变元可分离核函数对非线性支持向量分类机的影响*)

原峰山[1] 朱思铭[2]
(广州航海高等专科学校现代化教学中心 广州 510725)[1] (中山大学数学与计算科学学院 广州 510275)[2]

摘 要 证明了变元可分离函数在 Hilbert 空间中满足 Mercer 定理的条件，为构造新的非线性支持向量分类机时选定核函数提供了一种新方法，并通过新方法构造的核函数与其它核函数构造的非线性支持向量分类机比较，得出了较好的结果。

关键词 变元可分离函数，核函数，分类算法

Influence to Nonlinear Classification Support Vector Machines by Separable Variables Kernel Function

YUAN Feng-Shan[1] ZHU Si-Ming[2]
(Guangzhou Maritime College, Guangzhou 510725)[1]
(School of Mathematics and Computing Science, Zhongshan University, Guangzhou 510275)[2]

Abstract It is proved that under Hilbert space separable variables function can satisfy condition of Mercer theorem as kernel function to provide a new method in selecting new kernel function of nonlinear calssification machines based on support vector machine. Compared with other nonlinear classification support vector machines structured by known kernel function, the kernel function selected by new method can give better result.

Keywords Separable variables function, Kernel function, Classification algorithm

1 非线性分类问题的 Hilbert 空间映射及分类优化

在分类问题中，可以通过非线性映射将输入向量 x 映射到一个高维特征空间，在这个空间中构造最优分类超平面。构造时无须刻意追求如何构造这个特征空间，只要能够求出支持向量与特征空间中向量的内积就可以解决问题，这个内积一般用函数 $K(x,x_i)$ 来表示，也称为核函数，它应该满足 Mercer 条件。

非线性分类问题的基本问题是：假设输入向量 x 所在的输入空间是 n 维欧氏空间，即 $X\subset R^n$，需要解决的分类问题在这个空间是非线性的，在构造算法时就会较大的困难[1]。根据支持向量机理论，可以把向量 x 映射到另一个空间 H，称为 Hilbert 空间，也称为特征空间，它通常是无穷维的，对应于一个完备的内积空间。当 n 维欧式空间中的问题映射到 Hilbert 空间后，可以在这个空间中构造一个线性的分类超平面，对问题进行分类，即将非线性问题转化为 Hilbert 空间中的线性分类。

这种非线性分类问题的映射关系可描述如下[2]：

引入从输入空间 $X\subset R^n$ 到 Hilbert 空间的变换：

$$\Phi:\begin{array}{l}X\subset R^n\to H\\ x\to\Phi(x)\end{array}\tag{1}$$

变换后，原来的输入空间 R^n 的训练集：

$$T=\{(x_1,y_1),\cdots,(x_l,y_l)\}\tag{2}$$

映射为 Hilbert 空间的新训练集：

$$\overline{T}=\{(\Phi(x_1),y_1),\cdots,(\Phi(x_l),y_l)\}\tag{3}$$

理论上讲，在原来的欧氏空间的非线性分类问题映射到更高维的 Hilbert 空间后，就转化为线性可分的。训练集样本被线性地正确划分，是由于在 Hilbert 空间构造了一个超平面 $(w\cdot x)+b=0$ 作为线性划分的依据。这个超平面既能够根据样本的特征将它们分类，又能使得训练集 $\overline{T}$ 关于这个超平面几何间隔达到最大，原始分类问题实际上是求解以下最优化问题：

$$\min_{w,b}\ \frac{1}{2}\|w\|^2\tag{4}$$

$$\text{s.t. } y_i((w\cdot x_i)+b)\geqslant 1, i=1,\cdots,l\tag{5}$$

上述问题的对偶问题表达为：

$$\min_{\alpha}\ \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}y_iy_j\alpha_i\alpha_jK(x_i,x_j)-\sum_{j=1}^{l}\alpha_j\tag{6}$$

$$\text{s.t. } \sum_{i=1}^{l}y_i\alpha_i=0,\tag{7}$$

$$0\leqslant\alpha_i\leqslant C\quad i=1,\cdots,l\tag{8}$$

从上述对偶问题的解就可以实现对原始问题的求解，那么由超平面决定的非线性分类问题的决策函数是：

$$\begin{aligned}f(x)&=\text{sgn}((W^*\cdot x)+b^*)\\&=\text{sgn}(\sum_{i=1}^{l}\alpha_i^*y_i(\Phi(x)\cdot\Phi(x_i))+b^*)\\&=\text{sgn}(\sum_{i=1}^{l}\alpha_i^*y_iK(x,x_i)+b^*)\end{aligned}\tag{9}$$

