http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (14): 2673-2682, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.2673.2682
© 2013 Asian Network for Scientific Information

Research of Bessel Kernel Function of the First Kind for Support Vector Regression

Li Xiang, Zhu Quanyin and Wang Liuyang
Faculty of Computer Engineering, Huaiyin Institute of Technology, Huai'an 223003, China

Abstract: Kernel method is an effective method to solve the non-linear model analysis and also a research focus in the current pattern recognition community. The selection of kernel functions plays an important role in the performance of kernel methods. The Support Vector Regression (SVR) had provided higher performance than traditional learning ma-chines and had been widely applied in real-world regression problems and nonlinear function estimation problems. In view of the regression performance for SVR is affected largely by the selected kernel function, a model of SVR based on the Bessel kernel function of the first kind were put forward and given the implementations with R and LibSVM. 8 data sets in the database of UCI and 4 common kernel functions were selected for the experiment, the Mean Square Error (MSE) and determination coefficient R^2 were used as the performance evaluation index. The experimental results show that Bessel kernel function of first kind has higher prediction accuracy and a stronger generalization ability in SVR, which provides references for the kernel functions selection of SVR.

Key words: Support vector machine, regression, bessel kernel function, prediction, machine learning

INTRODUCTION

Support Vector Machine (SVM) has been introduced into the field of machine learning on the conference of Computational Learning Theory in 1992 by Jia *et al*. (2012). SVM is a technology based on statistical learning theory and integrates a number of technologies, such as maximum interval hyperplane, Mercer kernel, convex quadratic optimization, sparse solution (Granitzer *et al*., 2009; Wang and Chen, 2012).

In the prediction of a limited sample, SVR has developed a set of complete and normative machine learning theory and method based on statistics, which reduced the randomness of algorithm design greatly and overcame the deficiency of the bigger difference that may exist between empirical risk and expected risk in the traditional statistics (Xu and Jia, 2012). SVR, which has good generalization ability, strong nonlinear processing ability and won't fall into the local minimum value, has been widely used in the research of the regression problem (Hao and Chiang, 2008; Abo-Khalil and Lee, 2008; Bi *et al*., 2011). The regression performance of SVR is affected by many factors and here are two key factors: (1) The selection of the error penalty factor C and insensitive loss function parameter; (2) The form of kernel function and the selection of its parameters (Cheng and Wang, 2011). There are four kinds of SVM common kernel functions: linear kernel function, polynomial kernel function, RBF kernel function, sigmoid kernel function. In the standard non-linear support vector regression formulation, input vectors are projected into a high dimensional feature space by means of a set of basis functions; linear regression is then carried out in this space. A kernel function is used to compute inner products between projections of input vectors in feature space (Jacobs, 2012). Like other methods based on kernels, the quality of the regression depends on the choice of the kernel function and its parameters, which must be suitable to the current data. In general, this choice, also known as kernel selection, is a difficult task: The kernel is often chosen by trial and error, genetic optimization, or user expertise (Bellocchio *et al*., 2012).

Many scholars have made improvements in the selection of SVR kernel function and optimization of parameters in order to improve the ability of SVR to process the regression problems and have achieved some success (Juang and Hsieh, 2012; Lahiri and Ghanta, 2009; Chen and Chen, 2012). Bi *et al*. (2005) and Liu *et al*. (2005) put forward to improve SVR performance by optimizing RBF nuclear parameters but the process of searching the optimal parameters still needs a lot of time. Liu *et al*. (2009) put forward to combine sigmoid kernel function with RBF kernel function, forming a new hybrid kernel function to improve the global regression performance of SVR but the method has a complex computation process, sigmoid kernel function is positive semiBdefinite only when the parameters satisfy the certain conditions and the regression result is unstable. This article has put forward a SVR model based on Bessel kernel function of the first kind and gives the implementations in R and verified the

Corresponding Author: Li Xiang, Faculty of Computer Engineering, Huaiyin Institute of Technology, Huai'an 223003, China

effectiveness of this method through the simulation experiments of data sets in the UCI database, which has an implication for promoting the application of SVR.

PRINCIPLE OF SVR

Based on the statistical learning theory and the minimum of structure risk, support vector machine has selected the discriminant function properly to minimize the real risk of learning, it defined the support vector space, realized the nonlinear mapping from the sample space into high-dimensional feature space through the kernel function and described the nonlinear dependency relationship between factor and objects by using support vector. This method has advantages for nonlinear mapping under the condition of small samples and it can limit overfitting, particularly suitable for the data processing of small sample sets, so it is widely used in pattern recognition, regression analysis and other fields (Liu *et al*., 2012).

