A New Compact Support Kernel of Support Vector Machines

Enqing Dong Xiao Huang
School of Information Engineering, Shandong University at Weihai, 264209 China

Abstract- **A new kernel based on the KCS (Kernels with Compact Support) kernel family is proposed. The training time of the support vector machines of the new kernel is shorter than that of the common kernels. The support vector rate of the classifying machines also keeps in a low level. Compared with other classifying kernels, a great deal of experimental results indicates that the new kernel has a better classifying performance in VAD (Voice Active Detection).**

Keywords: **support vector machines; Voice Active Detection; kernel functions**

I. INTRODUCTION

The basic idea of Support Vector Machine (SVM) algorithm is to map input data into some other dot product space (called the feature space) via a nonlinear map. So it can be shown to correspond to a linear method in a high-dimensional feature space nonlinearly related to input space [1-2]. By the use of kernels $K(x_i \cdot x)$, all necessary computations are performed directly in input space[3]. Suppose $(x_i \cdot x)$, $i = 1,\cdots,n$, $x \in R^d$, $x \in \{+1,-1\}$ is classifying sign. By solving a quadratic function extremum problem with restrained inequality, an optimum classification function can be written as follows,

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{n} a_i^* y_i K(x_i \cdot x) + b^*) \qquad (1)$$

In SVM model, the kernel function is a key factor. Different kernels will directly affect the classification performance of SVM. At first, for understanding kernel function classification, we discuss the normal kernel functions[3], and then research the characteristics of mixture kernel function[4] based on Smits et al. research[5]. A kernel function（Kernels with Compact Support, KCS[6]）proposed by Lakhdar Remaki and Mohamed Cheriet et al. is based on the advantage of the RBF kernel, but the complexity of the kernel affects its performance. So we present a new kind of the KCS kernel function to solve the problem. Voice active detection experiments in different SNR and noise environment show the new corrected KCS kernel (called as CKCS in the following) has a better performance than the RBF and the mixture kernels (MIX kernels).

II. COMMON KERNELS

Kernel function is the key factor of SVM. A common kernel function can be defined as a function $K(x_i \cdot x_j) = \phi(x_i) \cdot \phi(x_j)$, here $\phi(\cdot)$ is a non-linear transform function. Since then, SVM becomes more practical. At present, the common-used kernel functions are as follows:

(a) Poly: $K(x_i \cdot x) = ((x_i \cdot x) + 1)^q$ (2)

(b) RBF: $K(x_i \cdot x) = \exp(- \| x_i - x \|^2 / \sigma^2)$ (3)

(c) Sigmoid): $K(x_i \cdot x) = \tanh(v(x_i \cdot x) + c)$ (4)

(d) Linear: $K(x_i \cdot x) = x_i \cdot x$ (5)

in formula (2) - (5) q, σ, c are real number. In practical applications, the selection of kernel function and correspond parameters is dependent on the real requirements.

III. MIXTURE KERNELS

The types of SVM kernels play a crucial role in SVM classification. The kernels must satisfy the Mecer condition. There are many kinds of kernels, such as text kernel, structure kernel and so on[3].It is not easy to explain their characteristics. However, as a whole, there are two kinds of kernels, local kernels and global kernels. Because the local kernel function has a good learning and bad generalized ability, whereas the global kernel function has a reversed fact, so a MIX kernel function is constructed as follows,

$$K_{mix} = aK_{local} + bK_{global} \qquad (6)$$

Poly is a typical global kernel function, such as formula (2). In Fig.1(a), when q is selected from 1, 2, 3, 4 respectively, and the test point is 0.5, the figure shows that global kernel function not only have an influence to data near the test point, but also has influence to data away from the test point. On the other hand, the RBF kernel function is a typical local kernel function, such as formula (3). In Fig.1(b), when σ is selected from 0.1, 0.2, 0.3, 0.4 respectively, and the test point is 0.5 too, the figure shows that local kernel function only has an influence to data in a small range of the test point.

In Fig.1(c), the kernel function has an influence to both of data near and far from the test point. The result shows the MIX kernel function has local and global characteristics at the same time. In addition, for selecting the MIX kernel parameters, the kernel will get better performance when the parameter a is close to 1(a+b=1). Formula (6) is a MIX kernel function expression which satisfies the Mecer condition.