2 变元可分离核函数关系的定义和分析

2.1 核函数和 Mercer 定理

在非线性分类问题中，无论是求解对应对偶的最佳解还是分类决策函数，其中一个非常重要的函数就是 $K(x,x_i)$ 或 $K(x_i,x_j)$，即前面提到的核函数，它的关系式为：$K(x,x_i)=(\Phi(x)\cdot\Phi(x_i))$，即对于非线性分类问题，转化成 Hilbert 空间的线性问题后，影响决策函数的是向量的点积 $\Phi(x)\cdot\Phi(x_i)$，这在构造算法时有很大的困难，支持向量机的巧妙之处就在于尽管需要解决的是 Hilbert 空间内的线性分类问题，但在解决问题时并不需要求解这个内积，只要找到一个可以代替的函数 $K(x,x_i)$，就能够完成分类机的构造。因此，寻找合

*)本课题受到国家自然科学基金项目资助(编号：10371135)；原峰山 博士，高级工程师，研究方向：人工智能与计算机网络。

适的核函数往往是非线性支持向量分类机的一个十分重要的问题。Vapnik 在其著作《统计学习理论》[3]中指出，对分类问题，如果数据是可以被分类的，已经被认可的三种学习机器的核函数是：

1)多项式支持向量机的核函数为：

$$K(x,x_i)=((x\cdot x_i)+1)^d \tag{10}$$

2)径向基函数支持向量机的核函数为：

$$K_\gamma(|x-x_i|)=\exp(-\gamma|x-x_i|^2) \tag{11}$$

3)两层神经网络支持向量机的核函数为：

$$K(x,x_i)=S[(x\cdot x_i)] \tag{12}$$

$S(u)$是 Sigmoid 函数。

核函数对于构造支持向量分类机起决定性的作用，因此，对非线性支持向量分类机，能否找出其他类型的核函数，往往成为构成新的非线性分类学习机的关键。现在，将采用一种新的方法力图找到新的核函数，扩充构造新的非线性分类机的可能性。

Mercer 定理[3]指出，为了保证一个连续对称函数 $K(u,v)$在 $L_2(C)$空间中有一个含正系数 $\alpha_k>0$ 的展开式：

$$K(u,v)=\sum_{k=1}^{\infty}\alpha_k z_k(u)z_k(v) \tag{13}$$

(即 $K(u,v)$描述了某一特征空间的内积)，充分必要条件是，下列条件：

$$\int_C\int_C K(u,v)g(u)g(v)\mathrm{d}u\mathrm{d}v\geqslant 0 \tag{14}$$

对于所有 $g\in L_2(C)$成立(C 为 R^n 的一个紧子集)。

以上定理决定了满足 Mercer 条件的核函数 $K(u,v)$，存在一个特征空间$(z_1(u),z_2(u),\cdots,z_k,\cdots)$，在这一空间中这个核函数生成内积，即上述表述的展开式，能够保证构造的支持向量机在 Hilbert 空间中内积结构具备好的性质。

2.2 变元可分离核函数的性质

定义一种新的概念，即变元可分离函数，一个函数 $K(u,v)$是变元可分离的，即 $K(u,v)=k(u)k(v)$。在 $L_2(C)$的空间中，对于所有 $g\in L_2(C)$，就可以保证$\int_C\int_C k(u)k(v)g(u)g(v)\mathrm{d}u\mathrm{d}v\geqslant 0$，从而保证$\int_C\int_C K(u,v)g(u)g(v)\mathrm{d}u\mathrm{d}v\geqslant 0$，那么由 $k(u)k(v)=K(u,v)$定义的函数是核函数，也满足：

$$k(u)k(v)=K(u,v)=\sum_{k=1}^{\infty}\alpha_k z_k(u)z_k(v) \tag{15}$$

上式说明，如果找到一个函数，在定义的 $L_2(C)$空间(C 是有界的 *Hilbert* 空间)中，若它的一元变量函数的乘积可以表示成对应的两元函数时，那么这个函数可以成为核函数。

