, while different kinds of kernel functions can lead to different

performance of classification results [5]. Kernel function display a very important role in SVM, linear algorithm can be non-linear and non-linear pattern recognition problem in original input space can be solved by using it. The curse of dimensionality problem could be avoided cleverly by introducing kernel function into SVM.

It is generally believed that there is no more important problems than the nonlinearity of examples from actual space and dimensionality reduction of examples space during the course of the research of recognition [6]. As a key factor of the success of the support vector machine, a more novel theory and more valid method to such problems as nonlinearity of swatch and dimensionality reduction, kernel function theory has attracted sweeping researchers to apply it to existing pattern recognition field and develop it.

The kernel model's scarcity is an important criterion to evaluate a kernel model, good or bad, so how to construct a sparse kernel model is a hot topic in machine learning field. There are two main strategies to establish a kernel model: convex optimization method and greedy method [7, 8]. Using convex optimization method can get the only one global optimal solution and wouldn't get the local optimal solution. The representative model using convex optimization is the support vector machines. The greedy method can get the optimal solution very fast, which produces the optimal solution is not always the global optimal solution, is the global optimal solution or the approximation of global optimal solution. Kernel matching pursuit, projection pursuit and orthogonal least squares; the three models are the examples that use the greedy method [9]. SVM showed excellent performances in solving small size samples, nonlinear and high dimensional pattern recognition problems. Compared with support vector machine models, the kernel matching pursuit has equivalent performance and better scarcity and less computational complexity. The orthogonal least squares model is a simple and high-efficiency model, and it has the characteristics of non-linear model with linear weight, good generalization performance and scarcity.

Next, we describe the SVM and its algorithm.

2. SVM and its algorithm

Support vector machine is a new machine learning technique developed from the middle of the 1990s by Vlaimir Vapnik. Support vector machine is a very specific class of algorithms, characterized by the use of a maximal margin hyper-plane the theory of kernels, the absence of local minima, convex optimization the sparseness of the solution, Mercer's theorem and the capacity control obtained by acting on the margin [10]. A large number of experiments have shown that support vector machine has not only simpler structure, but also better performance, especially its better generalization ability. The support vector machine approach was originally developed to solve binary classification problems.

The basic idea of SVM is that [11]: Select a set of feature subset from the training set and make the feature subset linear division is equivalent to the entire data set divided. Based on limited sample information, SVM finds the best

compromise in the complexity of the model (Learning accuracy of a particular sample) and the learning ability (the ability to identify any sample with no error), to obtain the best generalization ability [12]. Theoretically, this is approximation of structural risk minimization principle. In the condition of fixed machine learning experience risk, minimize VC confidence, and then minimize expected risk sector, to obtain linearly separable classification surface optimization.

In SVM, different kernel function and parameter selection has important implications for algorithm. There are four categories kernel functions in researches and applications.

1) Linear kernel function

(1)
$$K(x, x') = xx'.$$

The SVM in the function is the hyper plane in sample space.

2) Polynomial kernel function

(2)
$$K(x, x_i) = \left[(xx_i) + 1\right]^q.$$

The SVM in the function is q-order polynomial classifier.

3) RBF kernel function

(3)
$$K(x, x_i) = \exp\left[-\frac{|x - x_i|^2}{2\sigma^2}\right].$$

Each general function centre corresponds to a support vector. Output power value is automatically determined by the algorithm.

4) Sigmoid kernel function

(4)
$$K(x, x_i) = \tanh\left(v(x, x_i) + c\right).$$

The SVM is multilayer perceptron containing a hidden layer. Hidden nodes are determined automatically by the algorithm.

3. Improvement of kernels function

Currently, humans usually construct a kernel model using only one type of kernel function, but it is inadequate for some data that the data has different structures, so we will use mixed kernels for this question. We construct a regression model using the orthogonal least squares method and mixed kernels and use the repeated weighted boosting research to speed up the search. It has been proved that the repeated weighted boosting search is a simple and effective global optimization search algorithm, so the model we constructed by using the orthogonal least squares regression, tenable mixed kernels and repeated weighted boosting search has a good scarcity and generalization performance.

