AMSE JOURNALS-AMSE IIETA publication-2017-Series: Advances B; Vol. 60; N°2; pp 338-356
Submitted Feb. 28, 2017; Revised Apr. 18, 2017; Accepted May 18, 2017

Local Capability Analysis and Comparative Study of Kernel Functions in Support Vector Machine

*Hailun Wang, **Daxing Xu

*Logistics Engineering College, Shanghai Maritime University, Shanghai 200000, China;
College of Electrical and Information Engineering, Quzhou University, Quzhou 324000, China
(xiaohong1920@126.com)
**College of Electrical and Information Engineering, Quzhou University, Quzhou 324000, China
(daxingxu@163.com)

Abstract

Kernel function, the centrepiece of Support Vector Machine (SVM), is classified into local kernel function and global kernel function. The features of the local and global kernel functions can be demonstrated all at once in a combined kernel function. This paper analyses the local capability of SVM kernel function through comparative analysis. Specifically, the local capability of combined kernel function was defined and analysed for the first time; the local capability features of typical kernel functions and combined kernel function were detailed and compared with each other. Finally, the correctness and rationality of the analysis was verified through simulation.

Key words

Local capability, Combined kernel function, Local kernel function, Global kernel function, Support Vector Machine (SVM).

1. Introduction

In 1995, Corinna Cortes and Vapnik et al. created the concept of Support Vector Machine (SVM), a general learning method based on the theory of statistical learning [1]. With strong nonlinear processing and generalization capacities, the method offers a desired solution to such problem as small sample set, non-linearity and high-dimension pattern recognition. Over the years, the SVM has been successfully utilized to tackle various non-linear and non-separable machine learning problems, and thus classification and regression problems. At present, the SVM is attracting extensive attention from academia and being introduced to fault diagnosis. Its role in fault diagnosis is increasingly prominent thanks to the continuous development of SVM technology.

As the core of the SVM, the kernel function can greatly enhance the nonlinear processing capability of the SVM without sacrificing the intrinsic linearity in high-dimensional space. Owing to its major impact on the performance of the SVM, the kernel function has become a focal point in the current SVM research. Scholars at home and abroad have explored the construction, type selection and parameter identification of the kernel function. The construction of the kernel function has to satisfy only one premise: Mercer's theorem [2]. Under this premise, the function was constructed through function approximation, wavelet transform and various other means [3-5]. One of the most popular type selection methods is cross validation, by which different kernel functions are respectively tried to train samples and the kernel function with the smallest overall error is selected as the optimal kernel function. The parameter identification and optimization are the main orientation of the kernel function research.

The SVM constructed by single kernel functions based on single-feature space has a lot of defects in processing unevenly distributed samples. For example, suppose there is a feature containing a sub-feature obeying the polynomial distribution, and another sub-feature following the normal distribution. If the feature is treated with single kernel functions, it is impossible to represent the two different distributions properly but to depict a fraction of the feature [6]. Previous studies [7-9] have attempted to handle the classification problems with a combined kernel function featuring strong local information and global information treatment ability. This strategy can indeed make up for the defects of single kernel function in processing local and global information. However, there is not yet an effective method to optimize the weighting coefficients of the two basic kernel functions within the strategy.

Depending on the local and global capabilities, the kernel function is classified into local kernel function and global kernel function. The former is good at learning, and the latter does well in extrapolation [10-11]. Currently, there are numerous different types of kernel functions, each of which has its unique features and nonlinear processing abilities. Without the loss of the basic features of the original kernel functions, the local kernel functions and the global kernel function can be linearly combined into a new kernel function, namely, combined kernel function. The combined kernel function integrates the strengths of local and global kernel functions to reflect the exact features of actual samples [12-13].

Despite the definition and feature analysis of classical kernel functions in [14], the existing research on the kernel function mainly concentrates on the construction, type selection and parameter identification of the function. There is rare report on the specific local capability of kernel function and its effect on SVM performance. In fact, the local capability of kernel function is helpful to reveal the mechanism of SVM, and enhance the latter's classification and regression abilities.

To this end, this paper analyses the local capability of SVM kernel function through comparative analysis. Specifically, the local capability of combined kernel function was defined and analyzed for the first time; the local capability features of typical kernel functions and combined kernel functions were detailed and compared with each other. Finally, the correctness and rationality of the analysis was verified through simulation.

The remainder of this paper is arranged as follows. Section 2 describes the problem; Section 3 defines and analyses the local capability of a combined kernel function; Section 4 discusses and compares the local capabilities of kernel functions; Section 5 gives the simulation examples; Section 6 wraps up the research with some meaningful conclusions.