证明：根据文[2]，若 C 为 R^n 的一个紧子集，则 C 是闭有界的。而 $L_2(C)$是定义的一个 Hilbert 空间，当定义的连续函数 f 和 g 的内积为：$(f\cdot g)=\int_C f(u)g(u)\mathrm{d}u$，则满足 $\|f\|=\int_C f(u)^2\mathrm{d}u<\infty$。对于所有 $g\in L_2(C)$，并且 $K(u,v)=k(u)k(v)$，在 C 上求积分：

$$\begin{aligned}&\int_C K(u,v)\left[\int_C g(u)g(v)\mathrm{d}u\right]\mathrm{d}v\\&=\int_C\int_C K(u,v)g(u)g(v)\mathrm{d}u\mathrm{d}v\\&=\int_C k(u)k(v)\left[\int_C g(u)g(v)\mathrm{d}u\right]\mathrm{d}v\\&=\int_C\int_C k(u)k(v)g(u)g(v)\mathrm{d}u\mathrm{d}v\\&=\int_C k(u)g(u)du\int_C k(v)g(v)\mathrm{d}v\\&=(k\cdot g)(k\cdot g)\\&=(k\cdot g)^2\geqslant 0\end{aligned} \tag{16}$$

满足了 Mercer 定理的条件，所以满足 $K(u,v)=k(u)k(v)$关系的变元可分离函数可以成为核函数，由此引入的核函数称为变元可分离核函数。

这个证明说明，若将原来的 R^n 向量空间映射到无穷维的 Hilbert 空间，其定义的内积可以用变元可分离函数作为核函数描述特征向量，扩展了核函数的使用空间和意义。同时，还提供了一种寻找新的核函数的方法。下面将看到变元可分离核函数在实现非线性支持向量分类机算法方面的意义。

3 变元可分离核函数构造支持向量分类机的算例实现及特性分析

3.1 算例实现及与其它核函数的比较

根据前面的结论，变元可分离核函数可以实现新核函数的筛选，按定义 $\exp\left[-\left(\frac{u+v}{2\sigma^2}\right)\right]$就是一个变元可分离函数，突破了几种常见的核函数的局限性，为确定新的核函数提供了一种新的方法。现在以新方法选出的核函数 $\exp\left[-\left(\frac{u+v}{2\sigma^2}\right)\right]$为例进行基于支持向量机的分类处理运算，并与常见的几种核函数[5]进行分类结果的比较。

第一种：线性可分支持向量分类机，核函数为 $K(u,v)=(u\cdot v)$。

对线性不可分而非线性可分的有：

第二种：多项式核函数非线性支持向量分类机，核函数为：$K(u,v)=(u\cdot v)^d$。

第三种：Gauss 径向基核函数非线性支持向量分类机，核函数为：$K(u,v)=\exp\left(\frac{-\|u-v\|^2}{2\sigma_1^2}\right)$。

第四种：新选出的可变元函数构成的核函数形成的非线性支持向量分类机，核函数为：$K(u,v)=\exp\left(-\left\|\frac{u+v}{2\sigma_2^2}\right\|\right)$。

为了比较更为合理，在实现分类时对第三种和第四种参数选择尽量一致，取 $\sigma_1=\sigma_2=0.8$，并且四种分类 C 值都取 200，而多项式分类取 $d=2$。

算例实现如下：这是一个对不同颜色点分类的问题，被分类点见图 1。利用 Matlab，参考 Steve Gunn 开发的软件[3]经过改进后进行分类，其中被分类的数据是相同的，但是采用的核函数不同。

