

# 基于欧式距离的一种新的核函数

李 雷,任宇飞

(南京邮电大学 理学院,江苏 南京 210046)

**摘 要:**支持向量机表现的好坏很大程度上取决于核函数的选取,因此最近几年关于核函数的研究有许多。越来越多的核函数也被提了出来!但是选取合适的核函数往往却不容易,因为数据的特征往往不知道。文中利用函数的 Taylor 展开思想,提出了一种新的核函数,叫 T-KMOD,基于 KMOD 提出的。该核函数的灵活性更好,可以处理很多分类的问题。用网络入侵的数据对该核函数进行了仿真,从仿真的结果可以看出,和一些常用的核函数相比,它的鲁棒性更好,有更强的分类能力。同时该函数的分类效果更好。所以该核函数和一般常用的核函数相比,可能更具有一般选择性。

**关键词:**支持向量机;核函数;T-KMOD 核;分类

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2012)11-0157-04

## A New Kind of Kernel Function Based on Euclidean Distance

LI Lei, REN Yu-fei

(College of Science, Nanjing University of Posts and Telecommunications, Nanjing 210046, China)

**Abstract.** The behaviors of SVM largely depends on its adopted kernel function, so there are many researches on kernel function in recent years. More and more kernel functions are presented. It's very difficult to select a proper kernel function mentioned above, because the nature of the data is usually unknown. It presents a new kind of kernel function by the thought of function's Taylor expansion, called T-KMOD, based on KMOD. Its flexibility is better, so it may deal with lots of mapping problems. From the performance of T-KMOD investigated on network intrusion dates, obtain that it is robust and has stronger mapping ability comparing with commonly applied kernel functions. And it can obtain better generalization performance as well. So the proposed kernel function may be served as a generic alternative for some commonly applied kernel functions.

**Key words:** SVM; kernel function; T-KMOD kernel; classification

### 0 Introduction

**Definition 1**<sup>[1]</sup> A mapping  $\Phi: R^m \rightarrow H, \forall x \in R^m, x \rightarrow \Phi(x) (\Phi(x) \in H)$ , where  $m \geq 1, m \in Z$  and  $R^m, H$  is Euclidean space. A function  $k: R^m \times R^m \rightarrow R$ , with the characteristic  $\forall x_i, x_j \in R^m$ , s. t.  $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$ , Where  $\langle \cdot, \cdot \rangle$  is dot products in the space  $H$ , is called kernel function. Then call  $\Phi$  feature mapping and  $H$  feature space.

**Theorem 1**<sup>[2-4]</sup> if  $k_1, k_2 \dots$  are kernel functions in  $R^m \times R^m$ , and  $a \geq 0$  then

- (1)  $k = k_1 + k_2$ ; (2)  $k = ak_1$ ;
- (3)  $k = k_1 \times k_2$ ; (4)  $k = \lim_{n \rightarrow \infty} k_n$ ; (5)  $k = \exp(k_1)$ .

If  $k$  exists all the functions above are kernel functions.

An ideal SVM kernel function yields an inner product of given two vectors in a high dimensional vector space where all the input data can be linearly separated. Therefore, the inner product of any given pair of transformed vectors in the higher dimensional space can be found by applying the kernel function onto the input vectors directly without the need of an appropriate transformation function as  $\Phi(\cdot)$ . If kernel function is used by SVM, construct an optimization problem as follows<sup>[1]</sup>:

$$\begin{aligned} w(\alpha) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j k(x_i \cdot x_j) \\ \text{s. t. } \sum_{i=1}^N y_i \alpha_i &= 0 \\ \alpha_i &\geq 0, i = 1, \dots, N \end{aligned} \quad (1)$$

If  $\alpha^* = \{\alpha_1^*, \dots, \alpha_N^*\}$  are the optimal solution,  $w^* = \sum_{i=1}^N \alpha_i^* y_i \Phi(x_i)$ . And then decision function  $f(x) =$

收稿日期:2012-03-16;修回日期:2012-06-20

基金项目:国家自然科学基金(10371106,10471114);江苏省自然科学基金(04KJB110097)