2. Problem Description

2.1 Typical Kernel Function

The structure diagram of the SVM is shown in Figure 1. The upside of the SVM is attributable to the introduction of kernel function, which tactfully solves the problem of nonlinear classification. Through inner product operation, the kernel function converts high-dimensional primitive space into low-dimensional feature space. The introduction of kernel function greatly facilitates the learning control, as it improves the nonlinear processing ability of the SVM without sacrificing the inherent linearity of the SVM in the high-dimensional space. Due to the unique features of each kernel function, the SVMs based on different kernel functions differ in generalization ability. As mentioned above, the kernel function is classified into local kernel function and global kernel function. The local kernel function boasts strong learning ability in that it is good at extracting the locality of the sample, its value is only affected by neighbouring data points, and it does well in interpolation, while the global kernel function enjoys strong generalization ability in that it is good at extracting the global features of the sample, and its value is only affected by distant data points [14].

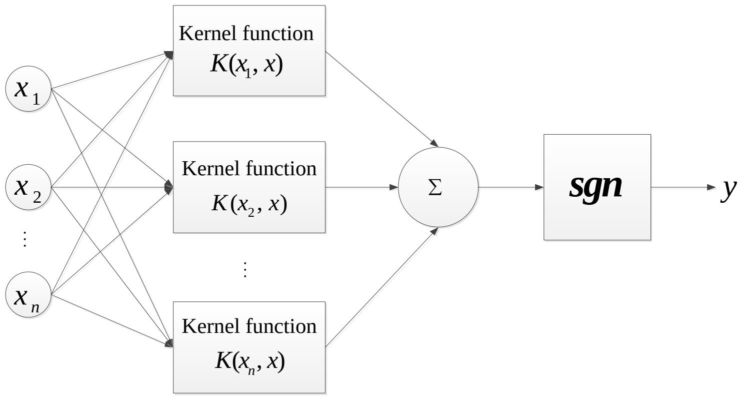

Fig.1. The Structure Diagram of the SVM

Let $K(x_i, x_j)$ denote the kernel function, where x_i and x_j represent the sample data. There are six types of typical kernel functions, including two local kernel functions and three global kernel functions.

The two typical local kernel functions are as follows:

(1) Gaussian kernel function (RBF kernel)

$$K(x_i, x_j)=\boldsymbol{exp}(-\left\|x_i - x_j\right\|^2/\sigma^2) \tag{1}$$

where σ>0 are the kernel parameters. The RBF kernel has stronger locality than general kernel functions. Since the function only needs to determine one parameter, it is relatively simple to establish the kernel function model. That is why it is one of the most widely used kernel functions [15].

(2) Fourier kernel function

$$K(x_i, x_j)=\frac{(1\text{-}q^2)(1\text{-}q)}{2(1-2q\,\boldsymbol{cos}(x_i - x_j)+q^2)} \tag{2}$$

where 0<q<1. The SVM based on Fourier kernel function has also been extensively applied [16].

The three typical global kernel functions are as follows:

(3) Linear kernel function

$$K(x_i, x_j) = x_i \cdot x_j \tag{3}$$

Featuring few parameters and fast speed, the linear kernel function finds the optimal generalization for the SVM in the original space [17].

(4) Polynomial kernel function

$$K(x_i, x_j) = ((x_i \cdot x_j)+c)^q \tag{4}$$

where c and q are kernel parameters ($c \geq 0; q \in N$). This kernel function helps to derive the q^{th} order polynomial classifier. When $c=1$, the function is a common polynomial kernel, in which the mapping dimension and computing load are positively correlated with the value of q. Excessively high value of q will increase the VC dimension of the function set, complicate the learning machine, and weaken the generalization ability of the SVM. The phenomenon of "over-fitting" is very likely to occur in this case [18].

(5) Sigmoid kernel function

$$K(x_i, x_j)=\boldsymbol{tanh}(\lambda(x_i \cdot x_j') + \varphi) \tag{5}$$

where λ and φ are kernel parameters ($\lambda>0$, $\varphi<0$). This function makes it possible to develop a multilayer perceptron with a hidden layer for the SVM so that the learning machine finds the global optimum instead of the local minimum. The function also guarantees that the SVM can generalize unknown samples well and avoid over-learning [19].

2.2 Research Motivations

Whereas each kernel has its own strengths, weaknesses and unique features, the decision function of the SVM varies substantially in classification performance depending on the specific kernel function adopted by the SVM. However, it is difficult for the SVM to achieve high performance based on a single kernel function only. The existing research on the kernel function mainly focuses on the selection of combined parameters and internal parameters, and seldom examines the SVM performance from the local capability of kernel function. In fact, the local capability of kernel function helps to reveal the mechanism of SVM, and enhance the latter's classification and regression abilities. For this reason, the local capability of kernel function was taken as the object of this research. Hereinto, the local capability of the combined kernel function was defined and analysed for the first time, and the local capability was discussed for each kernel function. The research lays a solid basis for ascertaining the relationship between local capability and kernel function.