Assume the training set that contains l samples is $\{(x_i, y_i), i = 1, 2, \cdots, l\}$, where x_i ($x_i \in R^d$) is the input column vector of the first i training sample, $x_i = [x^1_i, x^2_i, \cdots, x^d_i]^T$, $y_i \in R$ is the corresponding output value. Set the linear regression function that is established in high-dimensional feature space to:

$$f(x) = w\phi(x) + b \tag{1}$$

where, $\phi(x)$ is nonlinear mapping function:

$$L(f(x), y, \varepsilon) = \begin{cases} 0, & |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon, & |y - f(x)| \le \varepsilon \end{cases} \tag{2}$$

where, f (x) in Eq. 2 is the predicted value of the regression function, y is the corresponding real value, ε is the parameter of linear insensitive loss function. Introducing slack variables ξ_i, ξ_i^* and describe the above problems for searching w, b in mathematical language, that is:

$$\begin{cases} \min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ \text{s.t.} \begin{cases} y_i - w\phi(x_i) - b \le \varepsilon + \xi_i, \quad i = 1, 2, \cdots, l \\ -y_i + w\phi(x_i) + b \le \varepsilon + \xi_i^* \\ \xi_i \ge 0, \xi_i^* \ge 0 \end{cases} \end{cases} \tag{3}$$

C in Eq. 3 is the error penalty factor, the greater C the greater the penalty of samples with the training error more than ε, ε specified the error requirement of regression function, the smaller ε is the smaller the error of regression function is. In Eq. 3, Lagrange function is introduced and the dual conversion is done, finally the regression

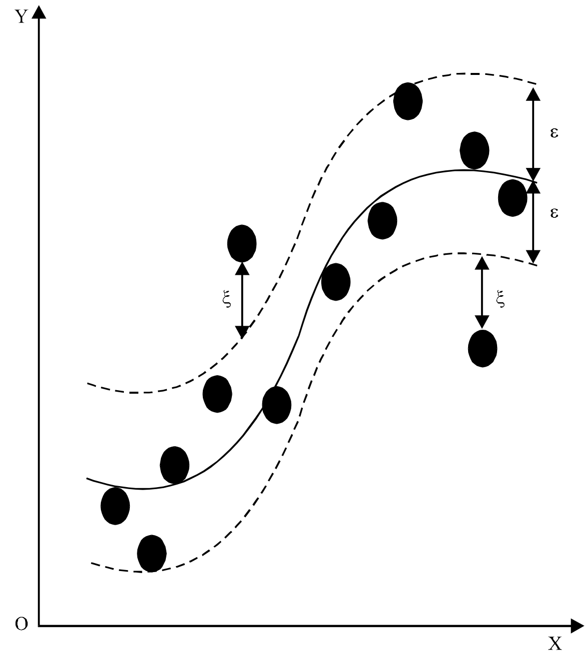

Fig. 1: Basic principle diagram of SVR

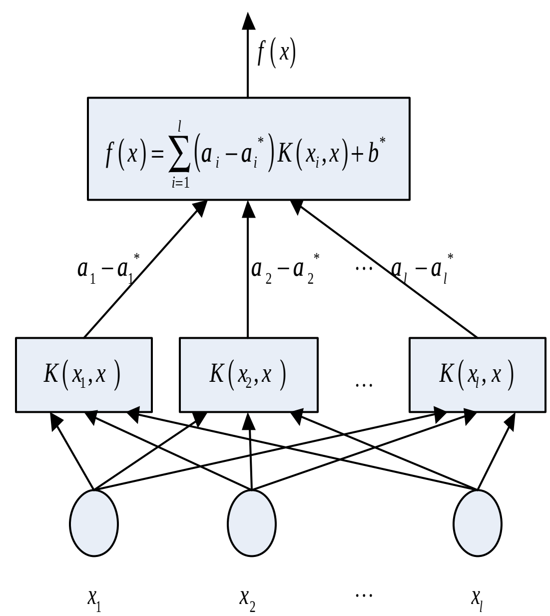

Fig. 2: Structure of SVR

function of SVR is as shown in Eq. 4. Figure 1 is the basic principle of SVR and Fig. 2 is the structural chart of SVR:

$$f(x) = \sum_{i=1}^{l}(a_i - a_i^*)K(x_i, x) + b^* \tag{4}$$

In Eq. 4:

$$0\ a=[a_1,a_2,\cdots,a_l],\ a^*=\left[a_1^*,a_2^*,\cdots,a_l^*\right],w^*=\sum_{i=1}^{l}\left(a_i-a_i^*\right)\phi(x_i)$$

$$b^*=\frac{1}{N_{nsv}}\left\{\sum_{0<a_i<C}\left[y_i-\sum_{x_i\in SV}\left(a_i-a_i^*\right)K(x_i,x_j)-\varepsilon\right]+\sum_{0<a_i<C}\left[y_i-\sum_{x_i\in SV}\left(a_j-a_j^*\right)K(x_i,x_j)+\varepsilon\right]\right\}$$

$$K(x_i,x)=\phi(x_i)\phi(x)$$

BESSEL KERNEL FUNCTION OF THE FIRST KIND

Introduction of bessel function: When solving definite problems in the circular area or cylindrical area, the second order linear ordinary differential equation will appear in the form of Eq. 5:

$$x^2\frac{d^2y}{dx^2}+x\frac{dy}{dx}+\left(x^2-n^2\right)y=0 \tag{5}$$

In Eq. 5, n is constant and the equation of Eq. 9 is called n order Bessel equation. The function y (x) as the standard solution in the equation of Eq. 5 is called Bessel function. The specific form of Bessel function will change as n in the equation changes and the corresponding solution is called n order Bessel function.

Introduction of Bessel Function of the First Kind: When n in Eq. 5 is integer and it is finite when x = 0, $J_n(x)$ as the solution of Eq. 5 is called n order Bessel function of the first kind:

$$J_n(x)=\sum_{m=0}^{\infty}\frac{(-1)^m}{m!\Gamma(m+n+1)}\left(\frac{x}{2}\right)^{n+2m} \tag{6}$$

Where:

$$\Gamma(x)=\int_0^{+\infty}e^{-t}t^{x-1}dt$$

When x>0, integration convergence will be got.