IV. CORRECTED KERNEL WITH COMPACT SUPPORT

Except constructing kernel function, using MIX kernel function, there are other methods such as the Gaussian recursive filter [7]. But they can only solve a part of the problems, so Lakhdar Remaki et al. proposed KCS kernel function to overcome the limitation of the RBF kernel.

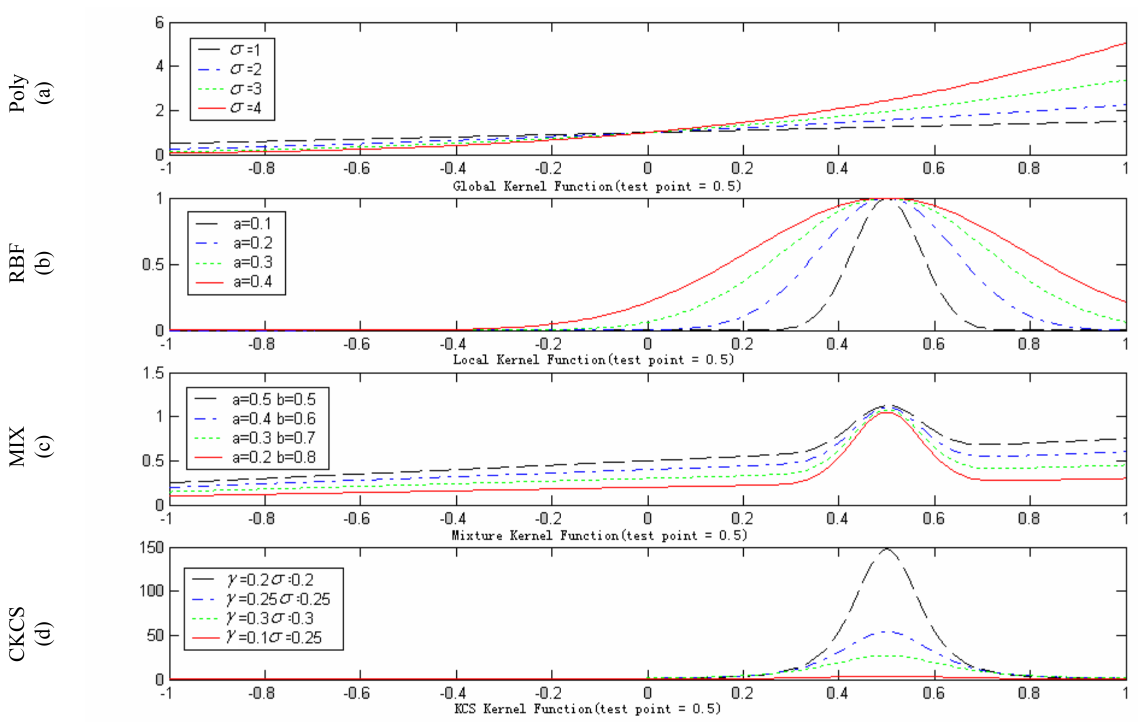

Figure 1. Compared Graph of Kernel Functions

The RBF kernel has two evident limitations: ① Information loss due to kernel truncation. ② Increasing processing time due to the long side.

In addition, the RBF kernel also has four important properties as follows:

(1) Recovering the initial signal when the scale parameter tends toward zero;

(2) Continuity with respect to the scale parameter;

(3) Stronger regularization property;

(4) Zero-crossing diminishing property.

The KCS is proposed by Lakhdar Remaki based on the RBF kernel, which retains properties (1) to (3). In addition, in fact, the property (4) is conserved overall. The KCS is as follows

$$K_{\sigma,\gamma}(x,y)=\begin{cases}\dfrac{1}{C_\gamma\sigma^2}e^{\left(\gamma\sigma^2/\left(x^2+y^2-\sigma^2\right)+\gamma\right)}, & x^2+y^2<\sigma^2\\ 0 & ,\text{elsewhere}\end{cases} \quad (7)$$

the parameter σ controls the bandwidth of the KCS, and C_γ is calculated by formula (8).

$$C_\gamma=\Delta x\Delta y e^{\gamma/2}\sum_{j=0}^{M}\sum_{i=0}^{N}e^{\gamma/2(x_i+y_i)}1_{B(0,1)}(x_i,y_i) \quad (8)$$