Many different types of kernel functions are combined in a certain way; the new mixed kernel function is suitable for non-linear relationship problems.

Polynomial kernel function is a common kernel function, with features of good generalization ability and weak learning ability. Gaussian kernel is typically localized kernel. With the increase of parameters, its scalability gradually weaken.

If these two kernels are combined linearly, the new mixed kernel function is

(5)
$$K = \alpha K_1 + (1 - \alpha)K_2,$$

where K_1 is the Gaussian kernel function, K_2 is the Polynomial kernel function. If $\alpha = 1$, K is the Gaussian kernel function. If $\alpha = 0$ it is Polynomial kernel function.

The sample collection is

(6)
$$\{x_i, y_i\}_{i=1}^{N}.$$

Establish the regression model to fit sample collection:

(7)
$$y(x) = \hat{y}(x) + e(x) = \sum_{i=1}^{M} \omega_i g_i(x) + e(x),$$

where x is the variable of m-dimensional, $y(x)$ is the actual output, $\hat{y}(x)$ is the approximate output of regression model, $e(x)$ is the model error, ω_i is the weighting factor of sub-regression, M is the several of regression model, $g_i(x)$ is the return item;

(8)
$$g_i(x) = \alpha_i k_{1i}(x) + (1 - \alpha)k_{2i}(x), \ \alpha \in (0, 1),$$

where α is the right combination weighting factor of mixed kernel function.

(9)
$$\vec{y} = [y_1, y_2, y_3, \ldots, y_N]^{\mathrm{T}},$$

(10)
$$\vec{e} = [e_1, e_2, e_3, \ldots, e_N]^{\mathrm{T}},$$

(11)
$$\vec{\omega} = [\omega_1, \omega_2, \omega_3, \ldots, \omega_N]^{\mathrm{T}},$$

(12)
$$\vec{G} = [g_1, g_2, g_3, \ldots, g_N]^{\mathrm{T}},$$

(13)
$$g_k = [g_k(x_1), g_k(x_2), g_k(x_3), \ldots, g_k(x_N)]^{\mathrm{T}}.$$

Based on these function

(14)
$$\vec{y} = \vec{G}\vec{\omega} + \vec{e},$$

(15)
$$\vec{G} = \vec{P}\vec{A},$$

(16)
$$\vec{y} = \vec{P}\vec{\theta} + \vec{e},$$

(17)
$$\vec{\theta} = [\theta_1, \theta_2, \theta_3, \ldots, \theta_N]^{\mathrm{T}},$$

(18)
$$\vec{\theta} = \vec{A}\vec{\omega}.$$

The MSE is

(19)
$$J = \vec{e}^{\mathrm{T}}\vec{e} / N = \vec{y}^{\mathrm{T}}\vec{y} / N - \sum_{i=1}^{M} \vec{p}_i^{\mathrm{T}}\vec{p}_i\theta_i^2 / N.$$

So

(20)
$$J_k = J_{k-1} - \frac{1}{N}\vec{p}_k^{\mathrm{T}}\vec{p}_k\theta_k^2.$$

After k times iteration, $J_k < \xi$, ξ is the selected threshold. Then a regression of mixed kernels function is obtained.

4. Results

The experiment is carried on with classic artificial data. Because of strong signal shock of artificial data, it is used to check the reliability of thinning model.

8

Generate 400 equally spaced data points from noisy data. Using cross-validation method, select 200 even point for training, and the remaining 200 odd points for testing.

As illustrated by Figs 1-3, compared with the Gaussian kernel, in the mixed kernel function of Gaussian and Polynomial kernel function, time consumption reduced with low accuracy. It improves time efficiency at the expense of accuracy. Compared with the Gaussian kernel, in the mixed kernel function of Gaussian and Wavelet function, time performance improved significantly, and scarcity and generalization ability of the model remains in the same performance.