3. Definition of Local Capability and Feature Analysis of Kernel Function

3.1 Definition of Local Capability of Kernel Function

In the existing research, there is no clear definition about the local capability of kernel function. Some scholars expressed the concept of "local kernel function" as "a kernel function whose influence mainly concentrates in the neighbourhood of the test point, and dwindles substantially on samples far away from the test point." In other words, a kernel function can be referred to as a local kernel function, provided that its function value fluctuates significantly near the test point and varies slightly at points far from the test point. Considering the difference between the left and right sides near the neighbourhood and the difference between left and right samples far from the test point, the author selected two endpoints near the test point, simulated the "variation of function value near the test point" with the slope change of the two endpoints, and took account of the interval length of the endpoints. In light of this, the local capability of kernel function is defined as follows:

Definition 1: Let $K(x_i, x_j)$ be a kernel function. Select two endpoints x_1 and x_2 in the vicinity of a test point x_0. Denote the slope of the kernel function at the endpoints as $K`(x_0, x_1)$ and $K`(x_0, x_2)$, respectively, and denote the local capability of the kernel function as G. The specific structure diagram of the kernel function is shown in Figure 2. Then, the local capability of the kernel function is believed to hinge on two quantities: endpoint slope difference and endpoint interval length. Hence, the local capability of the kernel function can be expressed as:

$$G = \frac{\left| K^{'}(x_0, x_2) - K^{'}(x_0, x_1) \right|}{x_2 - x_1} \tag{6}$$

By this definition, the local capability is positively correlated with the endpoint slope. The specific value of local capability depends on the test point, the two endpoints, and the internal parameters of the kernel function.

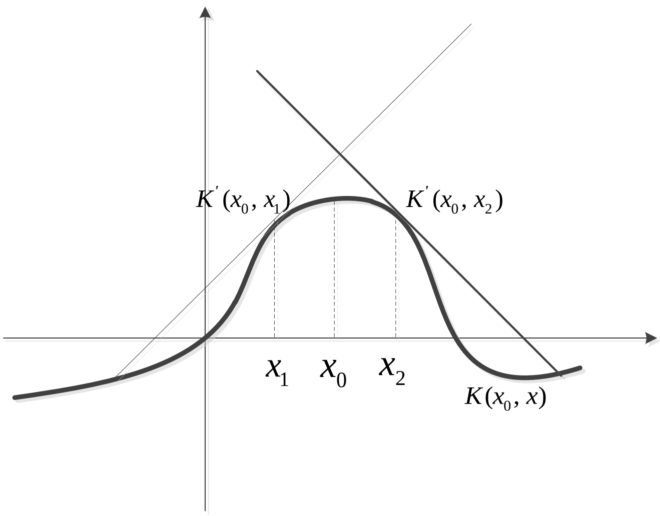

Fig.2. Structural Diagram of Local Capability of the Kernel Function

3.2 Local Capability of Combined Kernel Function

This sub-section gives the expression for the local capability of combined kernel function. With varied interpolation and extrapolation abilities, different kernel functions have different capabilities of learning and generalization. If these kernel functions are combined into a new kernel function, the resulting combined function will do well in both learning and generalization, which fits the purpose of the SVM. The performance of the combined kernel function can be controlled by integrating the prior knowledge of the process into the function through parameter adjustment. The local capability of combined kernel functions are expressed in the formula below.

Theorem 1. If $K_1(x_i, x_j)$, $K_2(x_i, x_j)$, …, $K_n(x_i, x_j)$ are n arbitrary kernel functions, the test point and the two endpoints are x_0, x_1 and x_2, respectively, and P_i is the weighting coefficient of the i-th kernel function, then the local capability of the combined kernel function $\sum_{i=1}^{n} p_i K_i\,(x_0, x)$ is expressed as:

Proof: Assuming that the derivative of the i-th kernel function is $K_i^{'}(x_i, x_j)$, the local capability G_i of the i-th kernel function can be expressed as follows according to Definition 1:

$$G_i = \frac{\left|(K_i^{'}(x_0, x_2) - K_i^{'}(x_0, x_1))\right|}{x_2 - x_1} \tag{7}$$

Then, the local capability of the combined kernel function is:

$$G = \frac{\left|\sum_{i=1}^{n} p_i K_i^{'}(x_0, x_2) - \sum_{i=1}^{n} p_i K_i^{'}(x_0, x_1)\right|}{x_2 - x_1} = \frac{\left|\sum_{i=1}^{n} p_i (K_i^{'}(x_0, x_2) - K_i^{'}(x_0, x_1))\right|}{x_2 - x_1} \tag{8}$$

As for the combined kernel function, there is the following theorem:

Theorem 2. If the weighting coefficient in Theorem 1 satisfies $\sum_{i=1}^{n} p_i = 1$ in Theorem 1, the local G capability of the combined kernel function $p_i \sum_{i=1}^{n} K_i(x_0, x)$ will not exceed the maximum local capability of single kernel functions: $G \le \max_{i=1,2...n}(G_i)$

where G_i is the local capability of the i-th kernel function.