In the research of the KCS, we make a correction about the kernel because the complexity of the KCS affects classifying speed and increasing classifying error. By generalizing the formula (7) we propose a corrected KCS (CKCS) as follows:

$$K'(x_i\cdot x)=G\cdot\exp(\gamma/(\|x_i-x\|^2+\sigma^2)-\lambda) \quad (9)$$

Experiments and theory show that the KCS not only overcomes the shortcomings of the RBF kernel, but also keeps the four advantages of the RBF kernel[6]. Fig.1(d) is the curve of the corrected KCS kernel(G=1, λ=1).

V. Performance Evaluation of the Corrected KCS Kernel in VAD

In order to assess the performance of the corrected new kernel in SVM classifying detection, we adopt a MIX kernel and other common kernels respectively to test using the IEEE Sentence Database[8] voice database. Speech data with car noise and babble noise are used in Voice Activity Detection(VAD)[9]respectively. The four characteristic parameters used by G.729B's VAD algorithm are used for training support vector machines in the experiments. The SNR of the experimental speech is 0dB, 5dB, 10dB, 15dB, 20dB respectively. The compared results show in Fig. 2 and Fig. 3.

In the Fig. 2 and Fig. 3, the MIX kernel and the CKCS kernel have a better performance than other common kernels. In particular, the CKCS kernel does much better than the traditional RBF kernel in generalization when the SNR of the speech is under 0. The property is very important for a classifier, which can satisfy the real requirement of the predictive and classifying ability. Moreover, the training time and support vector rate of the corrected CKCS kernel is near 1/20 and 1/10 of the RBF kernel in VAD respectively, and the MIX kernel is the second place. Therefore, the CKCS kernel can effectively overcome the disadvantage of the RBF kernel, and the MIX kernel can keep the most advantage of the global kernel and the local kernel. All of them increase the performance of the SVM classifier; there will be significant in real applications.

Furthermore, by comparing the training time in 15dB car noise environment with different parameters, Table 1. shows that the error rate of the traditional RBF kernel is lower than that of the Sigmoid kernel and the Poly kernel. However, the two disadvantages of the RBF kernel in the measure space cause a long training time and more support vector rate. Because the MIX kernel retains both of the characteristics of the RBF kernel and the Poly kernel, so the MIX kernel has a shorter training time and lower support vector rate related to the RBF kernel. On the basis of containing the characteristics of the RBF kernel, the corrected KCS kernel can effectively overcome the disadvantages of the RBF kernel so that there is a better performance in the training time, support vector rate and detection error rate.

We use K-fold cross validation to find the best parameters of the corrected KCS kernel[10-11]. The K-fold cross validation is a common way to calculate the generalized error. The training sample set is divided into k equal subsets in stochastic manner. After finding a judge function by training k-1 subsets, and then test the judge function using the remaining subset. The holdout method is repeated k times. In each time, one of the k subsets is used as the test set and the other k-1 subsets are put together to form a training set. Then the average error of all k trials is calculated as the generalized error. Table 2. shows that the corrected KCS kernel can have a best performance when $\gamma=1, \sigma=10$.

VI. CONCLUSIONS

A new corrected KCS kernel for Support Vector Machine based on the research of common kernels and Mix kernel is proposed. The training time of the support vector machines of the new kernel is shorter than that of the common kernels. The support vector rate of the classifying machines also keeps in a low level. Compared with other classifying kernels, the new kernel has a better classifying performance in VAD. In addition, the best parameter of corrected KCS kernel is given in the VAD experiment.

ACKNOWLEDGMENT

This work was supported in part by the national natural science foundation of China under Grant 60572074.