Table 1. One-dimensional Data fitting parameters with the New mixed kernel function of Gaussian and Polynomial kernel function and Gaussian kernel function

K	The mixed kernel function of Gaussian and Polynomial kernel function					Gaussian kernel			
	μ_K	α_K	θ_K	σ_K	MSE	μ_K	σ_K	θ_K	MSE
1	–0.0667	1.7524	1.9867	0.8439	0.038	0.095	1.4008	1.0494	0.0237
2	-1.76×10^{-4}	0.0746	0.7134	–0.0676	0.0332	–4.7673	1.0517	–0.2366	0.0201
3	0.8867	0.6558	0.0337	0.0022	0.0327	4.4735	0.6216	–0.1999	0.0164
4	–3.9484	0.669	0.5761	0.226	0.0282	7.451	0.9633	0.1676	0.0148
5	–	–	–	–	–	0.1248	0.7153	–0.0681	0.0144
6	–	–	–	–	–	–7.8283	0.6534	–0.8148	0.014
7	–	–	–	–	–	0.1888	0.1695	–0.0013	0.0136
8	–	–	–	–	–	1.1826	–0.0164	–0.0218	0.0131
	Training error: 0.0281					Training error: 0.0131			
	Test error: 0.0278					Training error: 0.0245			
	Run time: 3.326 s					Training error: 5.057 s			

Table 2. One-dimensional Data fitting parameters with the New mixed kernel function of Gaussian and Wavelet function and Gaussian kernel function

K	The new mixed kernel function of Gaussian and Wavelet function					Gaussian kernel			
	μ_K	α_K	θ_K	σ_K	MSE	μ_K	σ_K	θ_K	MSE
1	0.112	–2.995	1.776	–0.4152	0.026	0.095	1.4008	1.0494	0.0237
2	–4.3976	1.6922	1.6405	2.1945	0.0215	–4.7673	1.0517	–0.2366	0.0201
3	4.301	0.2804	1.3282	–0.2561	0.0165	4.4735	0.6216	–0.1999	0.0164
4	7.8341	0.3582	0.8484	–0.2837	0.0158	7.451	0.9633	0.1676	0.0148
5	–	–	–	–	–	0.1248	0.7153	–0.0681	0.0144
6	–	–	–	–	–	–7.8283	0.6534	–0.8148	0.014
7	–	–	–	–	–	0.1888	0.1695	–0.0013	0.0136
8	–	–	–	–	–	1.1826	–0.0164	–0.0218	0.0131
	Training error: 0.0154					Training error: 0.0131			
	Test error: 0.0207					Training error: 0.0245			
	Run time: 3.067 s					Training error: 5.057 s			

The data is generated in the way:

(21)
$$y(k) = (0.8 - 0.5\exp(-y^2(k-1)))y(k-1) -$$
$$- (0.3 + 0.9\exp(-y^2(k-1)))y(k-2) + 0.1\sin(\pi y(k-1)) + \varepsilon(k).$$

After noise data added, set $y(0) = y(-1) = 0$, produce 1000 sample points, 500 points are selected as the training data, and the remaining 500 points as the test data.

The input data is

(22)
$$\mathrm{x}_k = \left[y(k-1),\ y(k-2) \right]^{\mathrm{T}}.$$

In three-dimensional graphics drawing, take the two components of x as *x* and *y*, output is *z*. Fitting experiments are carried on with Gaussian kernel function, the mixed kernel function of Gaussian and Polynomial kernel function, the mixed kernel function of Gaussian and Wavelet function.
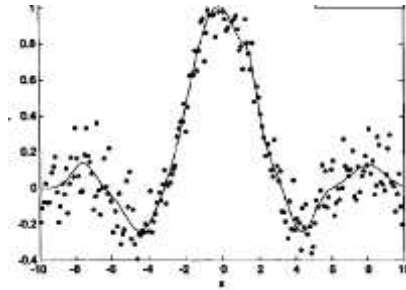
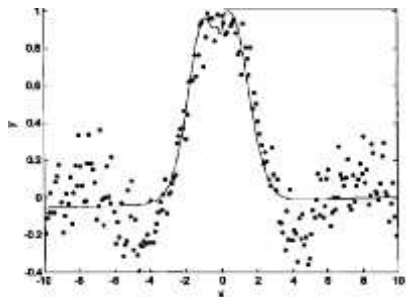Fig. 1. One-dimensional data fitting with Gaussian kernel function