Proof: The following formula is obtained by scaling Formula (8):

$$G \le \sum_{i=1}^{n} \frac{\left|p_i (K_i^{'}(x_0, x_2) - K_i^{'}(x_0, x_1))\right|}{x_2 - x_1} = \sum_{i=1}^{n} p_i G_i \le \sum_{i=1}^{n} p_i \max_{i=1,2...n}(G_i) \tag{9}$$

Substituting $\sum_{i=1}^{n} p_i = 1$ into the above formula, we can get: $G \le \max_{i=1,2...n}(G_i)$

According to Theorem 2, the local capability of the combined kernel function is no greater than the maximum local capability of any constitutive kernel function. Since the combined kernel function is the weighted average of all constitutive kernel functions, the combined kernel function is bound to have a smaller local capability than the maximum constitutive kernel function. However, the remaining capabilities of the combined kernel function will be enhanced. Moreover, the combined kernel function will carry more diverse local and global features thanks to the varied weighting coefficients.

4. Comparative Analysis of Local Capabilities of Kernel Functions

This section mainly calculates and compares the local capabilities of kernel functions.

4.1 Calculation of Local Capabilities for Single Kernel Functions

If the test point, the left endpoint, and the right endpoint are denoted as x_0, x_1 and x_2, respectively, the load capability of each typical kernel function can be calculated by Formula (6).

The local capability of RBF kernel is:

$$\begin{aligned} G_G &= \frac{\left|(K^{'}(x_2,x_0)-(K^{'}(x_1,x_0))\right|}{x_2-x_1} \\ &= \frac{2(x_2-x_0)e^{-(x_2-x_0)^2/\sigma^2}+2(x_0-x_1)e^{-(x_1-x_0)^2/\sigma^2}}{(x_2-x_1)\sigma^2} \end{aligned} \tag{10}$$

The local capability of Fourier kernel function is:

$$\begin{aligned} G_L &= \frac{\left|(K^{'}(x_2,x_0)-(K^{'}(x_1,x_0))\right|}{x_2-x_1} \\ &= \frac{x_2x_0-x_1x_0}{(x_2-x_1)}=0 \end{aligned} \tag{11}$$

The local capability of linear kernel function is:

$$\begin{aligned} G_F &= \frac{\left|(K^{'}(x_2,x_0)-(K^{'}(x_1,x_0))\right|}{x_2-x_1} \\ &= \frac{\left|\frac{(1\text{-}q^2)(1\text{-}q)\,\boldsymbol{sin}(x_2-x_0)}{(1-2q\,\boldsymbol{cos}(x_2-x_0)+q^2)^2}-\frac{(1\text{-}q^2)(1\text{-}q)\,\boldsymbol{sin}(x_1-x_0)}{(1-2q\,\boldsymbol{cos}(x_1-x_0)+q^2)^2}\right|}{x_2-x_1} \end{aligned} \tag{12}$$

The local capability of polynomial kernel function is:

$$\begin{aligned} G_P &= \frac{\left|(K^{'}(x_2,x_0)-(K^{'}(x_1,x_0))\right|}{x_2-x_1} \\ &= \frac{q\left|x_0\right|\left|(x_0x_2+c)^{q-1}-(x_0x_1+c)^{q-1}\right|}{x_2-x_1} \end{aligned} \tag{13}$$

The local capability of Sigmoid kernel function is:

$$G_S = \frac{\left|(K'(x_2, x_0) - (K'(x_1, x_0))\right|}{x_2 - x_1} = \frac{\lambda\left|x_0\right|\left|\tanh^2(\lambda(x_0 \cdot x_1) + \varphi)\text{-}\tanh^2(\lambda(x_0 \cdot x_2) + \varphi)\right|}{x_2 - x_1} \tag{14}$$

Note: According to Formulas (10)~(14), the local capability of kernel function depends on the internal parameters of the kernel function, in addition to the selected test point and left or right endpoint.

4.2 Comparison of Local Capabilities among Various Kernel Functions

4.2.1 Comparison of Local Capability between RBF Kernel and Fourier Kernel Function

Theorem 3: If the test point, the left endpoint, and the right endpoint are denoted as x_0, x_1 and x_2, and the distance between x_0 and x_1 equals that between x_0 and x_2, i.e., $\Delta x = x_2 - x_0$, then the local capability of RBF kernel is stronger than that of Fourier kernel function when $\sigma < \frac{\sqrt{2(1-q)}}{\sqrt{(1+q)}}$, weaker than the latter when $\sigma > \frac{\sqrt{2(1-q)}}{\sqrt{(1+q)}}$, equal to the latter when $\sigma = \frac{\sqrt{2(1-q)}}{\sqrt{(1+q)}}$.

