

文章编号: 1001-0920(2004)05-0525-04

一种贝叶斯证据框架下支持向量机建模方法的研究

阎威武, 常俊林, 邵惠鹤

(上海交通大学 自动化系, 上海 200030)

摘 要: 研究贝叶斯证据框架下标准支持向量机和最小二乘支持向量机的估计算法和建模方法, 给出一种高斯核支持向量机估计算法的参数选择和调整方法 将贝叶斯证据框架下支持向量机的建模方法应用于非线性系统的辨识, 仿真结果表明, 该方法对于工业过程建模是十分有效的

关键词: 支持向量机; 贝叶斯; 建模; 证据框架

中图分类号: TP274

文献标识码: A

Modeling method based on support vector machines within the Bayesian evidence framework

YAN Wei-wu, CHANG Jun-lin, SHAO Hui-he

(Department of Automation, Shanghai Jiaotong University, Shanghai 200030, China Correspondent: YAN Wei-wu, E-mail: yanww sjtu@hotmail.com)

Abstract Regression algorithms and modeling method based on standard support vector machine (Standard SVM) and least squares support vector machine (LS SVM) within the Bayesian evidence framework are discussed A parameter selecting and tuning method within the Bayesian evidence framework is proposed for SVM regression with Gauss kernel SVM modeling method within the Bayesian evidence framework is applied to nonlinear system identification, and effective results are obtained

Key words: support vector machine; bayesian; modeling; evidence framework

1 引 言

统计学习理论是由 Vapnik^[1]建立起来的一种专门研究小样本情况下机器学习规律的理论, 支持向量机(SVM)是在这一理论上发展而来的一种新的通用学习方法 支持向量机较好地解决了小样本、非线性、高维数、局部极小点等实际问题, 已在模式识别、信号处理以及函数逼近等领域得到了应用, 其中支持向量机估计算法及其应用是相对较新的研究领域 在支持向量机中, 正规化参数和核参数是非常重要的参数, 因此, 寻找最佳正规化参数和核参数的问题也就是最佳模型选择问题 这些参数的选择可

以根据经验、确认集、交叉验证、误差界方法以及统计方法等进行确定^[2].

Mackay^[3]提出一种贝叶斯证据框架, Kowk^[4]则对贝叶斯证据框架下标准支持向量机分类问题进行了研究 本文主要研究贝叶斯证据框架下支持向量机估计算法, 给出一种高斯核标准支持向量机(Standard SVM)和最小二乘支持向量机(LS SVM)估计算法的参数选择和调整方法, 同时给出一种贝叶斯证据框架下支持向量机估计算法的建模方法 将该建模方法应用于非线性系统的辨识, 取得了十分有效的结果

收稿日期: 2003-05-12; 修回日期: 2003-07-28

基金项目: 国家 863 计划资助项目(2001AA 4113130).

作者简介: 阎威武(1971—), 男, 甘肃陇西人, 博士, 从事人工智能、机器学习等研究; 邵惠鹤(1936—), 男, 浙江宁波人, 教授, 博士生导师, 从事工业过程控制等研究

2 支持向量机回归估计算法

设样本集为

$$(x_1, y_1), \dots, (x_l, y_l) \quad R^n \times R.$$

首先用一非线性映射 $\Psi(\bullet)$ 把样本从原空间 R^n 映射到 k 维特征空间 $\Psi(x) = (\mathcal{Q}_{x_1}, \mathcal{Q}_{x_2}, \dots, \mathcal{Q}_{x_l})$, 然后在这个高维特征空间中构造最优决策函数 $f(x) = w \mathcal{Q}_x + b$, 这样非线性估计函数便转化为高维特征空间中的线性估计函数. 利用结构风险最小化原则寻找 w 和 b , 即最小化 $R = \frac{1}{2} w^2 + cR_{\text{emp}}$, 其中: w^2 为控制模型的复杂度, c 为正规化参数, R_{emp} 为误差控制函数. R_{emp} 中选取不同的损失函数, 可构造不同形式的支持向量机. 标准的支持向量机在优化目标中的损失函数为误差 $\xi^{[1]}$. 最小二乘支持向量机在优化目标中的损失函数为误差 ξ 的二次项^[5].