Fig. 2. One-dimensional data fitting with the mixed kernel function of Gaussian
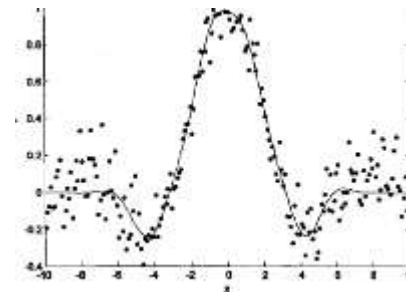and Polynomial kernel function

Fig. 3. One-dimensional data fitting with the mixed kernel function of Gaussian and Wavelet function

10

Table 3. Two-dimensional Data fitting parameters with Gaussian kernel function

K	μ_K		σ_K	θ_K	MSE
1	0.5753	−1.6996	0.7768	−1.6457	0.2779
2	−0.4772	2.1518	1.0084	0.953	0.1659
3	−0.8676	1.6896	0.741	1.5085	0.0115
4	0.4971	−2.5029	1.1832	−3.4364	0.0034
5	−1.486	1.497	0.8769	0.5759	0.0019
6	0.1606	0.4371	0.4477	−0.2939	0.0016
7	−0.665	1.3407	0.4986	−0.4294	0.0015
8	0.7791	−0.9903	0.4005	−0.1076	0.0013
Training error	0.013				
Test error	0.022				
Run time	15.089 s				

Table 4. Two-dimensional Data fitting parameters with the Mixed kernel function of Gaussian and Polynomial kernel function

K	μ_K		σ_K	θ_K	MSE
1	1.65×10^4	-2.70×10^4	-6.69×10^3	-2.28×10^3	−0.4152
2	0.2489	−76.48	−0.192	22.9276	2.1945
3	−0.0184	−0.2376	0.6741	0.6142	−0.2561
4	0.0397	−1.5991	0.6389	0.8691	−0.2837
Training error	0.012				
Test error	5.906×10^{-06}				
Run time	9.861 s				

Table 5. Two-dimensional Data fitting parameters with the Mixed kernel function of Gaussian and Wavelet function

K	μ_K		σ_K	θ_K	MSE
1	1.4351	−2.3469	−0.3797	0.8643	4.2845
2	−0.8774	1.17192	0.4951	0.5691	−2.4526
3	−0.4383	−0.5519	0.4531	0.3902	0.6352
4	0.551	1.2446	0.1189	0.4953	−1.1829
Training error	0.0154				
Test error	0.014				
Run time	11.889 s				

As shown in Tables 3-5, compared with the Gaussian kernel, in the mixed kernel function of Gaussian and Polynomial kernel function, time consumption reduced with low accuracy. It improves time efficiency at the expense of accuracy. Compared with the Gaussian kernel, in the mixed kernel function of Gaussian and Wavelet function, time performance improved significantly, and scarcity and generalization ability of the model remains in the same performance.
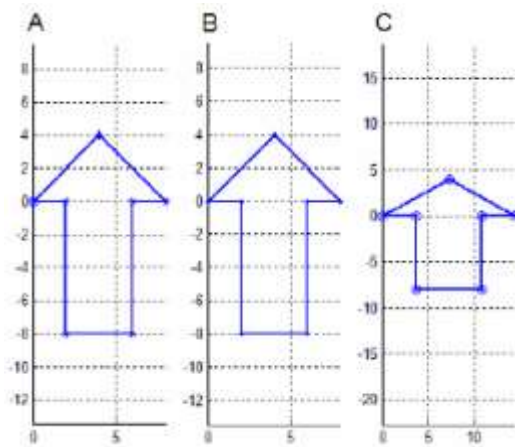
11

Fig. 4. An up arrow and its optimal pattern under improved SVM. Improved SVM is applied on the convex hull of the vertexes of the arrow

Image target recognition is one of the most attractive areas in the field of image understanding and computer vision. It has wide usability in commercial area, civil area and military area. The pattern analysis technology based on kernel methods is a new technology of good performance and strict theory, and has been used in image target recognition.

