一类变量可分离的支持向量分类机的研究与应用

黄文强[1,2] 黄榕波[3] 朱思铭[1]

(中山大学数学与计算科学学院 广州 510275)[1] (中国南方航空股份有限公司计算机中心 广州 510406)[2]

(广东药学院数学系 广州 510226)[3]

摘 要 针对传统SVC方法在样本容量大时存在训练时间过长的不足,建立了一种变量可分离的支持向量分类模型DCSVC,并将其应用于随机生成的模拟数据的学习与航空公司旅客运输数据的预测中。实践证明,DCSVC算法预测的误差小于传统SVM算法,具有较高的精度,且训练时间比传统SVC短。

关键词 支持向量机,分类,变量可分离,预测

Research on Application of a Kind of Support Vector Classification Machine with Variables Seperable

HUANG Wen-Qiang[1,2] HUANG Rong-Bo[3] ZHU Si-Ming[1]

(School of Mathematices and Computational Science, Zhongshan University, Guangzhou 510275)[1]

(Computer Center of China Southern Airlines, Guangzhou 510406)[2]

(Maths Department of Guangdong Medicine Academy, Guangzhou 510226)[3]

Abstract This paper sets up a new kind of support vector regression machine model "DCSVM" which variables can be separated. And the experiment demonstrates that the error of DCSVM is less than traditional support vector machine. We also get a good regression result in the airline passenger volumn forecast using DCSVM method.

Keywords SVM, Classfication, Variable separable, Airline forecast

作为一种以结构风险最小化原理为基础的新算法,支持向量机(Support Vector Machine,简称SVM)具有其他以经验风险最小化原理为基础的算法难以比拟的优越性,它可以转化为求解一个凸二次优化算法,能够保证得到的极值解是全局最优解。当样本数量为n时,该二次规划问题包含$2n$个优化变量、1个等式约束、$4n$个不等式约束,同时涉及到n平方维的核函数矩阵的计算等,因此求解的规模与样本数量有密切关系。为此,如何减少训练时间和降低内存消耗成为SVM研究的一个重要方向。近年来,很多人提出了不同的方法来解决该优化问题。1995年,Cortes和Vapnik提出了选块算法[6],Osuna于1997年提出了分解算法;Joachims提出了求解大型支持向量机中优化问题算法,称为SVM^{light}。John C. Platt于1998年提出了序列最小最优化(SMO)算法[5]。黄榕波在其博士毕业论文[7]中,对一类径向基函数提出了一种分工协作混合系统方法,利用输入向量具有的直和分解性质将优化问题分解为几个子问题,得到了比较好的效果。

本文将分工协作混合的思想推广到支持向量机中,讨论了它的构造以及算法,并利用此思想,对航空旅客运输量相关数据进行机器学习和回归,作出预测,得到的结果比传统SVM方法的结果好。

1 支持向量机及其回归算法

1.1 算法描述

SVM的理论基础是统计学习理论[2,3]。下面以回归问题说明它的算法[4,6,7]。对回归问题,就是考虑用函数$f(x)=wx+b$拟合数据(x_i,y_i),其中,$i=1,\cdots,n$,$x_i\in R^d$,$y_i\in R$。考虑到允许拟合误差的情况,引入松弛因子$\xi_i\geqslant 0$,$\xi_i^*\geqslant 0$,优化问题是最小化

$$R(w,\xi,\xi^*)=\frac{1}{2}w\cdot w+C\sum_{i=1}^{n}(\xi+\xi^*) \tag{1}$$

上式中,常数$C>0$,是对超出误差ϵ样本的惩罚程度。约束条件为

$$\begin{cases} f(x_i)-y_i\leqslant\xi_i^*+\epsilon \\ y_i-f(x_i)\leqslant\xi_i+\epsilon \\ \xi_i,\xi_i^*\geqslant 0 \end{cases} \tag{2}$$