作者简介:李 雷(1958-),男,博士,教授,研究方向为智能信号处理、非线性分析与计算智能;任宇飞(1986-),男,安徽合肥人,硕士研究生,研究方向为信息处理理论与应用。

$$\text{Sign} \left\{ \sum_{i=1}^N y_i \alpha_i^* k(x, x_i) + b^* \right\}.$$

## 1 The Presented Kernel Function -- T-KMOD

**Definition 2**<sup>[5-8]</sup>. A function  $f: (0, \infty) \rightarrow R$  is completely monotonic if it is  $C^\infty$  and, for all  $r > 0$  and  $k \geq 0$ ,  $(-1)^k f^{(k)}(r) \geq 0$ . Here  $f^{(k)}$  denotes the  $k$ th derivative of  $f$ .

**Theorem 2**<sup>[1,9-11]</sup>. Let  $X \subset R^n$ ,  $f: (0, \infty) \rightarrow R$  and  $K: X \times X \rightarrow R$  be defined by  $K(x, t) = f(\|x - t\|^2)$ . If  $f$  is completely monotonic, then  $K$  is positive definite.

For most commonly distance based kernels (e. g. RBF), points closing to each other are strongly correlated, whereas far apart points have uncorrelated images in the augmented space. The concern will focus on the image of the original points being linearly separable in the augmented space, therefore a kernel must turn very close points in the original space into weakly correlated elements (as weak as possible) while still maintaining the closeness information from vanishing. Then need the following couple of features: a quick decrease in the neighborhood of zero point and a moderate decrease toward infinity<sup>[4]</sup>. Based on this reason, Ayat, Cheriet, Remaki and Suen (2001) proposed an KMOD<sup>[2]</sup> kernel:

$$\text{KMOD}(x, y) = K \left[ \exp \frac{\gamma}{\|x - y\|^2 + \sigma^2} - 1 \right] \quad (2)$$

Where  $K$  is a normalization constant;  $\gamma$  and  $\sigma$  are two parameters controlling respectively the decreasing speed around zero and the width of the kernel.

In this section, present a kernel function, namely T-KMOD, Taylor Kernel with Moderate Decreasing. The T-KMOD can be seen as an approximate function of KMOD kernel function by using mathematical formula  $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \in R$ . This function is formulated as:

$$\text{T-KMOD}(x, y) = L \sum_{i=1}^n \frac{\left( \frac{\gamma}{\|x - y\|^2 + \sigma^2} \right)^i}{i!},$$

where  $i, n \in Z, L \geq 0$  (3)

where  $L$  is a normalization constant. The parameters  $n$  and  $\gamma, \sigma$  control decreasing speed around zero point and width, respectively. Prove T-KMOD  $(x, y)$  is a Mercer kernel.

**Corollary 1** Let  $\sigma \neq 0$ , T-KMOD  $(x, y)$  stated above is a Mercer kernel.

**Proof:** Clearly, for  $\sigma \neq 0$ , this function is continu-

ous, symmetric and  $C^\infty$ . Let  $f(r) = \frac{\gamma}{r + \sigma^2}$ ,  $\gamma = \|x - y\|^2$ , the T-KMOD  $(r) = \sum_{i=0}^n \frac{f^{(i)}(r)}{i!}$ . Because  $f^{(k)} = (-1)^k L (r + \sigma^2)^{-k-1}$ ,  $(-1)^k f^{(k)} > 0$ .  $f(r)$  is completely monotonic. Therefore,  $f(r)$  is positive definite by Theorem 2. So  $f(r)$  is a Mercer kernel. Then  $f^{(i)}(r), i \in Z$  is also a Mercer kernel by Theorem 1. Last, T-KMOD  $(x, y) = L \sum_{i=1}^n \frac{f^{(i)}(r)}{i!}$  is a Mercer kernel by Theorem 1 as well.