The shape of Bessel function of the first kind is roughly similar to sine or cosine function that decays according to the rate of :

$$\frac{1}{\sqrt{x}}$$

but zeros are not periodic, as x increases, the interval of zeros will be closer to the periodicity. The curve is about $J_n(x)$ of Zero-order, First-order and Second-order Bessel function of the first kind, as shown in Fig. 3.

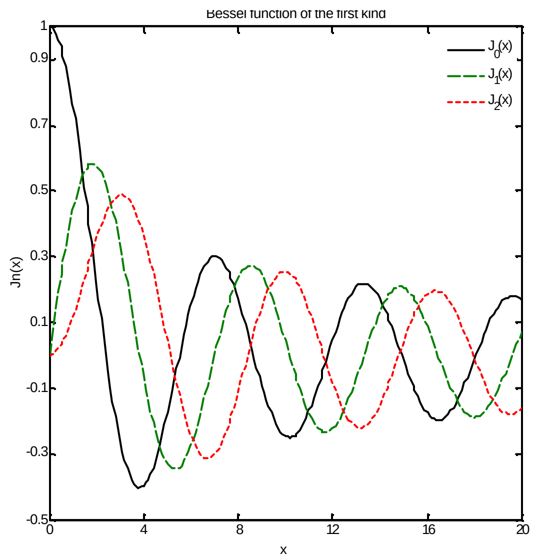

Fig. 3: Zero-order, first-order and second-order bessel function of the first kind

Instruction of R and Implementation of Bessel Kernel Function of the First Kind in R: R is an open statistical programming tool and is also a mathematical computing environment. R has provided a flexible, interactive environment for analyzing and processing data and meanwhile R provided a number of statistical program packages, as well as some integrated statistical tools and various functions of mathematical calculation and statistical calculation.

LibSVM is a simple, easy-to-use and fast effective package of SVM pattern recognition and regression that professor Chih-Jen Lin from Taiwan University and etc developed and designed. This package can run in the environment of Java, Matlab, R, Python, Ruby,. net and etc.

Interface program package number of R for LibSVM is e1071, this package can solve C-SVC (C-support vector classication), nu-SVC (nu-support vector classication), one-class SVM (distribution estimation), epsilion-SVR (epsilon-support vector regression), nu-SVR (nu-support vector regression)l.

In order to use Bessel Kernel function of the first kind in R and LibSVM, we also need to install Kernel-based Machine Learning Lab package, which contains Hyperbolic tangent kernel (4Laplacian kernel (4Bessel kernel (4ANOVA RBF kernel (4Spline kernel (4String kernel and other kernel functions.

The help document of R gives the Bessel kernel function of the first kind to implement the function in R:

$$K(x,x')=-\mathrm{Bessel}_{(nu+1)}^{n}\left(\sigma|x-x'|^2\right) \tag{7}$$

Table 1: Number of training sample and test sample

Data sets	Feature dimension	Training samples	Test samples
Computer hardware	8	169	40
Concrete compressive strength	9	800	230
AutoMPG	8	300	98
Concrete slump test	10	80	23
Housing	14	450	56
Red wine quality	12	4000	898
White wine quality	12	4000	898
Year prediction MSD	90	400000	115345

EXPERIMENT AND RESULT ANALYSIS

Experimental data: In order to contrast the performance of Bessle kernel function of the first kind and other common nuclear functions in the regression of SVR. This experiment has selected eight data sets of Computer Hardware, Concrete Compressive Strength and etc in UCI database for comparison, which covered the situation of small sample and large sample, more and less categories, more and less feature dimension, uniform and non-uniform sample distribution, etc. The number of training sample and test sample that each data set randomly selected is as shown in Table 1.

Experiment and result analysis: The simulation experiment has been conducted by using R and LibSVM, training samples and test samples have been selected randomly in the data set according to the number as shown in table and SVR regression experiment has been done for different kernel functions. The cross validation method and the optimum parameter training model have been used to search the optimal error penalty factor C and kernel function parameter g and the algorithm process is as shown in Fig. 4. The Mean Square Error (MSE) that the function svmpredict returned and determination coefficient R^2 have been used to evaluate the performance of SVR regression prediction model. The closer MSE value gets to 0, the better and the closer R^2 value gets to 1, the better:

$$MSE = \frac{1}{l}\sum_{i=1}^{l}(\hat{y}_i - y_i)^2 \tag{8}$$

$$R^2 = \frac{\left(l\sum_{i=1}^{l}\hat{y}_i y_i - \sum_{i=1}^{l}\hat{y}_i \sum_{i=1}^{l} y_i\right)^2}{\left(l\sum_{i=1}^{l}\hat{y}_i^2 - \left(\sum_{i=1}^{l}\hat{y}_i\right)^2\right)\left(l\sum_{i=1}^{l} y_i^2 - \left(\sum_{i=1}^{l} y_i\right)^2\right)} \tag{9}$$