式(1)中第一项使函数更为平坦,提高泛化能力,第二项则为减少误差,C对两者作出折衷。对这一凸二次优化问题,引入Lagrange函数:

$$L(w,b,\xi,\xi^*,\alpha,\alpha^*,\gamma,\gamma^*)=\frac{1}{2}w\cdot w+C\sum_{i=1}^{n}(\zeta_i+\xi_i^*)-\sum_{i=1}^{n}\alpha_i[\xi_i+\epsilon+y_i-f(x_i)]-\sum_{i=1}^{n}\alpha_i^*[\xi_i^*+\epsilon-y_i+f(x_i)]-\sum_{i=1}^{n}(\xi_i\gamma_i+\xi_i^*\gamma_i^*) \tag{3}$$

上式中,α_i、α_i^*为Lagrange乘子,且α_i、$\alpha_i^*\geqslant 0$,γ_i、$\gamma_i^*\geqslant 0$,$i=1,\cdots,n$。

对(3)式进行偏微分,并令各式为0,得到

$$\begin{cases} \sum_{i=1}^{n}(\alpha_i-\alpha_i^*)=0 \\ w=\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)x_i \\ C-\alpha_i-\gamma_i=0 \\ C-\alpha_i^*-\gamma_i^*=0 \end{cases} \tag{4}$$

将式(4)代入式(3),就得到优化问题的对偶形式,最大化函数

黄文强 博士生,主要研究方向:人工智能与计算机网络、数据挖掘;朱思铭 教授,博士生导师,主要研究方向:常微分方程、机器学习。

$$W(\alpha,\alpha^*)=-\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)(x_i\cdot x_j)+\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)y_i-\sum_{i=1}^{n}(\alpha_i+\alpha_i^*)\varepsilon \quad (5)$$

约束条件为

$$\begin{cases}\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)=0\\ 0\leqslant\alpha_i,\alpha_i^*\leqslant C\end{cases} \quad (6)$$

这是一个二次优化问题，w 可以从(4)式得到，再由 $y_i(w\cdot x_i+b)=1$ 和 $\alpha_i\in(0,C)$ 对应的样板计算出 b 的值。

对于非线性回归，一般是先使用非线性映射把数据映射到一个高维的特征空间 H，通过变换 $\Phi:R^d\to H$ 将 x 映射为 $\Phi(x)$，再在高维特征空间 H 中进行回归。此时，优化问题化为在式(6)约束下的最大化函数：

$$W(\alpha,\alpha^3)=-\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i\cdot x_j)+\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)y_i-\sum_{i=1}^{n}(\alpha_i+\alpha_i^*)\varepsilon \quad (7)$$

此时 $w=\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)\Phi(x_i)$，函数 $f(x)$ 可以表示为：

$$f(x)=\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)K(x,x_i)+b$$

根据 Kuhn tucker 定理，在鞍点有

$$\begin{cases}\alpha_i(\varepsilon+\xi_i+y_i-f(x_i))=0 & i=1,\cdots,n\\ \alpha_i^*(\varepsilon+\xi_i^*-y_i+f(x_i))=0 & i=1,\cdots,n\end{cases} \quad (8)$$

可见，对任意一组 α_i,α_i^*，都不会同时为 0，结合 Kuhn tucker 定理和式(4)，得到

$$\begin{cases}\varepsilon+\xi_i+y_i-f(x_i)=0 & \alpha_i\in(0,C)\\ \varepsilon+\xi_i^*-y_i+f(x_i)=0 & \alpha_i^*\in(0,C)\end{cases} \quad (9)$$

由式(9)可以求得 b。令

$\beta_i=\alpha_i-\alpha_i^*$，当 $\beta_i\neq0$ 时，其对应的训练样本就是支持向量。又由于 $\alpha_i,\alpha_i^*\geqslant0$，所以支持向量也就是有一个 Lagrange 乘子(α_i 或 α_i^*)大于 0 的训练样本。