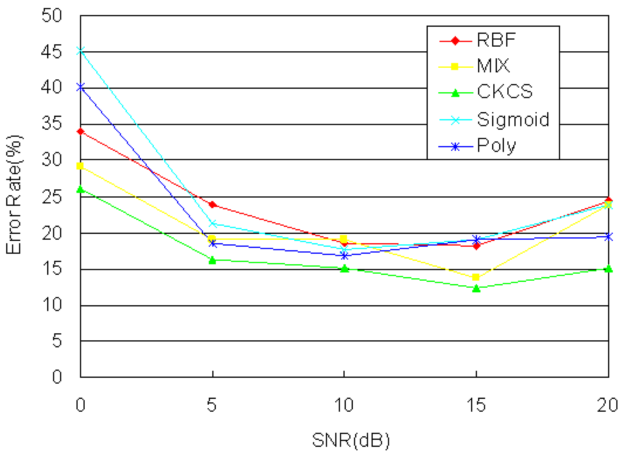

Figure 2. Performance Comparison in Car Noise

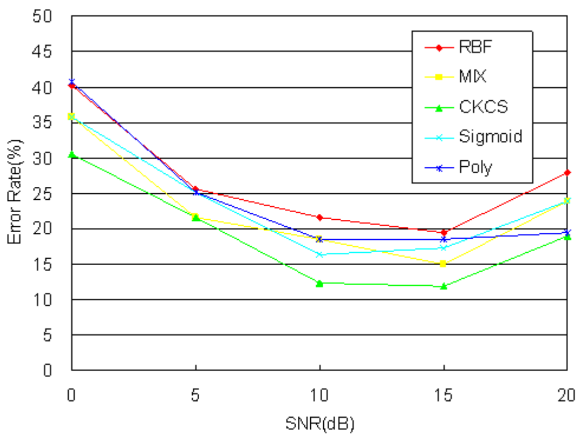

Figure 3. Performance Comparison in Babble Noise

Table 1. Performance Comparison in 15dB Car Noise

	Error(%)	SVs(%)	Time(s)
RBF	15.0442	30	434.6
Poly	19.0265	1.4	11
CKCS	11.5044	1.4	6.2
MIX	11.9496	5.2	203.4
Sigmoid	19.0265	2.3	15.4

Table 2. Parameter Selection Comparison in 15dB Car Noise

	Error(%)	SVs(%)	Time(s)
$\gamma=0.2, \sigma=0.2$	39.3782	43.4	24.9
$\gamma=0.1, \sigma=0.5$	17.6166	15	144.4
$\gamma=0.1, \sigma=0.9$	12.9534	6.4	46.4
$\gamma=3, \sigma=3$	11.3990	8.12	44
$\gamma=1, \sigma=10$	14.5078	0.7	5.3
$\gamma=5, \sigma=5$	18.1347	4.1	22.3

REFERENCES

[1] V. N. Vapnik. Statistical Learning Theory. New York: John Wiley & Sons, 1998.
[2] V. N. Vapnik. The Nature of Statistical Learning Theory, 2nd edition. New York: Springer-Verlag, 2000.
[3] Shawe-Taylor, J. Cristianini, N. KernelMethods for Pattern Analysis. Cambridge University Press, 2004.
[4] Zheng Sheng, Liu Jian, Tian Jin-Wen. "An SVM-based Small Target Segmentationand Clustering Approach," Proceding of the Third International Conference on Machine learning and Cybernetics. Shanghai: IEEE, 3318 - 3323, 2004.
[5] Smits G F, Jordaan E M., "Improved SVM Regression using Mixture of Kernels," Proceedings of the 2002 International Joint Conference on Neural Networks. Hawaii: IEEE, 2785 - 2790, 2002.
[6] Remaki, L. Cheriet, M. Member, "KCS-New Kernel Family with Compact Support in Scale Space: Formulation and Impact," IEEE Transaction on Image processing, vol.9 (6), 970 - 981, 2000.
[7] R. Horaud and O. Monga, Vision par Ordinateur, outils fondamentaux, Paris, France:Hermes, ,425 - 427, 1995.
[8] IEEE Subcommittee, IEEE Recommended Practice for Speech Quality Measurements. IEEE Trans. Audio and Electroacoustics, 225-246, 1969.
[9] Enqing, D. Guizhong, L. Yatong, Z. Xiaodi, Z., "Applying support vector machines to voice activity detection," 6th International Conference on Signal Processing. vol.2, 1124 - 1127, 2002.
[10] Chang, C. Lin, C J. "LIBSVM: a library for support vector machines," Dept. of Computer Science and Information Engineering, National Taiwan University 2001.
[11] Hsu, C W, Chan, C C, Lin, C J. , " Practical Guide to Support Vector Classification," Dept. of Computer Science and Information Engineering, National Taiwan University 2003.