The linear kernel function prediction for ComputerHardware training set is shown in Fig. 5 and the linear kernel function prediction for test set is shown in Fig. 6. The polynomial kernel function prediction for ComputerHardware training set is shown in Fig. 7 and the polynomial kernel function prediction for test set is shown in Fig. 8. The RBF kernel function prediction for ComputerHardware training set is shown in Fig. 9 and the RBF kernel function prediction for test set is shown in Fig. 10. The sigmoid kernel function prediction for ComputerHardware training set is shown in Fig. 11 and the sigmoid kernel function prediction for test set is shown in Fig. 12. The Bessel kernel function of the first kind prediction for ComputerHardware training set is shown in Fig. 13 and the Bessel kernel function of the first kind prediction for test set is shown in Fig. 14. SVR

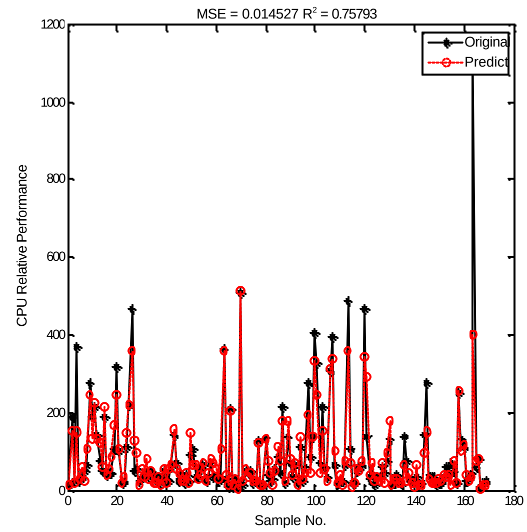

Fig. 4: Flow diagram of SVR model

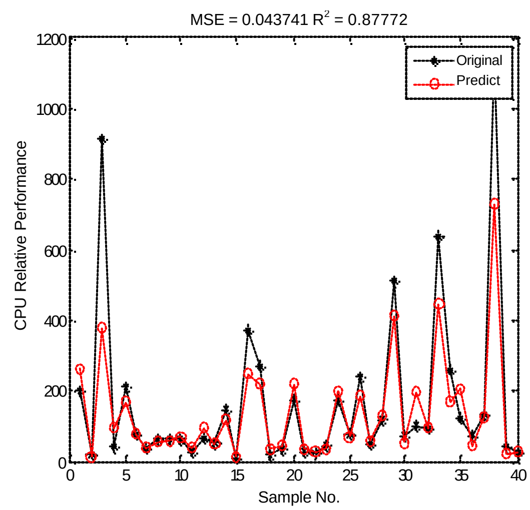

Fig. 5: Prediction of linear kernel function for computerhardware training set

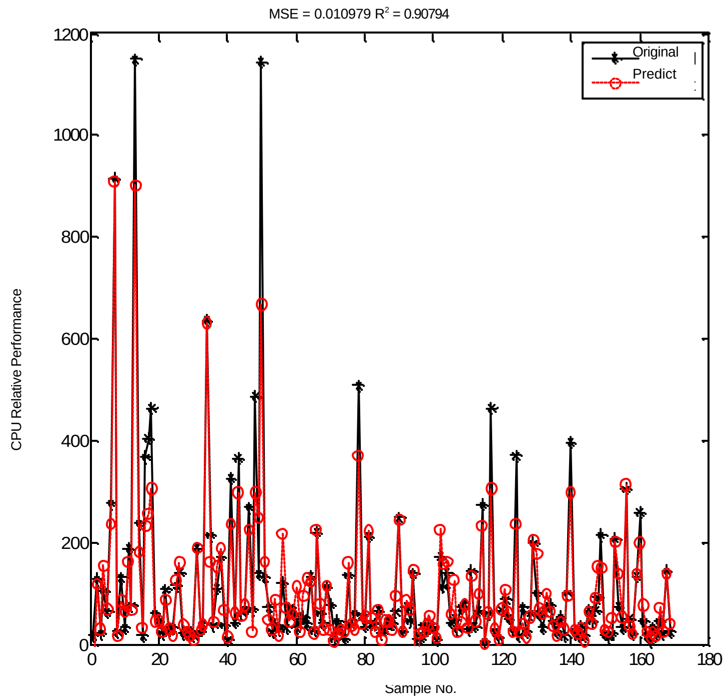

Fig. 6: Prediction of linear kernel function for computerhardware test set

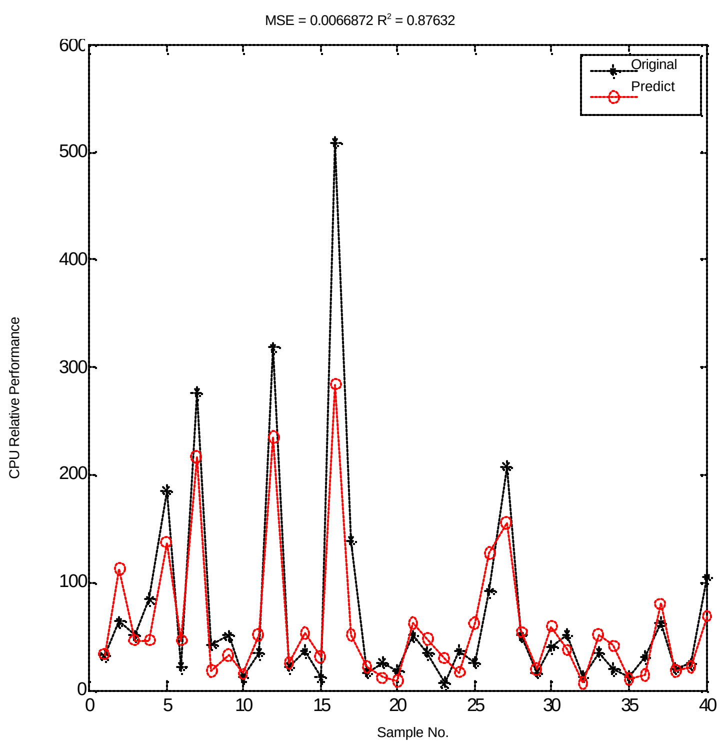

Fig. 7: Prediction of polynomial kernel function for computerhardware training set