在上面的对偶问题中，寻优函数只设计训练样本之间的内积，所以在高维空间中实际上只需要内积运算(可使用原空间中的函数实现，不必知道变换方式)。

1.2 核函数

SVM 中，不同的内积核函数将形成不同的算法。目前研究比较多的核函数主要有以下四类：

1)多项式核函数

$K(x,y)=(xy+1)^d \quad d=1,2,\cdots$

2)RBF 核函数

$$K(x,y)=\exp(-\frac{\|x-y\|^2}{2\sigma^2})$$

3)Sigmoid 函数

$K(x,y)=\tanh(b(xy)-c)$

4)线性核函数

$K(x,y)=xy$

2 变量可分离支持向量机算法与分析

定义 1(变量可分离函数) 设 $y=f(x)$ 是向量函数，其中 $y\in R,x=(x^{(1)},x^{(2)},\cdots,x^{(d)})^T\in V\subset R^d$。如果存在 V 的分解 $T:V=V_1\oplus V_2\oplus\cdots\oplus V_q$，记 $x(r)=(x^{(1)}(r),x^{(2)}(r),\cdots,x^{(d_r)}(r))^T\in V_r,V_r\subset R^{d_r},r=1,2,\cdots,q$，而且 $\sum_{r=1}^{q}d_r=d$，使得向量函数具有如下形式：

$y=f(x)=\sum_{r=1}^{q}\alpha_r f_r(x(r)),\sum_{r=1}^{q}\alpha_r=1,\alpha_r\in[0,1]$ 则称函数 $y=f(x)$ 关于直和分解 T 是变量可分离函数。

对于训练集合 $\{x_i,y_i\},i=1,2,\cdots,n,x_i\in V\subset R^d,y\in R$，为方便计算，这里我们假定 $x_i,y_i>0$。设存在分解 $x_i=x_i(1)\oplus x_i(2)\oplus\cdots\oplus x_i(q)$，其中 $x_i(r)=(x_i^{(1)}(r),x_i^{(2)}(r)\cdots,x_i^{(d)}(r))^T\in V_r,V_r\subset R^{d_r},r=1,2,\cdots,q$，则训练集合 $\{x_i,y_i\}$ 可以分解为 q 个训练集合 $\{x_i(r),y_i\},r=1,2,\cdots,q$。分别对 q 个训练集合进行支持向量回归，得到 q 个回归函数 $f(x(r))=w(r)\phi(x(r))+b(r)$，再对 q 个回归函数进行加权求和，得到最终回归函数：

$y=\sum_{r=1}^{q}\alpha_r f(x(r))=\sum_{r=1}^{q}\alpha_r(w(r)\phi(x(r))+b(r))$。我们称这种算法为分而治之的支持向量机算法，简称 DCSVM。

在 DCSVM 算法中，加权平均参数 α_r 的选择将直接影响 DCSVM 算法的精度，一般可取分组数量的倒数。作者将另文论述 α_r 的选择方法。

DCSVM 的算法结构如图 1。

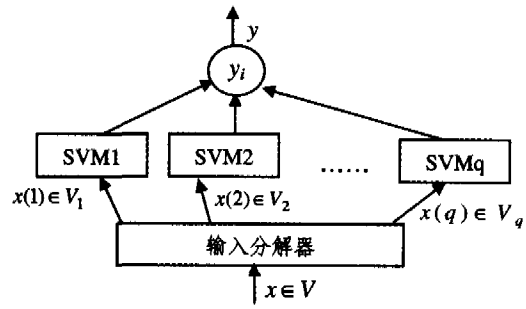

图 1 DCSVM 的结构图

3 数据实验

以下实验中，我们主要考虑预测结果的精度比较，从而体现 DCSVM 与传统 SVM 的区别。

主要精度公式如下：

平均相对误差公式为：

$$MAPE=\frac{1}{N}\sum_{i=1}^{n}|\hat{y}_i-y_i|/|y_i|$$

最大相对误差公式为：

$$MXE=\max_{i=1}^{n}|\hat{y}_i-y_i|/|y_i|$$

均方误差公式为：

$$RMSE=\sqrt{\sum_{i=1}^{N}(\hat{y}_i-y_i)^2/N}$$

均等系数 EC 值公式为：

$$EC=1-\sqrt{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2}/\sqrt{\sum_{i=1}^{n}(\hat{y}_i^2+y_i^2)}$$