Proof: According to Formulas (10) and (11), we have:

$$G_G\text{-}G_F = \frac{2(x_2 - x_0)e^{-(x_2-x_0)^2/\sigma^2} + 2(x_0 - x_1)e^{-(x_1-x_0)^2/\sigma^2}}{(x_2 - x_1)\sigma^2} - \frac{\left|\frac{(1\text{-}q^2)(1\text{-}q)\,\boldsymbol{sin}(x_2 - x_0)}{(1-2q\,\boldsymbol{cos}(x_2 - x_0)+q^2)^2} - \frac{(1\text{-}q^2)(1\text{-}q)\,\boldsymbol{sin}(x_1 - x_0)}{(1-2q\,\boldsymbol{cos}(x_1 - x_0)+q^2)^2}\right|}{x_2 - x_1} \tag{15}$$

Assuming that x_0, x_1 and x_2 are fixed and $x_2 - x_0 = x_0 - x_1$, then:

$$G_G\text{-}G_F = \frac{2}{\sigma^2}e^{-(x_2-x_0)^2/\sigma^2} - \frac{(1\text{-}q^2)(1\text{-}q)\,\boldsymbol{sin}(x_2 - x_0)}{(1-2q\,\boldsymbol{cos}(x_2 - x_0)+q^2)^2(x_2 - x_0)} \tag{16}$$

When the endpoints x_1 and x_2 are approaching the test point x_0,

$$\lim_{x_2 \to x_0} (G_G\text{-}G_L) = \frac{2}{\sigma^2} - \lim_{x_2 \to x_0} \frac{(1\text{-}q^2)(1\text{-}q)\,\boldsymbol{sin}(x_2 - x_0)}{(1-2q\,\boldsymbol{cos}(x_2 - x_0)+q^2)^2 (x_2 - x_0)} \\ = \frac{2}{\sigma^2} - \frac{1+q}{(1\text{-}q)^2} \tag{17}$$

If $\frac{2}{\sigma^2} - \frac{1+q}{(1\text{-}q)^2} > 0$, the solution $\sigma < \frac{\sqrt{2}(1\text{-}q)}{\sqrt{1+q}}$ is obtained, indicating that the RBF kernel has stronger local capability than the Fourier kernel function.

If $\frac{2}{\sigma^2} - \frac{1+q}{(1\text{-}q)^2} < 0$, the solution $\sigma > \frac{\sqrt{2}(1\text{-}q)}{\sqrt{1+q}}$ is obtained, indicating that the RBF kernel has weaker local capability than the Fourier kernel function.

If $\frac{2}{\sigma^2} - \frac{1+q}{(1\text{-}q)^2} = 0$, the solution $\sigma = \frac{\sqrt{2}(1\text{-}q)}{\sqrt{1+q}}$ is obtained, indicating that the RBF kernel has an equal local capability to the Fourier kernel function.

4.2.2 Local Capability Comparison between Local Kernel Function and Global Kernel Function

By the definition of local capability $G = \frac{|k_{x2} - k_{x1}|}{x_2 - x_1} = \geq 0$, the linear kernel function has the weakest local capability with G_L=0. Below is a pairwise comparison of local capability between local and global kernel functions.

Theorem 4: If the test point, the left endpoint, and the right endpoint are denoted as x_0, x_1 and x_2, then RBF kernel has stronger local capability than polynomial kernel function.

Proof: According to Formulas (10) and (13), we have:

$$G_G\text{-}G_p = \frac{2}{\sigma^2} e^{-(x_2 - x_0)^2/\sigma^2} - \frac{q|x_0|\left|(x_0 x_2 + c)^{q-1} - (x_0 x_1 + c)^{q-1}\right|}{2(x_2 - x_0)} \tag{18}$$

When the endpoints x_1 and x_2 are approaching the test point x_0, $\lim_{x_2 \to x_0} (G_G - G_P) = \frac{2}{\sigma^2} > 0$, indicating that the RBF kernel has stronger local capability than the polynomial kernel function.

Theorem 5: If the test point, the left endpoint, and the right endpoint are denoted as x_0, x_1 and x_2, then RBF kernel has stronger local capability than sigmoid kernel function.

Proof: According to Formulas (10) and (14), we have:

$$G_G\text{-}G_S = \frac{2}{\sigma^2} e^{-(x_2-x_0)^2/\sigma^2} - \frac{\lambda|x_0|\left|\tanh^2(\lambda(x_0 \cdot x_1)+\varphi)\text{-}\tanh^2(\lambda(x_0 \cdot x_2)+\varphi)\right|}{2(x_2-x_0)} \tag{19}$$

When the endpoints x_1 and x_2 are approaching the test point x_0, $\lim_{x_2 \to x_0}(G_G - G_s) = \frac{2}{\sigma^2} > 0$, indicating that the RBF kernel has stronger local capability than the sigmoid kernel function.

Theorem 6: If the test point, the left endpoint, and the right endpoint are denoted as x_0, x_1 and x_2, then Fourier kernel function has stronger local capability than polynomial kernel function.