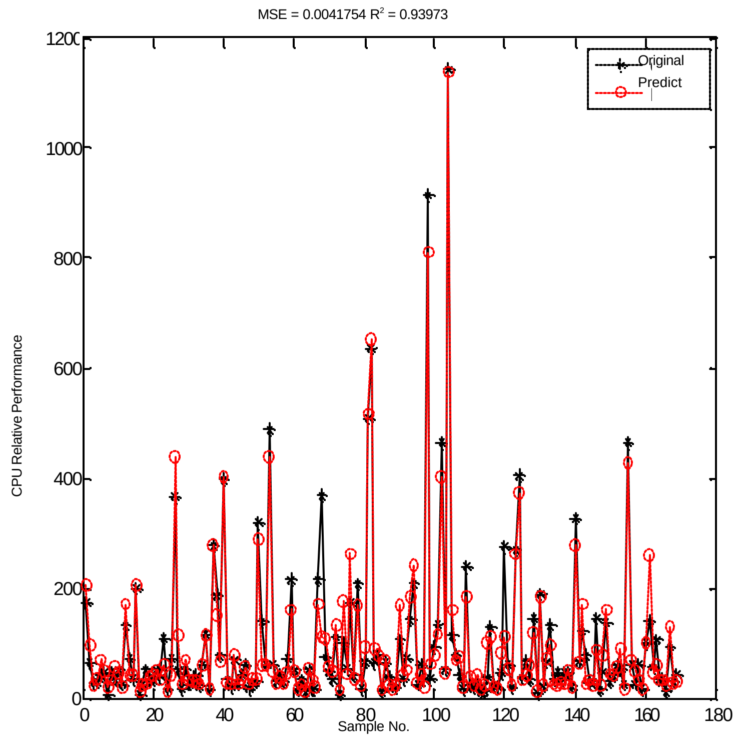

Fig. 8: Prediction of polynomial kernel function for computerhardware test set

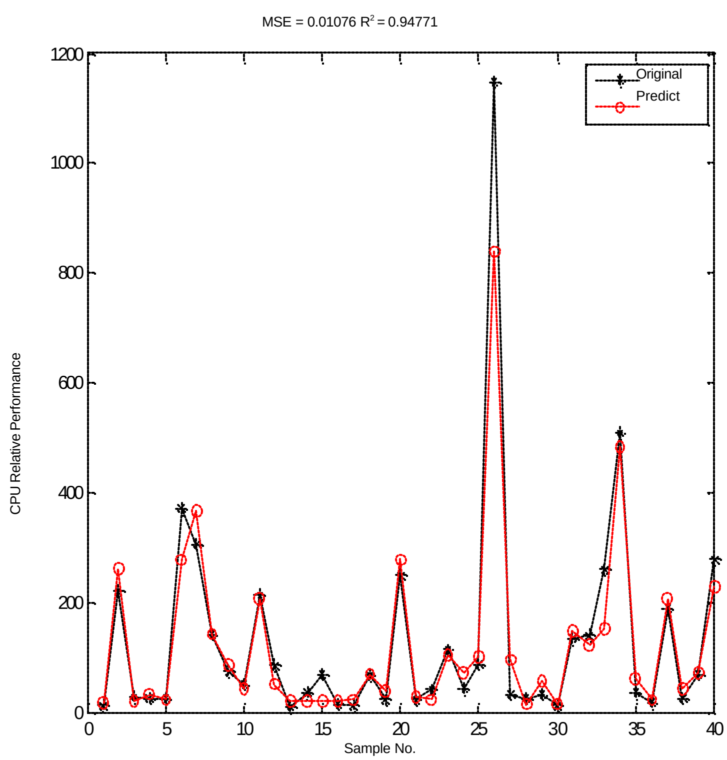

Fig. 9: Prediction of RBF kernel function for computerhardware training set

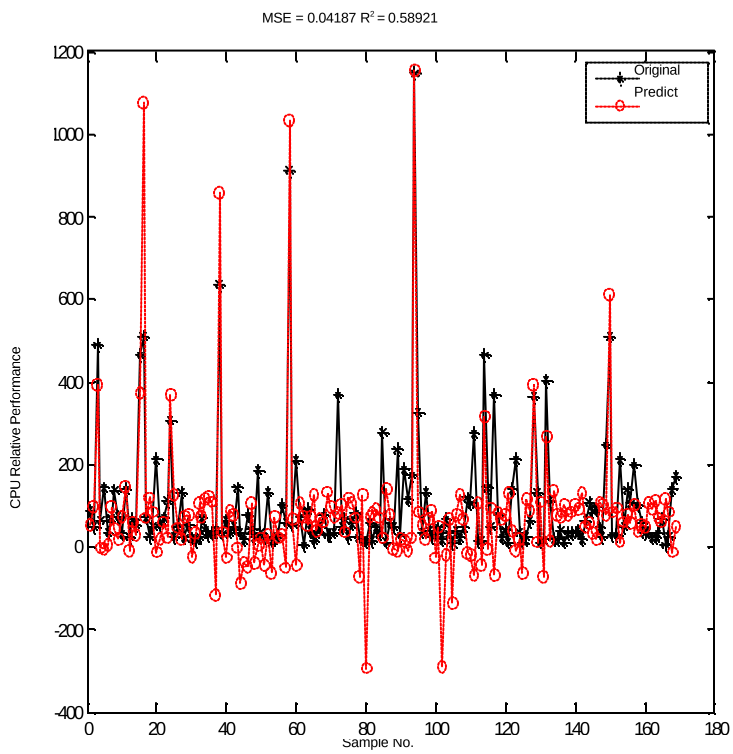

Fig. 10: Prediction of RBF kernel function for computerhardware test set

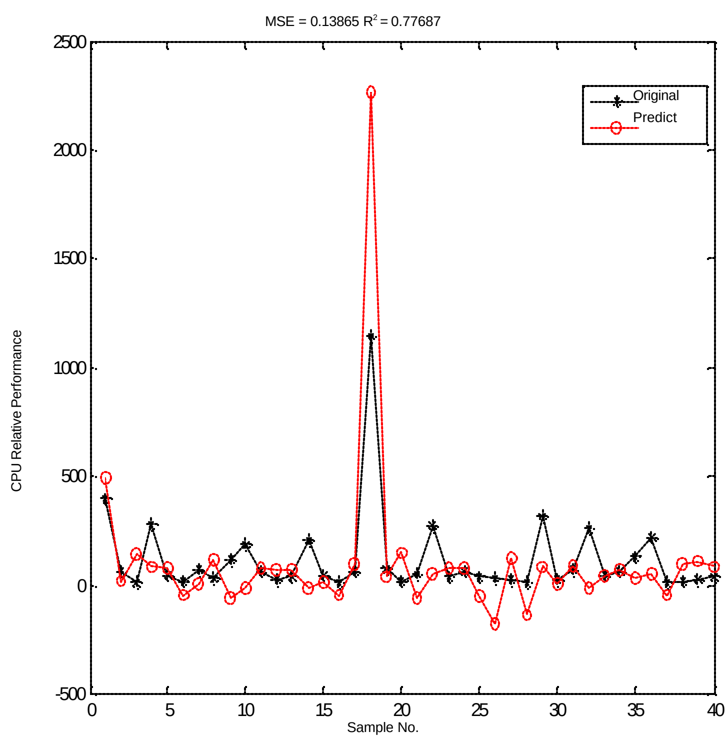

Fig. 11: Prediction of sigmoid kernel function for computerhardware training set

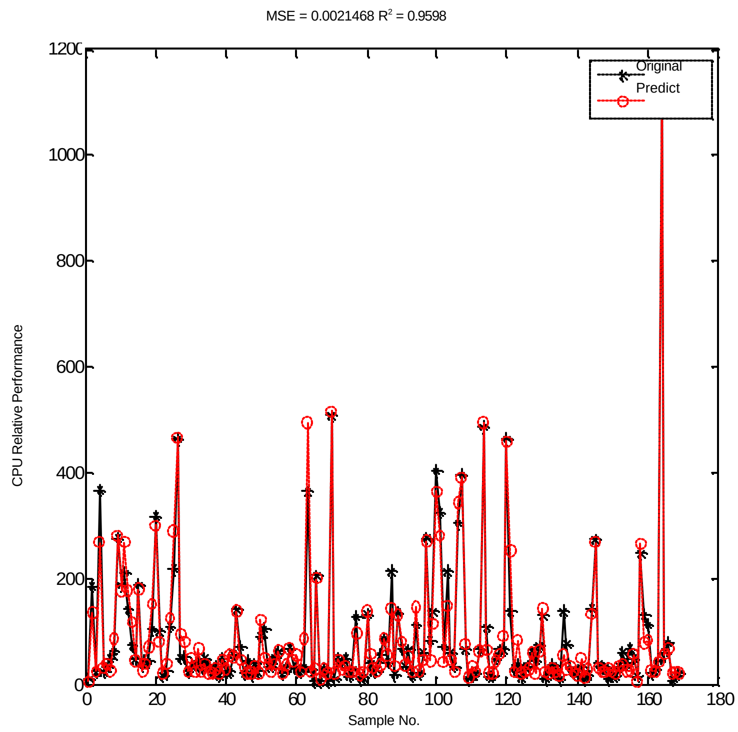

Fig.12: Prediction of sigmoid kernel function for computerhardware test set

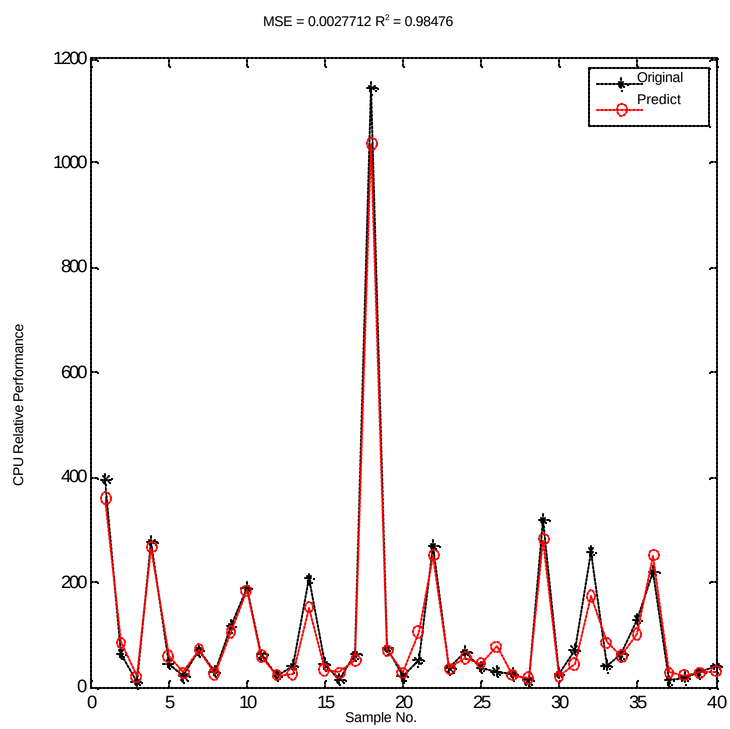

Fig.13: Prediction of bessel kernel function of the first kind for computerhardware training set

Table 2: SVR performance of different kernel functions for computer hardware data set

Data set			Linear kernel function	Polynomial kernel function	RBF kernel function	Sigmoid kernel function	Bessel kernel function of the first kind
Computer hardware	Training set	MSE	0.0145	0.0109	0.0041	0.0418	0.0021