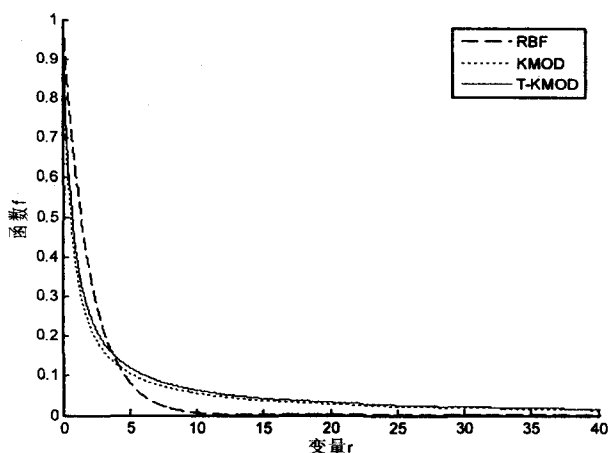


Fig 1 The trends near the origin point for KMOD, RBF and UKF kernels

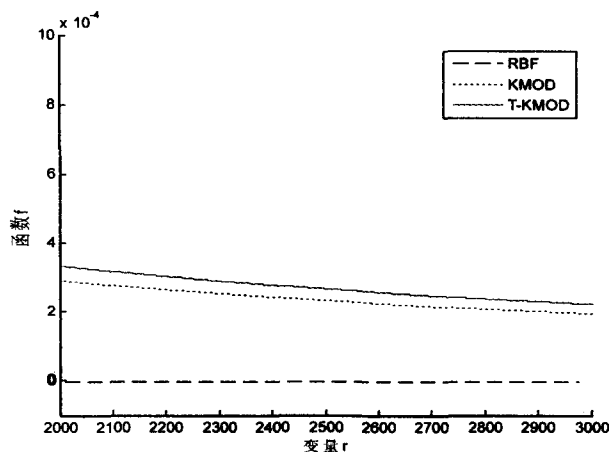


Fig 2 The trends far away the origin point for KMOD, RBF and UKF kernels

Then, in order to compare T-KMOD with RBF and KMOD kernels, draw their figures in the near and far neighborhood of zero in Figs. 1 and 2. It can be seen that T-KMOD kernel satisfies quick decrease in the neighborhood of zero point and moderate decrease toward infinity, whereas the RBF kernel may satisfy the first requirement but not the second, where  $n = 2, \sigma = 1, \gamma = 1$ .

However, from Figs. 1 and 2, see that T-KMOD decreases more slowly than KMOD in the neighborhood of

zero point. Extend (3) as follows:

$$T\text{-KMOD}(x,y)=L\sum_{i=1}^n\frac{(\frac{\gamma}{\|x-y\|^2+\sigma^2})^{k+i}}{i!},i,n\in Z,k\in R,n\geq 1,k\geq 0\quad (4)$$

where  $L$  is a normalization constant. The parameters  $n,k$  and  $\gamma,\sigma$  control decreasing speed around zero point and width, respectively. In fact, Eq. (4) is just a special case of Eq. (3). For simplicity, use the same symbol T-KMOD as Eq. (4). Prove T-KMOD in Eq. (4) is a Mercer kernel.

Then, in order to compare the presented kernel (T-KMOD) with RBF and KMOD kernels, draw their figures in the near and far neighborhood of zero corresponding to different parameter values in Fig. 3, Fig. 4. From Fig. 3 and 4, it can be seen that T-KMOD kernel decreases more quickly than KMOD in the neighborhood of zero point, where  $n=2,\sigma=1,\gamma=1,k=1.2$ .

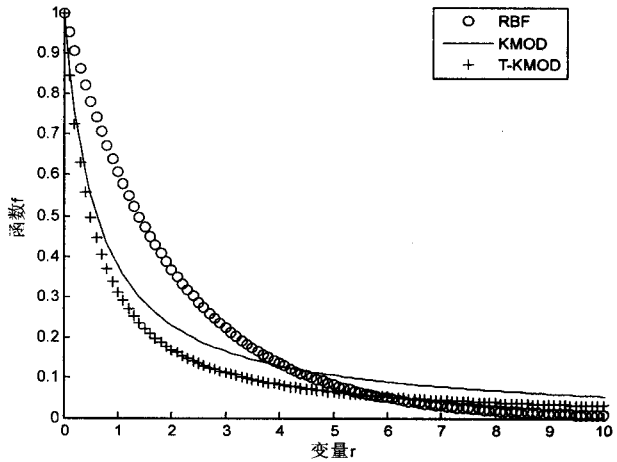


Fig 3 The trends near the origin point with different parameter values for KMOD,RBF and UKF kernels