3.1 随机函数生成数据回归

我们由以下式子生成 50 个数据点：$\{x_i,y_i\},i=1,2,\cdots,50$。其中 $\{x_i^{(j)}\}_{j=1}^{4},i=1,2,\cdots,50$ 是随机生成的[0,50]均匀分布数据。y_i 的函数值为

$$y_i=\frac{1}{4}\left(\frac{x_i^{(1)}}{x_i^{(1)}+1}+\frac{x_i^{(2)}}{x_i^{(2)}+1}+\frac{x_i^{(3)}}{x_i^{(3)}+1}+\frac{x_i^{(4)}}{x_i^{(4)}+1}\right)+\varepsilon_i$$

其中，ε_i 是服从[$-u,+u$]的均匀分布白噪声。

显然，可以由 $x_i=x_i^{(1)}\oplus x_i^{(2)}\oplus x_i^{(3)}\oplus x_i^{(4)}$ 分解成 4 个 DCSVM 子训练集合 $\{x_i^{(r)},y_i\},r=1,2,\cdots,4$，最后得到学习结果 $\hat{y}=\sum_{i=1}^{4}\alpha_i\hat{y}^i$。其中 α_i 的选择比较关键，将影响整个 DCSVM 模型的学习结果。

这里给出 α_i 的选取方式如下：

1)求每个 DCSVM 分组得到的训练误差及其误差总和；

2)求每个 DCSVM 分组训练误差占误差总和的百分比；

3)将误差百分比由小到大排序，记录每个分组对应误差序列的组号；

4)将组号对应的误差率确定为每个分组的 α_i。

选取前 40 个数据作为训练集合，后 10 个数据作为测试集合。选择不同的核函数分别为多项式核 poly、高斯径向基函数 rbf、线性核 linear、anova 核、一次及二次 ε 不敏感损失函数，ε 值范围为[0.001 0.01 0.1 0.5 1]、C 值范围为[0.1 1 10 100]，ε_i 是服从[−0.5，+0.5]的均匀分布白噪声。选择 MAPE、RMSE、MXE、EC 值进行比较，DCSVM 的训练效果均比传统 SVM 的训练效果好。表 1 选择了两种方法的最优 8 种结果(按照 MAPE 排序)。

3.2 航空公司旅客运输数据预测

航空收益管理理论中，旅客流量的预测是一个重要步骤。航空旅客流量预测的准确与否，关系到下一步航空公司如何发展。旅客流量受到影响的因素很多，价格、当时可销售的座位数量等是主要的影响因素。由于航空运输市场中供需非均衡性的客观存在，因此旅客运输预测比较复杂。常用的旅客量预测方法包括时间序列方法(移动平滑法、指数平滑法、随机时间序列方法)、相关(回归)分析法、神经网络方法及综合的组合预测方法等[1]，这些方法大都集中在对其因果关系回归模型和时间序列模型的分析上，所建立的模型不能全面、科学和本质地反映所预测动态数据的内在结构和复杂特性，丢失了信息量。使用 SVM 方法能够克服这些弱点。一般来说，旅客运输量与飞行小时、飞行班次、平均航线长度、载运率、飞机数量等因素有关系。本文将以上 5 个因素作为输入参数，旅客人数作为输出，选择中国南方航空公司 1978 年到 2002 年的历史运输统计数据作为网络的全部学习样本。选取前 20 个样本进行拟合训练，取后 5 个样本进行外推预测检验，构建传统 SVM 模型及 DCSVM 模型。我们将上述 5 个变量分离为两组变量：第一组变量为飞行小时、飞行班次、平均航线长度，第二组变量为载运率、飞机数量。分别将两组变量作为输入，旅客人数作为输出，将两组 DCSVM 模型的预测结果之和取平均，作为 DCSVM 模型的预测输出。