Proof: According to Formulas (11) and (13), we have:

$$G_F\text{-}G_p = \frac{(1\text{-}q^2)(1\text{-}q)\,\boldsymbol{sin}(x_2-x_0)}{(1-2q\,\boldsymbol{cos}(x_2-x_0)+q^2)^2(x_2-x_0)} - \frac{q|x_0|\left|(x_0x_2+c)^{q-1}-(x_0x_1+c)^{q-1}\right|}{2(x_2-x_0)} \tag{20}$$

When the endpoints x_1 and x_2 are approaching the test point x_0, $\boldsymbol{lim}_{x_2 \to x_0}(G_F\text{-}G_P) = \frac{1+q}{(1\text{-}q)^2} > 0$, indicating that the Fourier kernel function has stronger local capability than the polynomial kernel function.

Theorem 7: If the test point, the left endpoint, and the right endpoint are denoted as x_0, x_1 and x_2, then Fourier kernel function has stronger local capability than sigmoid kernel function.

Proof: According to Formulas (11) and (14), we have:

$$G_F\text{-}G_S = \frac{(1\text{-}q^2)(1\text{-}q)\,\boldsymbol{sin}(x_2-x_0)}{(1-2q\,\boldsymbol{cos}(x_2-x_0)+q^2)^2(x_2-x_0)} - \frac{\lambda|x_0|\left|\tanh^2(\lambda(x_0 \cdot x_1)+\varphi)\text{-}\tanh^2(\lambda(x_0 \cdot x_2)+\varphi)\right|}{2(x_2-x_0)} \tag{21}$$

When the endpoints x_1 and x_2 are approaching the test point x_0, $\boldsymbol{lim}_{x_2 \to x_0}(G_F\text{-}G_S) = \frac{1+q}{(1\text{-}q)^2} > 0$, indicating that the Fourier kernel function has stronger local capability than the sigmoid kernel function.

4.2.3 Local Capability Comparison between Polynomial Kernel Function and Sigmoid Kernel Function

Theorem 8. If the test point, the left endpoint, and the right endpoint are denoted as x_0, x_1 and x_2, then polynomial kernel function has the same local capability with sigmoid kernel function.

Proof: According to Formulas (13) and (14), we have:

$$G_P\text{-}G_S = \frac{q|x_0|\left|(x_0x_2+c)^{q-1}-(x_0x_1+c)^{q-1}\right|}{2(x_2-x_0)} - \frac{\lambda|x_0|\left|\tanh^2(\lambda(x_0\cdot x_1)+\varphi)\text{-}\tanh^2(\lambda(x_0\cdot x_2)+\varphi)\right|}{2(x_2-x_0)}$$

When the endpoints x_1 and x_2 are approaching the test point x_0, $\lim_{x_2\to x_0}(G_p\text{-}G_S)=0$, indicating that the polynomial kernel function has the same local capability with the sigmoid kernel function.

4.2.4 Local Capability Comparison between Combined Kernel Function and Local Kernel Function

Theorem 9. If the test point, the left endpoint, and the right endpoint are denoted as x_0, x_1 and x_2, and the combined kernel function consists of a typical local kernel function and a typical global kernel function, then the local kernel function has greater local capability than the combined kernel function.

Proof: Let $K_1(x_0, x)$ be a local kernel function and $K_2(x_0, x)$ be a global kernel function. If the weighting coefficients of $K_1(x_0, x)$ and $K_2(x_0, x)$ are denoted as P_1 and P_2, respectively, and the local capabilities of the two functions are denoted as G_l and G_g, respectively, then:

When the endpoints x_1 and x_2 are approaching the test point x_0, $\lim_{x_2\to x_0}(G_l)>0$. Considering that the local capability of the combined kernel function is $G\le \max_{i=1,2\ldots n}(G_i)$ (Theorem 1), it is concluded that the local kernel function has stronger local capability than the combined kernel function.

Theorem 10. If the test point, the left endpoint, and the right endpoint are denoted as x_0, x_1 and x_2, and the combined kernel function consists of two typical local kernel functions, then the local capability of the combined kernel function falls between those of two kernel functions.

Proof: Let $K_1(x_0, x)$ and $K_2(x_0, x)$ be the two local kernel functions. If the weighting coefficients of $K_1(x_0, x)$ and $K_2(x_0, x)$ are denoted as P_1 and P_2, respectively, and the local capabilities of the two functions are denoted as G_l and G_g, respectively, then:

When the endpoints x_1 and x_2 are approaching the test point x_0, $\lim_{x_2\to x_0}(G_{l1})=G_1$ and $\lim_{x_2\to x_0}(G_{l2})=G_2$. Considering that the local capability of the combined kernel function is

$\lim_{x_2 \to x_0}(G) = p_1G_1 + p_2G_2$ (Theorem 8), we have $\min_{i=1,2}(G_i) < p_1G_1 + p_2G_2 < \max_{i=1,2}(G_i)$, which means the local capability of the combined kernel function falls between those of two kernel functions.