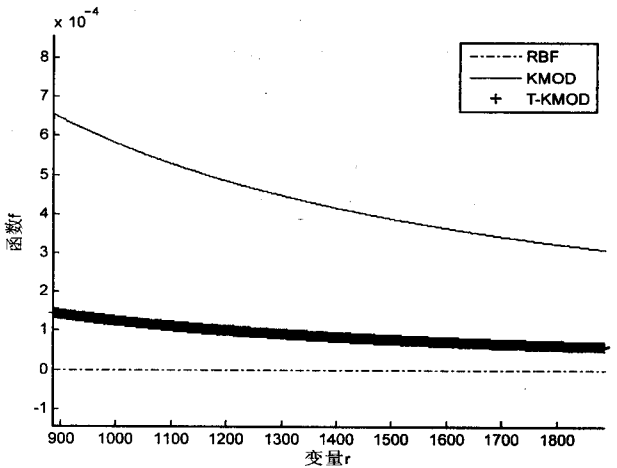


Fig 4 The trends far away the origin point with different parameter values for KMOD, RBF and UKF kernels

T-KMOD kernel is flexible to select. Selecting different  $n,\sigma,\gamma$ , will construct different kernel functions, which satisfies both of the requirements: a quick decrease in the neighborhood of zero point and a moderate decrease toward infinity. Hence it might serve as a kind of universal kernel which can replace (by selecting the appropriate parameters) the set of commonly applied kernel functions, i. e., the linear, polynomial and Radial Basis Function kernels, etc.

2 Experiment

It takes the benchmark KDD Cup 1999 (KDD99, KDD cup, 1999)<sup>[12]</sup> as the dataset of the experiments. Because the data set KDD is too large for the experiment, select a subset of KDD, about 5% of it. There are some performance indicators for the intrusion detection system as follows TP, FP, TN and FN:

$$\text{True Negative Rate: } TNR = \frac{TN}{TN+FP} \quad (5)$$

$$\text{False Negative Rate: } FNR = \frac{FN}{TP+FN} \quad (6)$$

Where TNR denotes the detection rate and FNR denotes the false alarm rate<sup>[13]</sup>.

Randomly select samples from the subset to form the training and test set. There are three data sets in Table 1. The percentage of the normal samples in each set is different to verify the effectiveness of T-KMOD.

Table 1 Three training and test sets

No.	Training set		
	Normal( % )	Abnormal( % )	Total
DS1	65.5	34.5	5023
DS2	73.8	26.2	9342
DS3	55.7	44.3	8487
No.	Test set		
	Normal( % )	Abnormal( % )	Total
DS1	45.3	54.7	12918
DS2	80.6	19.4	10249
DS3	68.9	31.1	13540

In this simulation experiment, adopt the soft margin SVM (C-SVM) as the classifier. Use matlab 7.1 and libsvm 3.1 as supplementary tools. The parameter  $\sigma$  of RBF is set to be 0.5,  $\gamma$  and  $\sigma$  of KMOD is set to be 0.5 and 0.5 respectively;  $\gamma,\sigma,k$  and  $n$  of T-KMOD is set to be 0.5, 0.5 and 1.5 respectively; and the parameters degree and coef. of POLY are set to be 3 and 1, respectively. The penalty parameter C of C-SVM is set to be 2000.

The results are listed in Table 2. First, from the table 2, see that the detection rates of T-KMOD is the highest. Second, the detection rate and the false alarm rate are two constraining indicators, so a higher detection rate may result in higher false alarm rate. T-KMOD has the highest detection rate, however it has a higher false alarm rate than KMOD, it's lower than POLY, and still in the acceptable range. Third, as shown in Table 2, the training time of T-KMOD is the least, which indicates T-KMOD has good performance in reducing the training time.