An experiment is designed, which carried on the Matlab7.4.0 in computer with Pentium1.50GHZ CPU and Windows XP operate system. In this experiment, we have established a number of shapes as simulation data to prove the ability of the algorithm. A is the data dealed with Gaussian kernel function. B is the data dealt with by the mixed kernel function of Gaussian and Polynomial kernel function. C is the data dealt with by the mixed kernel function of Gaussian and Wavelet function.
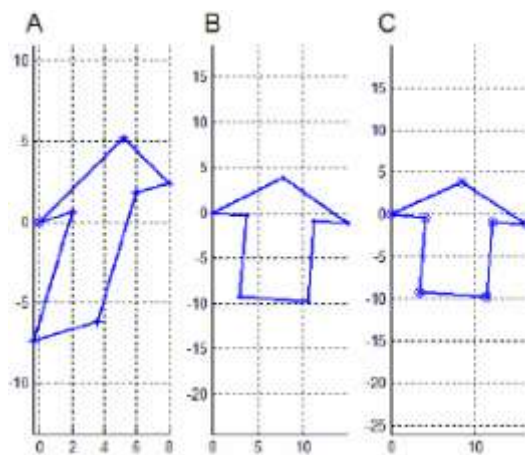

Fig. 5. An oblique arrow and its optimal pattern under improved SVM. Improved SVM is applied on the convex hull of the vertexes of the arrow
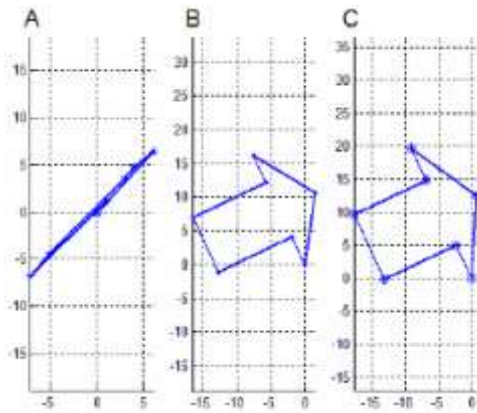
12

Fig. 6. A highly oblique arrow and its optimal pattern under improved SVM. Improved SVM is applied on the convex hull of the vertexes of the arrow
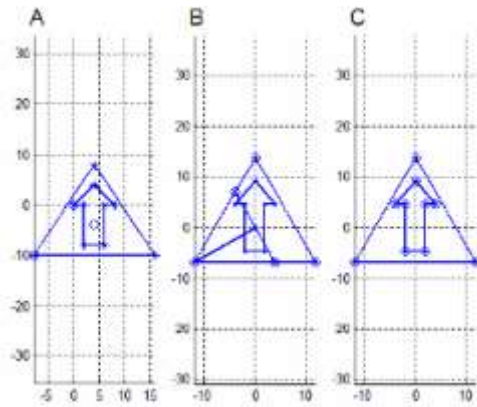

Fig. 7. An up shape of a traffic sign model and its optimal pattern under improved SVM. Improved SVM is applied on the convex hull of the vertexes of the shape
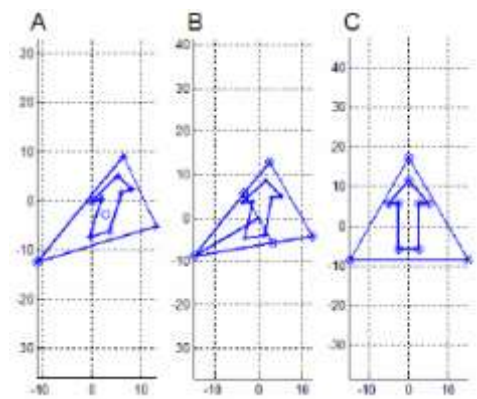

Fig. 8. An oblique shape of a traffic sign model and its optimal pattern under improved SVM. Improved SVM is applied on the convex hull of the vertexes of the shape

As shown in Figs 4-8, the pattern C gets better accuracy within the changing direction. It means that the mixed kernel function of Gaussian and Wavelet function is better in image recognition.