4.2.5 Comparative Analysis on Local Capabilities of Kernel Functions

Based on the previous theorems, this sub-section comprehensively compares the local capabilities of local kernel functions, global kernel functions and their combined kernel functions. As mentioned above, the local capability of kernel function is closely related to the test point, left and right endpoints and the kernel function parameters. Rather than acquire the specific value of local capability of a certain kernel function, the author aimed to express the relative strength of local capabilities of the kernel functions.

The following results are drawn from the comparative analysis: among the local kernel functions, RBF kernel has stronger local capability than Fourier kernel function; local kernel function possesses much stronger local capability than global kernel function; the linear kernel function has the weakest local capability (0); the local capabilities of the other global kernel functions are also 0 when the endpoints x_1 and x_2 are approaching the test point x_0. To sum up, the local kernel function boasts the strongest local capability; the local capability of the combined kernel function involving a local and a global kernel function is stronger than that of the global kernel function and weaker than that of the local kernel function; the global kernel function comes bottom in terms of local capability.

5. Experimental Simulation

5.1 Verification of Local Capability of Single Kernel Function

The kernel parameters are set as follows: RBF (σ=0.1), Fourier kernel function (q=0.8), polynomial kernel function (c=1; q=2); sigmoid kernel function (λ=2; φ=-P_i/4); the test point: $x_0=0.5$; the endpoints: $x_1=0.45$ and $x_2=0.55$. The calculated local capabilities are listed in Table 1.

Tab.1. Local Capabilities of Single Kernel Functions

Kernel function	Gaussian kernel function	Fourier kernel function	Linear kernel function	Polynomial kernel function	Sigmoid kernel function
Local Capability Value	155.7602	32.6401	0	0.5	0.5114

As shown in Table 1, the local kernel function has much stronger local capability than the global kernel function; RBF kernel possesses stronger local capability than Fourier kernel function; the local capability of linear kernel function is 0 because the slope is constant at any point of the function; the polynomial kernel function has the same local capability with the sigmoid kernel function. The simulation results are in good agreement with the results of the previous comparative analysis.

5.2 Verification of Local Capability Features of Rbf

The following experiment was conducted for the local kernel function RBF to test the local capability variation with the kernel parameter σ. The experimental procedures are presented below and its results are shown in Table 2.

Tab.2. Local Capability Variation with Kernel Parameter σ of RBF

Kernel parameter value	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$
Local Capability Value	155.7602	46.9707	21.6134	12.3062

It is obvious that the local capability gradually decreased with the increase in the kernel parameter σ. The trend verifies the correctness of Theorem 2.

5.3 Verification of Local Capability of Combined Kernel Functions

Following the parameter settings in 5.1, the local capabilities of combined kernel functions were calculated with the weighting coefficients of 0.5 and 0.5. The calculated results are listed in Table 3.

Tab.3. Local Capabilities of Combined Kernel Functions Involving Two Kernel Functions

Local Capability Value	Fourier kernel function	Linear kernel function	Polynomial kernel function	Sigmoid kernel function
Kernel function	94.2001	77.8801	77.6301	77.6244
Fourier kernel function	——	16.3201	16.0701	16.0644
Linear kernel function	——	——	0.25	0.2557
Polynomial kernel function	——	——	——	0.5057

As shown in Table 3, if the combined kernel functions are ranked in descending order of local capability, the combined kernel function involving two local kernel functions will come first, followed by the combined kernel function involving a local kernel function and a global kernel function, while the combined kernel function involving two global kernel functions will come bottom of the ranking. In the meantime, the combined kernel function outshines the two single constituent kernel functions in local capability.

Next, the author calculated the local capabilities of the combined kernel functions involving three kernel functions. Such a combined kernel function either consists of two local kernel functions and one global kernel function (two-local-one-global), or one local kernel function and two global kernel functions (one-local-two-global). The internal kernel parameters remain the same, and the weighting coefficients are 0.5,0.3 and 0.2. The results are presented in Tables 4 and 5.

Tab.4. Local Capabilities of Combined Kernel Functions Involving Two Local Kernel Functions and One Global Kernel Function

Local Capability Value	Linear kernel function	Polynomial kernel function	Sigmoid kernel function
Gaussian kernel function Fourier kernel function	115.5601	108.9706	115.5090

Tab.5. Local Capabilities of Combined Kernel Functions Involving One Local Kernel Function and Two Global Kernel Functions

Local Capability Value	Linear Kernel Function Polynomial kernel function	Linear kernel function Sigmoid kernel function	Polynomial kernel function Sigmoid kernel function
Gaussian kernel function	108.9821	108.9810	108.8810
Fourier kernel function	22.7981	22.7969	22.8031

It can be seen from Table 4 that the local capability of any of the three "two-local-one-global" combined kernel functions is below the maximum local capability of the single kernel functions, which validates Theorem 2. Moreover, the three combined kernel functions have almost equivalent local capabilities. This is because the local capability of each combined kernel function mainly comes from the two local kernel functions.