我们分别采用 Polynomial 核函数、RBF 核函数、ANOVA 核函数和线性核函数等，利用 ε 不敏感及 ε 平方不敏感损失函数，对学习样本进行拟合和预测。实验中，C 值的取值范围为[0.01 0.1 0.4 0.5 0.6 1 10 100 1000 5000]。采用 ε 一次及 ε 二次不敏感两种损失函数进行分别比较，ε 的取值范围为[0.001 0.01 0 0.1 1 5 6 7 8 9 10 15 20]。RBF 核函数中的 σ 取值范围为[0.5 0.75 1 2 5 10]，POLY 核函数中的多项式次数取值范围为[1 2 5]。利用 MATLAB 软件中 SVM 工具包进行实验。实验结果显示，SVM 及 DCSVM 方法中，线性核函数的结果最好，而 RBF 核出现过学习现象。为省略篇幅，表 2 仅列出了结果最好的 8 种情况(按照 MAPE 最小与 EC 最大的原则)。

表 1 利用 DCSVM 及传统 SVM 进行函数拟合的不同参数拟合结果比较

SVM 方法	核函数	C	损失函数次数	ε	MAPE	EC	RMSE	MXE
DCSVM	POLY	100	2	0.001	0.02075	0.98573	0.026365	0.052515
DCSVM	POLY	10	2	0.001	0.020789	0.98586	0.026107	0.050579
DCSVM	POLY	1	2	0.001	0.023883	0.9848	0.02799	0.053288
DCSVM	POLY	10	1	0.01	0.025297	0.98378	0.030233	0.079695
DCSVM	LINEAR	0.1	2	0.001	0.02103	0.98502	0.027672	0.055761
DCSVM	LINEAR	100	1	0.01	0.024538	0.98435	0.02909	0.066208
DCSVM	LINEAR	10	1	0.01	0.027182	0.98346	0.030822	0.075085
DCSVM	LINEAR	1	1	0.01	0.028058	0.9832	0.031319	0.076591
传统 SVM	LINEAR	10	1	0.001	0.026271	0.9827	0.032233	0.0788
传统 SVM	LINEAR	1	1	0.001	0.028507	0.98181	0.033973	0.084266
传统 SVM	LINEAR	0.1	1	0.001	0.028822	0.98171	0.034177	0.084806
传统 SVM	LINEAR	0.1	1	0.01	0.03432	0.97892	0.039423	0.090059
传统 SVM	POLY	100	1	0.01	0.033063	0.97756	0.042019	0.13192
传统 SVM	POLY	10	1	0.01	0.033087	0.97753	0.042066	0.13199
传统 SVM	POLY	10	1	0.001	0.038159	0.97651	0.044236	0.11888
传统 SVM	POLY	100	1	0.001	0.038817	0.9763	0.044645	0.11768

表 2 利用 DCSVM 及传统 SVM 进行航空旅客运输量预测的不同参数预测结果比较

SVM 方法	核函数	C	损失函数次数	ε	MAPE	EC	RMSE	MXE
DCSVM	线性	0.5	1	10	0.028813	0.9826	39.598	0.050885
DCSVM	线性	0.4	1	10	0.029112	0.98252	39.646	0.042384
DCSVM	ANOVA	100	2	0.001	0.030505	0.98118	43.088	0.072673
DCSVM	ANOVA	1	2	0.001	0.030532	0.98116	43.126	0.072624
DCSVM	ANOVA	0.6	2	0.001	0.030549	0.98115	43.151	0.072591
DCSVM	ANOVA	0.5	2	0.001	0.030558	0.98114	43.164	0.072575
DCSVM	线性	0.4	1	5	0.030592	0.98179	41.275	0.045773
DCSVM	线性	0.4	1	0.001	0.031632	0.9804	44.328	0.050347
传统	线性	1	1	5	0.082672	0.95695	100.22	0.1607
传统	ANOVA	1	1	5	0.083247	0.95766	98.904	0.16177
传统	线性	1	1	0	0.086275	0.95528	104.16	0.16614
传统	线性	1	1	0.001	0.086276	0.95528	104.16	0.16615
传统	线性	1	1	0.01	0.086283	0.95528	104.17	0.16616
传统	ANOVA	1	1	0.1	0.086804	0.95614	102.6	0.16757
传统	ANOVA	1	1	0.01	0.086861	0.95612	102.66	0.16767
传统	ANOVA	1	1	0.001	0.086867	0.95612	102.66	0.16768