5. Conclusion

This paper has shown an improved SVM with two mixed kernel functions. With the new mixed kernel functions, we check different parameters, showing good time efficiency and accuracy. And then, in image recognition we used SVM with two mixed kernel functions, the mixed kernel function of Gaussian and Wavelet function is suitable for more states.

In actual use of the SVM, we can choose different kernel function based on our needs.

References

1. M e g r i, A. C., I. E l N a q a. Prediction of the Thermal Comfort Indices Using Improved Support Vector Machine Classifiers and Nonlinear Kernel Functions. – Indoor and Built Environment, 2014, 1420326X14539693.
2. O z e r, S., C. H. C h e n, H. A. C i r p a n. A Set of New Chebyshev Kernel Functions for Support Vector Machine Pattern Classification. – Pattern Recognition, Vol. **44**, 2011, No 7, pp. 1435-1447.
3. Y o o n, C., D. K i m, W. J u n g et al. AppScope: Application Energy Metering Framework for Android Smartphone Using Kernel Activity Monitoring – USENIX Annual Technical Conference. 2012, pp. 387-400.
4. P r e e, H., B. H e r w i g, T. G r u b e r et al. On General Purpose Time Series Similarity Measures and their Use as Kernel Functions in Support Vector Machines. – Information Sciences, Vol. **281**, 2014, pp. 478-495.
5. L e e, Y. H., Y. Y. C h o, G. M. C h o. Interior-Point Algorithms for P_ {*}(\ kappa)-LCP Based on a New Class of Kernel Functions. – Journal of Global Optimization, Vol. **58**, 2014, No 1, pp. 137-149.
6. V i e i r a, M. V. C. Interior-Point Methods Based on Kernel Functions for Symmetric Optimization. – Optimization Methods and Software, Vol. **27**, 2012, No 3, pp. 513-537.
7. E l a n g o v a n, M., V. S u g u m a r a n, K. I. R a m a c h a n d r a n et al. Effect of SVM Kernel Functions on Classification of Vibration Signals of a Single Point Cutting Tool. – Expert Systems with Applications, Vol. **38**, 2011, No 12, pp. 15202-15207.
8. C h e n, Y., F. L i u, S. M e i, et al. Toward Adaptive Robust State Estimation Based on MCC by Using the Generalized Gaussian Density as Kernel Functions. – International Journal of Electrical Power & Energy Systems, Vol. **71**, 2015, pp. 297-304.
9. H o n g, S., Z. Z h o u, C. L u et al. 1547. Bearing Remaining Life Prediction Using Gaussian Process Regression with Composite Kernel Functions. – Journal of Vibroengineering, Vol. **17**, 2015, No 2.
10. S h a l e v-S h w a r t z, S., Y. S i n g e r, N. S r e b r o et al. Pegasos: Primal Estimated Sub-Gradient Solver for SVM. – Mathematical Programming, Vol. **127**, 2011, No 1, pp. 3-30.
11. H a n, K. L., S. V. M. T h o m a s, S. M. K o o n t z et al. Adenosine A2A Receptor Agonist-Mediated Increase in Donor-Derived Regulatory T Cells Suppresses Development of Graft-Versus-Host Disease. – Journal of Immunology, Vol. **190**, 2013, No 1, pp. 458-468.
12. T e s s o n, S. V. M., M. M o n t r e s o r, G. P r o c a c c i n i et al. Temporal Changes in Population Structure of a Marine Planktonic Diatom. – PloS One, Vol. **9**, 2014, No 12, e114984.