		R^2	0.7579	0.9079	0.9397	0.5892	0.9598
	TEST SET	MSE	0.0437	0.0066	0.0107	0.1386	0.0027
		R^2	0.8777	0.8763	0.9477	0.7768	0.9847

Table 3: SVR Performance of Different Kernel Functions for Other 7 Data Sets in UCI

Data set			Linear kernel function	Polynomial kernel function	RBF kernel function	Sigmoid kernel function	Bessel kernel function of the first kind
Concrete compressive strengt	Training set	MSE	0.0686	0.0572	0.0078	0.0688	0.0037
		R^2	0.6011	0.7212	0.9540	0.6000	0.9639
	Test set	MSE	0.0742	0.0653	0.0112	0.0744	0.0095
		R^2	0.7581	0.7590	0.9519	0.7577	0.9842
Auto MPG	Training set	MSE	0.0338	0.0215	0.0133	0.0322	0.0116
		R^2	0.8044	0.8532	0.9219	0.8120	0.9308
	Test set	MSE	0.0271	0.0631	0.0151	0.0467	0.0103
		R^2	0.8653	0.8701	0.9213	0.8366	0.9326
Concrete slumptest	Training set	MSE	0.0162	0.0033	0.0005	0.0143	0.0012
		R^2	0.8950	0.9779	0.9959	0.9015	0.9867
	Test set	MSE	0.0103	0.0144	0.0047	0.0171	0.0056
		R^2	0.9245	0.9079	0.9810	0.8660	0.9738