(下转第 181 页)

表 1 相容决策表

U	a	b	c	d	e	f	g	h	i	D1	D2	D3	D4	D5
1	A	T	T	T	T	C	G	T	T	1	1	1	1	1
2	A	T	T	T	T	G	A	T	T	1	1	1	1	10
3	A	T	T	T	T	G	T	A	G	1	1	1	1	10
4	A	T	T	T	T	T	T	T	C	1	1	1	5	5
5	A	T	T	T	T	T	T	T	G	1	1	1	5	11
6	C	T	T	G	G	A	G	A	G	1	1	1	5	11
7	C	T	T	G	G	C	G	C	T	1	1	1	5	11
8	C	T	T	G	G	C	T	C	T	1	1	3	3	3
9	C	T	T	G	T	G	C	T	A	1	1	3	3	12
10	C	T	T	T	A	A	G	A	A	1	1	3	3	12
11	C	T	T	T	A	T	T	C	T	1	1	3	6	6
12	C	T	T	T	T	T	A	T	A	1	1	3	6	13
13	C	T	T	T	T	T	C	T	T	1	1	3	6	13
14	C	T	T	T	T	T	T	T	T	1	1	3	6	13
15	G	C	A	G	G	A	A	T	G	1	2	2	2	2
16	G	C	A	G	G	A	G	A	C	1	2	2	2	14
17	G	C	A	G	G	G	A	G	A	1	2	2	2	14
18	G	C	A	G	G	G	C	A	C	1	2	2	2	14
19	G	C	A	G	G	G	G	G	A	1	2	2	7	7
20	G	C	A	G	G	G	T	C	C	1	2	2	7	15
21	G	C	A	G	T	T	G	G	T	1	2	2	7	15
22	G	C	A	T	A	G	A	A	A	1	2	2	7	15
23	G	C	A	T	C	A	A	G	C	1	2	4	4	4
24	T	T	T	T	G	G	G	C	C	1	2	4	4	4
25	T	T	T	T	T	A	T	T	C	1	2	4	4	16
26	T	T	T	T	T	C	C	C	T	1	2	4	4	16
27	T	T	T	T	T	T	C	T	T	1	2	4	9	9
28	T	T	T	T	T	T	G	A	G	1	2	4	9	9
29	T	T	T	T	T	T	T	C	A	1	2	4	9	17
30	T	T	T	T	T	T	T	T	T	1	2	4	9	17

表 2 相对 D 核

ϕ	$\{a\}$	$\{a,g\}$	$\{a,g\}$	$\{a,g,i\}$
$core_C(D_1)$	$core_C(D_2)$	$core_C(D_3)$	$core_C(D_4)$	$core_C(D_5)$

从表 2 和表 3 不难看出，完全符合定理 4 的结论。决策粒度越细，它的近似分类精度、分类质量和信息熵[6]就越大。但计算复杂度也相应变大，有时造成计算代价太大。所以，决策粒度不宜太细，决策粒度的细化程度应适度和符合实际需要。

表 3 相对 D 约简

$red_C(D_1)$	$red_C(D_2)$	$red_C(D_3)$	$red_C(D_4)$	$red_C(D_5)$
ϕ	$\{a\}$	$\{a,g,e\}$ $\{a,g,f,d\}$ $\{a,g,h,d\}$ $\{a,g,i\}$	$\{a,g,f,d\}$ $\{a,g,i,h,d\}$ $\{a,g,f,e\}$ $\{a,g,i,h,e\}$ $\{a,g,i,h,f\}$	$\{a,g,i,f,e\}$ $\{a,g,i,h,d\}$ $\{a,g,i,h,e\}$ $\{a,g,i,h,f\}$