分类前数据分布如下图，可以看出红色和蓝色的点在区域上有一定的分布规律，但是有一定的重叠(如图 1 所示)。

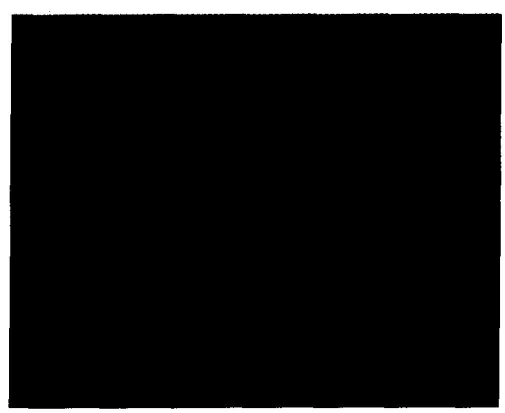

图 1

第一组：线性支持向量分类机组，分类结果如图 2 所示。

图中选定参数 $C=200$，结果：分类运算时间 0.4 秒，支持向量所占比例为 26%，最大间隔 0.216249，构造的分隔面较宽，但是错分度较高。

第二组：多项式核函数非线性支持向量分类机组，分类结

果如图 3 所示。

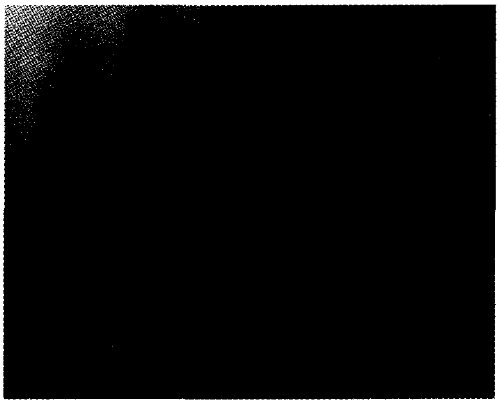

图 2

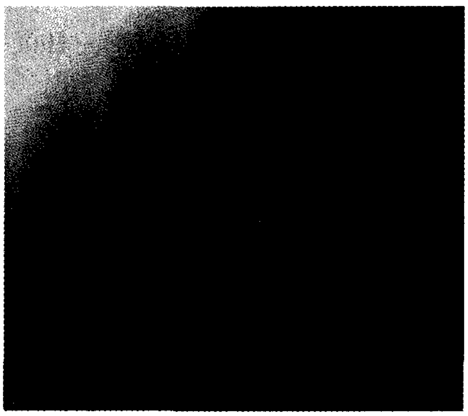

图 3

图中选定参数指数 $d=2,C=200$，结果：分类运算时间 0.3 秒，支持向量所占比例 26.0%，最大间隔 0.179460，构造的分隔面稍窄，有一定的错分度，但是对分隔线的分隔趋势有一定的显现。

第三组：Gaussian 径向基核函数非线性支持向量分类机组，分类结果如图 4 所示。

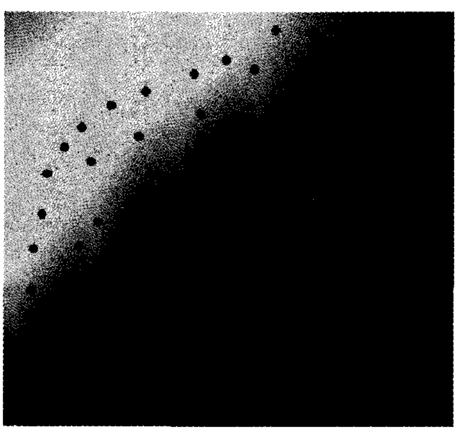

图 4

图中选定参数指数 $\sigma=0.8,C=200$，结果：分类运算时间 0.3 秒，支持向量所占比例 29.0%，最大间隔 0.097352，构造的分隔面窄于多项式分类，有一定错分度，较好地实现了非线性分类。但是，有两个支持向量被错划在分隔线以外。

第四组：新选出的可变元函数构成的核函数形成的非线性支持向量分类机组，分类结果如图 5 所示。

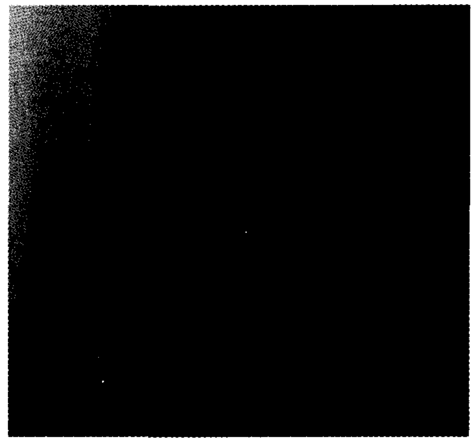

图 5

图中选定参数指数 $\sigma=0.8,C=200$，结果：分类运算时间 0.3 秒，支持向量所占比例 28.0%，最大间隔 0.164861，构造的分隔面窄于多项式分类但高于 Gaussian 核分类，有一定错分度，但好于 Gaussian 核分类，较好地实现了非线性分类。