Table 2. Comparison of the performance of the four kernel functions

Set	C-SVM (kernels)	Detection rate(%)	False alarm rate(%)	Train time(s)	Test time(s)
DS1	POLY	84.134	15.866	15.350	7.842
	RBF	83.544	16.456	13.739	10.948
	KMOD	87.451	12.549	5.823	5.382
	T-KMOD	91.692	8.308	2.431	2.732
DS2	POLY	85.832	14.168	30.552	6.930
	RBF	86.358	13.642	58.398	25.932
	KMOD	90.683	9.317	20.743	4.539
	T-KMOD	93.802	6.198	15.590	2.473
DS3	POLY	82.483	7.517	48.395	7.892
	RBF	85.928	4.072	68.932	28.392
	KMOD	91.842	8.158	32.478	6.821
	T-KMOD	94.515	5.485	10.892	3.589

The above results show that the performances on network intrusion detection of T-KMOD is more better than that of other kernel functions.

### 3 Conclusion

In this paper, T-KMOD kernel function is described. From the discussion above, know that the T-KMOD kernel has two main advantages: one is that it is not necessary to make a selection out of the above mentioned kernel functions, which can simplify the modeling process and will save much computing time. The other is

that due to its flexibility to vary, it has a stronger mapping ability and can properly deal with kinds of mapping problems.

### References:

- [1] Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods [M]. Cambridge, UK: Cambridge University Press, 2000.
- [2] Remaki L, Cheriet M. Kcs-new kernel family with compact support scale space [J]. IEEE Transactions on Image Processing, 2000, 9(6): 970-981.
- [3] 邓乃扬, 田英杰. 支持向量机-理论、算法与拓展 [M]. 北京: 科学出版社, 2009.
- [4] Ayat N E, Cheriet M, Remaki L, et al. KMOD-A New Support Vector Machine Kernel with Moderate Decreasing for Pattern Recognition, Application to Digit Image Recognition [C]//Proceedings of 6th International Conference on Document Analysis and Recognition. Seattle, USA: IEEE, 2001.
- [5] Schoenberg I J. Metric spaces and completely monotone functions [J]. Annals of Mathematics, 1938, 39(4): 811-841.
- [6] Scholkopf B, Burges C, Smola A. Advances in Kernel Methods: Support Vector Learning [M]. [s. l.]: MIT Press, 1999.
- [7] 张成伟, 郑 诚. 基于改进 SVM 的文本信息检索研究 [J]. 计算机技术与发展, 2009, 19(1): 71-73.
- [8] 王玉震, 李 雷. 基于 SVR 的图像增强方法 [J]. 计算机技术与发展, 2009, 19(1): 60-62.
- [9] 李 雷, 鲁延玲, 周蒙蒙. 基于核方法的一种新的模糊支持向量机 [J]. 计算机技术与发展, 2010, 20(2): 9-11.
- [10] 施其权, 李小明, 肖辞源. 一类新型快速模糊支持向量机 [J]. 计算机技术与发展, 2010, 20(2): 103-105.
- [11] 侯惠芳, 白莉媛, 刘素华. 基于支持向量机的图像边缘检测研究 [J]. 计算机工程与应用, 2007, 43(18): 32-33.
- [12] KDD99, KDD cup 1999 data [EB/OL]. 1999. <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>.
- [13] Yi Yang, Wu Jiansheng, Xu Wei. Incremental SVM based on reserved set for network intrusion detection [J]. Expert Systems with Applications, 2011, 38(6): 7698-7707.

(上接第 156 页)

Notes in Computer Science, 2002, 2487: 110-127.

- [12] 张琳琳, 应 时, 倪 友. 一种软件体系结构关注点分析方法 [J]. 计算机学报, 2009, 32(9): 1782-1791.
- [13] 王 伟, 杨 庚, 张迎周, 等. 基于程序切片和服务构件的

语义 Web 服务组合 [J]. 计算机技术与发展, 2011, 21(11): 141-144.

- [14] 李 翔, 怀进鹏, 曾 晋, 等. 一种 Java 遗留系统服务化切分和封装方法 [J]. 计算机学报, 2009, 32(9): 1804-1815.