Housing	Training set	MSE	0.0574	0.0627	0.0061	0.0544	0.0034
		R^2	0.6720	0.8675	0.9629	0.6753	0.9868
	Test set	MSE	0.0561	0.0639	0.0111	0.1438	0.0085
		R^2	0.5961	0.8528	0.9446	0.4717	0.9571
Red wine quality	Training set	MSE	0.0717	0.0615	0.0170	0.0719	0.0103
		R^2	0.5510	0.6591	0.8686	0.5500	0.8912
	Test set	MSE	0.0768	0.0688	0.0201	0.0770	0.0053
		R^2	0.6923	0.7131	0.8667	0.6919	0.8967
White wine quality	Training set	MSE	0.0331	0.0298	0.0237	0.0576	0.0301
		R^2	0.8321	0.8371	0.9657	0.8503	0.8729
	Test set	MSE	0.0593	0.0259	0.0296	0.1447	0.0345
		R^2	0.8099	0.8287	0.9229	0.8191	0.8536
Year prediction MSD	Training set	MSE	0.0514	0.0400	0.0324	0.0499	0.0298
		R^2	0.7681	0.8135	0.9374	0.7752	0.9715
	Test set	MSE	0.0452	0.0787	0.0340	0.0634	0.0247
		R^2	0.7247	0.7827	0.9168	0.7680	0.9362

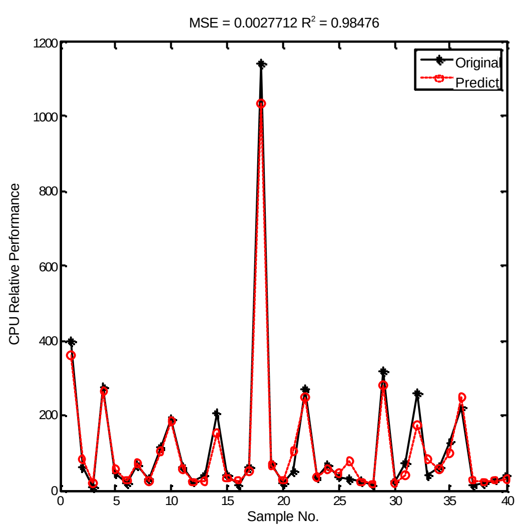

Fig. 14: Prediction of bessel kernel function of the first kind for computerhardware test set

regression performance of different kernel functions of ComputerHardware data set is shown in Table 2.