由以上的结果可以看出，新的核函数构造支持向量分类机有较好的非线性分类性能。将上述几组数据归纳如表 1。

表 1

分类器种类 / 参数和结果	线形分类器	多项式分类器
d,σ,C 等值	$C=200$	$d=2,C=200$
分类运算时间(秒)	0.4	0.3
支持向量所占比例	26%	26%
最大分类间隔	0.216249	0.17946
分类器种类 / 参数和结果	Gaussian 分类器	可变元核函数分类器
d,σ,C 等值	$\sigma=0.8,C=200$	$\sigma=0.8,C=200$
分类运算时间(秒)	0.3	0.3
支持向量所占比例	29%	28%
最大分类间隔	0.097352	0.164861

从表中所列可知，新选定的核函数构造的分类间隔有较大的分类间隔，大于 Gaussian 和分类器，与多项式分类器近似，并且分隔时所需的支持向量所占的比例也较小。

3.2 新核参数选择对分类机的影响

根据可变元核函数选定方法定出的新的核函数，虽然有较好的分类性能，但是随着它的两个重要参数 σ 和 C 选择不同，对分类的效果也有较大的影响。这里共选择了 24 组的 σ-C 组合搭配，构造新核，从中找出基于新核的较好的新的支持向量分类机。

第一组：σ-C-M 组合，即 σ 值-C 值与新的分类机最大分类间隔 M 之间的组合，其中 σ 的取值为[0.6,6]，C 的取值为[100,400]，经过新的分类机的处理，得到不同的最大分类间隔，从中找出较优的支持向量分类机。

第二组：σ-C-SV 组合，即 σ 值-C 值与新的分类机所需支持向量百分比之间的组合，其中 σ 的取值为[0.6,6]，C 的取值为[100,400]。在处理同样的数据组合中，如果所需的支持向量数占所有向量数的百分比越小，说明构造的支持向量分类机越好，从而找出较好的支持向量分类机。

(下转第 174 页)

类似上述过程,可以得到关于决策类Ψ的确定规则为

(d,1)→(e,Ψ);(a,2)∧(b,3)∧(d,1)→(e,Ψ);(b,0)∨(b,2)→(e,Ψ)

关于决策类Ψ的可能规则为

(a,2)∧(b,3)→(e,Ψ);(a,2)∧(d,0)→(e,Ψ);(a,3)∧(d,3)→(e,Ψ);(b,0)→(e,Ψ)

结束语 粗糙集理论发展至今,以其独特的优势在智能信息处理领域占据着重要的地位。由于现实世界中广泛存在的数据不完备性,促进了各种拓展粗集模型的产生与发展。本文所考虑的不完备信息系统既有遗漏型,又有丢失型未知属性值,综合了以往学者所讨论的不完备信息系统中未知属性值所具有的两种解释,因此具有更为广泛的意义。在此基础上,对信息系统中的数据不完备性进行了讨论,使用集对分析的方法建立了一种新的拓展粗集模型,并与传统的几种拓展二元关系进行了对比分析。最后,研究了不进行约简而直接从信息系统中推导出确定和可能性规则的方法,并用实例进行了验证。

在本文所做工作的基础上,下一步的工作就是对挖掘出来规则的可信度进行度量。另一方面,亦可进行广义不完备信息系统中高阶规则,即知识依赖的挖掘方法,并对部分知识依赖的度量方法进行更为深入的研究。

参考文献

1 Pawlak Z. Rough set theory and its applications to data analysis [J]. Cybernetics and Systems, 1998, 29: 661~688

2 Pawlak Z. Rough sets and intelligent data analysis [J]. Information Sciences, 2002, 147: 1~12

3 王国胤. Rough集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39(10): 1238~1243

4 Kryszkiewicz M. Rough set approach to incomplete information systems [J]. Information Sciences, 1998, 112: 39~49

5 Stefanowski J, Tsoukias A. Incomplete information tables and rough classification [J]. Computational Intelligence, 2001, 17: 545~566

6 Wu Weizhi, Zhang Wenxiu, Li Huaizu. Knowledge acquisition in incomplete fuzzy information systems via the rough set approach [J]. Expert Systems, 2003, 20(5): 280~286

7 Grzymala-Busse J W. On the unknown attribute values in learning from examples. In: Proceeding of the 6th International Symposium on Methodologies for Intelligent Systems (ISMIS-91), Charlotte, North Carolina, October, 1991. Lecture Notes in Artificial Intelligence. vol 542. Springer-Verlag, Berlin, Heidelberg, New York, 1991. 368~377