结束语 本文分析研究了相容决策表的嵌套决策粒度的相对 D 核的关系和相对 D 约简的关系，得出并证明了粗决策粒度的核一定是细决策粒度核的子集，粗决策粒度的一个相对 D 约简在满足相容性的条件下一定可以扩张为细决策粒度的一个约简，反过来，细决策粒度的一个约简一定可以缩减为粗决策粒度的一个约简。研究结果对知识约简和动态求解问题有一定的实际意义。

5 参 考 文 献

1 王国胤. Rough 集理论与知识获取[M]. 西安：西安交通大学出版社，2001
2 Zhang Wenxiu, Wu Weizhi, Liang Jiye, et al. The theory and methodology of Rough Set [M]. Lin Peng, et al, eds. Beijing: Science Publishing Company, 2000
3 Zhang Wenxiu, Liang Yi, Wu Weizhi. Information System and KDD [M]. Yang Bo Eds. Beijing: Science Publishing Company, 2000
4 张铃，张钹. 模糊商空间理论（模糊粒度计算方法）[J]. 软件学报，2003，14(4)：770～776
5 Skowron A. The rough sets theory and evidence theory [J]. Fundamental Information XIII Intelligence, 1995, 11: 371～388
6 苗夺谦，王珏. 粗糙集理论中概念与运算的信息表示[J]. 软件学报，1999，10(2)：113～116

（上接第 165 页）

根据以上拟合训练和外推预测的结果分析，DCSVM 及传统 SVM 方法均具有较好的推广能力，误差符合预测精度的要求，用于航空航线旅客运输量的预测是有效的。而无论从 MAPE 值还是从 EC、RMSE、MXE 进行比较，均表明 DCSVM 的结果比传统 SVM 要好。

结束语 理论上 SVM 能以任意精度逼近函数。本文建立了一种分而治之的 SVM 学习方法（称为 DCSVM 网络模型），并将该 DCSVM 方法应用于一般的函数逼近与实际中的航空旅客量预测问题，采用 ε 一次及二次不敏感损失函数，分别选用 Linear、RBF、ANOVA 和 Poly 等 4 种核函数建立了 DCSVM 网络模型，其实际结果具有相当理想的精度，比传统 SVM 方法预测的精度更高，可以满足预测要求。如果采用并行计算方法，计算时间将远远小于传统 SVM 模型。

到目前为止，对 SVM 模型的核函数及其参数以及损失函数的选择尚没有确定的方法和结论。对取各种核函数预测后的结果进行二次 SVM 预测，是一个可取的办法。在本文的变量可分离支持向量机 DCSVM 网络模型的基础上，值得考虑的方面包括：一般性的任意变量分离支持向量机、直接影响预测结果的 DCSVM 模型中加权平均系数 a_i 的选择；DCSVM 模型的逼近能力与复杂性分析的理论证明等。限于篇幅，作者将另文论述。

参 考 文 献

1 Adams W, Michael V. Short Term Forecasting of Passenger Demend and Some Application in Quantas . AGIFORS Symposium Proc, 27, Sydney, Australia, 1987
2 Vapnik V N. The Nature of Statistical Learning Theory [M]. NY: Springer2Verlag, 1995
3 边肇祺，张学工. 模式识别[M]. 第 2 版. 北京：清华大学出版社，2000
4 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
5 Platt J C . Fast Training of Support Vector Machines using Sequential Minimal Optimization [R]. Microsoft Research , 2000
6 Cortes C, Vapnik V. Support Vector Networks. Machine Learning, 1995, 20: 273～297
7 Huang Rongbo. A Divide-and-Conquer Hybrid System Based Radial Basis Function Networks: [Doctoral Thesis]. Zhongshan University, 2004
8 Gunn S R. Support Vector Machines for Classification and Regression [R]. University of outhampton , 1998
9 PROS5. 2 Training Course PROS Company 2001
10 中国南方航空股份有限公司. 中国南方航空股份有限公司 2003 年统计年鉴. 2004