As can be seen from Table 5, each of the "one-local-two-global" combined kernel functions has a stronger local capability than any of its constituent single kernel function, owing to the influence of the local kernel function. In addition, the combined kernel function involving RBF outperforms that involving Fourier kernel function in local capability.

Conclusion

The research of kernel function is an important way to improve the SVM performance. However, the existing studies on kernel function mainly focus on type selection and parameter identification. In light of the problem, this paper analyses the local capability of SVM kernel function through comparative analysis. Specifically, the local capability of combined kernel function was defined and analysed for the first time; the local capability features of typical kernel functions and combined kernel function were detailed and compared with each other. Finally, the correctness and rationality of the analysis was verified through simulation.

Acknowledgements

The authors are grateful to the anonymous referees for their healthy and useful comments towards the improvement of this paper. This work was supported by the Natural Science Foundation of China (61403229, 61503213, 61403113), Zhejiang Provincial Natural Science Foundation of China (LQ17F030005).

References

1. K. Sopyła, P. Drozda, Stochastic gradient descent with Barzilai–Borwein update step for SVM, 2015, Information Sciences, vol. 31, no. 6, pp. 218-233.
2. C. Cortesc, V.N. Vapnik, Support vector networks, 1995, Machine Learning, vol. 20, no. 1, pp. 273 -297.
3. G. Baudat, F. Anouar, Generalized discriminant analysis using a kernel approach, 2010, Neural computation, vol. 12, no. 10, pp. 2385-2404.
4. M. Szummer, T.S. Jaakkola, Kernel expansions with unlabeled examples, 2011, Advances in Neural Information Processing Systems, pp. 626-632.
5. P. Rujan, M. Marchand, Computing the Bayes kernel classifier, 1999.
6. C.W. Hsu, C.C. Chang, C.J. Lin, A practical guide to support vector classification, 2013, Department of information engineering, National Taiwan University, pp. 1-16.

7. E.B. Huerta, B. Duval, J.K. Hao, A hybrid GA/SVM approach for gene selection and classification of microarray data, 2006, Applications of Evolutionary Computing, vol. 3907, no. 1, pp. 34-44.
8. J. Zhou, T. Bai, C. Suo, The SVM optimized by culture genetic algorithm and its application in forecasting share price, 2008, IEEE International Conference on Granular Computing, Art. No. 4664698, pp. 838–843
9. J. Zhou, T. Bai, A. Zhang, J. Tian, The integrated methodology of wavelet transform and GA based-SVM for forecasting share price, 2016, IEEE International Conference on Granular Computing, Art. No. 4608094, pp. 729–733
10. F. Kuang, S. Zhang, Z. Jin, W. Xu, A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection, 2015, Soft Computing, vol. 19, no. 5, pp. 1-13.
11. Z. Xu, Z.Y. Dong, W.Q. Liu, Short-term electricity price forecasting using wavelet and SVM techniques, 2013, Dynamics of Continuous Discrete & Impulsive Systems, vol. 16, no. 1, pp. 372-377.
12. S.K. Aggarwal, L.M. Saini, A. Kumar, Electricity price forecasting in deregulated markets: a review and evaluation, International Journal of Electrical Power & Energy Systems, vol. 31, no. 1, pp. 13–22
13. T. Mu, A.K. Nandi, Automatic tuning of L2-SVM parameters employing the extended Kalman filter, 2009, Expert Systems, vol. 26, no. 2, pp. 160- 175.
14. H.L. Wang, D.X. Xu, Analysis of Local Capability Characteristics of Kernel Function, 2017, Journal of Residuals Science and Technology, vol. 13, no. 7, pp. 110-116.
15. X. Huang, The study on Kernel in Support Vector Machine, 2008, Soochow University, China.
16. H. Yi, X. Song, B. Jiang, Y. Liu, Z. Zhou, Flexible support vector regression and its application to fault detection, 2013, Acta Automatica Sinica, vol. 39, no. 3, pp. 272-284
17. O. Chapelle, V. Vapnik, O. Bousquet, S. Mukherjee, Choosing multiple parameters for support vector machines, 2002, Machine Learning, vol. 46, no. 1, pp. 131-159.
18. T.H. Loutas, D. Roulias, G. Georgoulas, Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression, 2013, IEEE Transactions on Reliability, vol. 62, no. 4, pp. 821-832.
19. J.P. Jacobs, Bayesian support vector regression with automatic relevance determination kernel for modeling of antenna input characteristics, 2012, IEEE Transactions on Antennas and Propagation, vol. 60, no. 4, pp. 2114-2118.

20. N. Wang, H.Z. Ai, Combining local and global information for hair shape modeling, 2014, Acta Automatica Sinica, vol. 40, no. 4, pp. 615-623.