8 Grzymala-Busse J W, Wang A Y. Modified algorithms LEM1 and LEM2 for rule induction from data with missing attribute values. In: Proceeding of the Fifth International Workshop on Rough Sets and Soft Computing (RSSC'97) at the Third Joint Conference on Information Sciences (JCIS'97), Research TrianglePark, NC, March, 1997. 69~72

9 赵克勤. 集对分析及其初步应用[M]. 杭州: 浙江科学出版社, 2000

10 黄兵,周献中. 不完备信息系统中基于联系度的粗集模型拓展[J]. 系统工程理论与实践, 2004, 24(1): 88~92

11 Grzymala-Busse J W. Data with Missing Attribute Values: Generalization of Indiscernibility Relation and Rule Induction. In: Transactions on Rough Sets I, Lecture Notes in Computer Science. vol 3100. Springer-Verlag, Berlin, 2004. 78~95

12 Shan Ning, Ziarko W. Data-based acquisition and incremental modification of classification rules [J]. Computational Intelligence, 1995, 11(2): 357~370

(上接第165页)

图6为第一组数据作出的关系图。

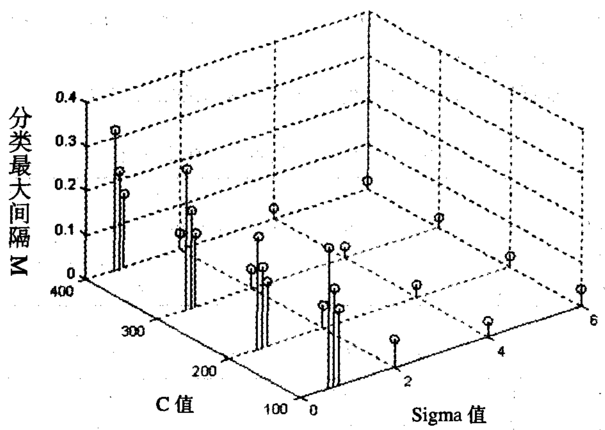

图6

从图6可以看出,当σ值大于2时,最大分类间隔下降,影响分类效果。而C值的变化对分类间隔有一定的影响,但不及σ的影响大。图中还表明,取较小σ的值有利于提高分类机的最大分类间隔。

图7为第二组数据作出的关系图。

从图7可以看出,当σ值大于2时,支持向量分类机中支持向量所占的百分比SV上升,构造的分类机分类效用下降。而C值的变化对SV有一定的影响,但不及σ的影响大,取较小的σ值有利于降低构造支持向量分类机所需的支持向量数,从而使分类机更优化。

结论 (1)经过证明,确定了如果一个函数是变元可分离的,在Hilbert空间中,存在$K(u,v)=k(u)k(v)$关系则可以满足Mercer条件,可以成为核函数,成为选择新的非线性支持向量分类机核函数的一种新方法。(2)根据本例选定的可变元核函数构造的非线性支持向量分类机可以实现较好的分类,参数σ的选择应该小一些,C值应该较大些,但不宜趋于无穷。σ对分类机的影响较敏感,选定时应该注意。

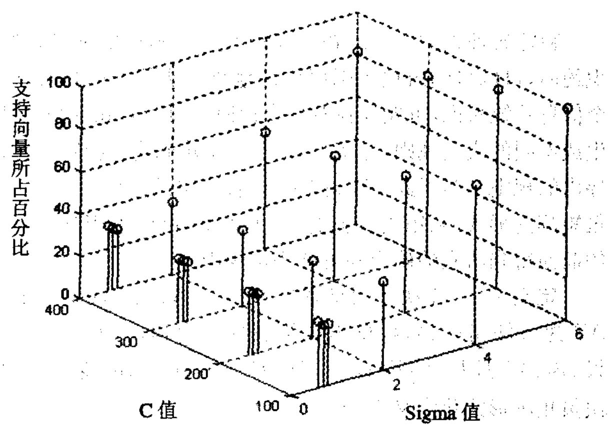

图7

参考文献

1 宋晓峰,陈德钊,俞欢军,胡上序. 支持向量机中的优化算法. 计算机科学,2003(1)

2 邓乃扬,田英杰. 数据挖掘中的新方法—支持向量机. 北京:科学出版社,2004,6

3 张学工译. 统计学习理论的本质. 北京:清华大学出版社,2000,9

4 http://www.kernel_machines.org

5 Boser B, Guyon I, Vapnik V. A training algorithm for optimal margin classifiers. In: Fifth Annual Workshop on Computer Learning Theory. Pittsburgh: ACM Press,1992