Figure 5-14 and Table 3, the corresponding MSE value of Bessel nuclear function of the first kind in ComputerHardware training set and test set, gets closest to 0 and R^2 gets closest to Eq. 1. Thus it can be seen the corresponding SVR model of Bessel kernel function of the first kind in ComputerHardware data set has the highest prediction accuracy and has the best generalization ability.

Similarly, SVR regression performance of each kernel function in other seven data sets is shown in Table 3.

Table 4, the corresponding MSE value of Bessel nuclear function of the first kind in Concrete Compressive Strengt, AutoMPG, Housing, RedWineQuality, Year PredictionMSD gets closest to 0, R^2 gets closest to 1 and the regression performance is the best; the corresponding MSE and R^2 in Concrete SlumpTest and WhiteWineQuality are second only to RBF kernel function. Thus it can be seen, SVR model that is built by Bessel kernel function of the first kind has relatively high prediction accuracy, a better generalization ability and a stronger robustness. In the use of SVR, Bessel kernel function of the first kind can be selected as a priority.

CONCLUSION

SVR has been widely used in the field of pattern recognition and it has achieved good application effect but its prediction effect is affected greatly by the selection of kernel function. Based on the research of Bessel function of the first kind, the paper has put forward SVR model based on Bessel kernel function of the first kind and given the method of using Bessel kernel function of the first kind in R and LibSVM, it is found Bessel kernel function of the first kind has a better regression prediction performance through the regression and comparative research of eight data sets in UCI database, to provide reference for the application of SVR.

ACKNOWLEDGMENT

This study was supported by the National Sparking Plan Project of China (2011GA690190) and the fund of Huai'an Science and Technology Projects, China (HASZ2013008).

REFERENCES

Abo-Khalil, A.G. and D.C. Lee, 2008. MPPT control of wind generation systems based on estimated wind speed using SVR. IEEE Trans. Ind. Electron., 55: 1489-1490.

Bellocchio, F., S. Ferrari, V. Piuri and N.A. Borghese, 2012. Hierarchical approach for multiscale support vector regression. IEEE Trans. Neural Networks Learn. Syst., 23: 1448-1460.

Bi, L., O. Tsimhoni and Y. Liu, 2011. Using the support vector regression approach to model human performance. IEEE Trans. Syst., 41: 410-417.

Bi, L.P., H. Huang, Z.Y. Zheng and H.T. Song, 2005. New heuristic for determination gaussian kernels parameter. Proceedings of the International Conference on Machine Learning and Cybernetics, Vol. 7, August 18-21, 2005, Guangzhou, China, pp: 4299-4304.

Chen, Y. and Q. Chen, 2012. Interval analysis using least squares support vector fuzzy regression. J. Control Theor. Appl., 10: 458-464.

Cheng, P. and X.L. Wang, 2011. Influence of SVR parameter on non-linear function approximation. Comput. Eng., 37: 189-191, 194.

Granitzer, M., A.S. Rath, M. Kroll, C. Seifert and D. Ipsmiller *et al*., 2009. Machine learning based work task classification. J. Digital Inform. Manage., 7: 305-312.

Hao, P.Y. and J.H. Chiang, 2008. Fuzzy regression analysis by support vector learning approach. IEEE Trans. Fuzzy Syst., 16: 428-441.

Jacobs, J.P., 2012. Bayesian support vector regression with automatic relevance determination kernel for modeling of antenna input characteristics. IEEE Trans. Antennas Propaagation, 60: 2114-2118.

Jia, Y.F., A.L. Zhang and Y.Z. Wu, 2012. Improved global optimization algorithm based on incremental support vector regression model. Comput. Sci., 39: 185-188.

Juang, C.F. and C.D. Hsieh, 2012. A fuzzy system constructed by rule generation and iterative linear SVR for antecedent and consequent parameter optimization. IEEE Trans. Fuzzy Syst., 20: 372-384.

Lahiri, S.K. and K.C. Ghanta, 2009. Support vector regression with parameter tuning assisted by differential evolution technique: Study on pressure drop of slurry flow in pipeline. Korean J. Chem. Eng., 26: 1175-1185.

Liu, H.J., Y.N. Wang and X.F. Lu, 2005. A method to choose kernel function and its parameters for support vector machines. Proceedings of the International Conference on Machine Learning and Cybernetics, Vol. 7, August 18-21, 2005, Guangzhou, China, pp: 4277-4280.

Liu, M., S.S. Zhou and H. Wu, 2009. SVM based on new mixed kernel function. J. Comput. Appl., 29: 167-168.

Liu, Y.B., Q.B. He and P. Zhang, 2012. Rotating machinery fault diagnosis method based on SVM. Comput. Eng., 38: 233-235.

Wang, T.H. and J.T. Chen, 2012. Survey of research on kernel selection. Comput. Eng. Design, 33: 1181-1186.

Xu, G.G. and Y. Jia, 2012. Pattern Recognition and Intelligent Computing in MATLAB. Beihang University Press, China, pp: 125-126.