对标准支持向量机, 优化问题为

$$\begin{aligned} \min J(w, \xi) &= \frac{1}{2} w^T w + c \sum_{i=1}^l (\xi_i + \xi_i^*), \\ \text{s.t. } & y_i - w \mathcal{Q}(x_i) - b \leq \xi_i, \\ & w \mathcal{Q}(x_i) + b - y_i \leq \xi_i^*, \\ & \xi_i^*, \xi_i \geq 0, i = 1, 2, \dots, l \end{aligned} \quad (1)$$

最后的非线性估计函数为

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) K(x_i, x_j) + b \quad (2)$$

对最小二乘支持向量机, 优化问题为

$$\begin{aligned} \min J(w, \xi) &= \frac{1}{2} w^T w + c \sum_{i=1}^l \xi_i^2, \\ \text{s.t. } & y_i = w \mathcal{Q}(x_i) + b + \xi_i, \\ & i = 1, 2, \dots, l \end{aligned} \quad (3)$$

最后的非线性估计函数为

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

在支持向量机估计算法中, 正规化参数 c 和核参数 (如高斯 $K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2})$) 中的参数 σ 是非常重要的参数. 下节将针对式(1)和式(3), 讨论贝叶斯证据框架下支持向量机估计算法最优参数的选择问题.

3 支持向量机参数的贝叶斯调整方法

证据框架将贝叶斯推断分为 3 个准则的推断^[3]. 贝叶斯证据框架的基本思想是最大化参数分布的后验, 而最佳参数值或模型是在参数分布后验最大化的情况下得到的. 在支持向量机中, 准则 1 可推断参数 a_i 和 b , 准则 2 可估计正规化参数, 准则 3 可估计核参数.

3.1 准则 1 推断

首先为便于处理, 将优化问题的目标函数除以 c , 并用 λ 代替 $1/c$. 假设 D 为数据空间, H 为模型空间. 准则 1 推断通过贝叶斯规则推断 w 的后验^[3,4]:

$$p(w/D, \lambda, H) = p(D/w, \lambda, H) p(w/\lambda, H). \quad (5)$$

假设训练数据是独立同分布的, $p(w/\lambda, H)$ 服从高斯分布, 则存在

$$p(D/w, \lambda, H) = \prod_{i=1}^l p(y_i/x_i, w, \lambda, H) p(x_i/w, \lambda, H), \quad (6)$$

$$p(w/\lambda, H) = \left[\frac{\lambda}{2\pi} \right]^{\frac{k}{2}} \exp\left[-\frac{\lambda}{2} w^T w\right], \quad (7)$$

其中 $p(x_i/w, \lambda, H)$ 可以看作是常数. 用 $L(y_i, f(x_i))$ 表示损失函数, 则有

$$p(y_i/x_i, w, \lambda, H) = \exp(-L(y_i, f(x_i))). \quad (8)$$

将式(6)~(8)代入式(9), 得

$$p(w/D, \lambda, H) = \exp\left[-\frac{\lambda}{2} w^T w - \sum_{i=1}^l L(y_i, f(x_i))\right]. \quad (9)$$

可以看出, 支持向量机的训练(1)和(3)也即最大化 w 的后验 $p(w/D, \lambda, H)$.

3.2 准则 2 推断

将贝叶斯规则应用于准则 2 推断, 通过最大化正规化参数 λ 的后验 $p(\lambda/D, H)$, 可以求得 λ 的最佳值 λ_{MP} . 由于

$$\begin{aligned} p(\lambda/D, H) &= \frac{p(D/\lambda, H) p(\lambda/H)}{p(D/H)} = p(D/\lambda, H) p(\lambda/H) \\ &= p(D/w, \lambda, H) p(w/\lambda, H) dw \\ &= \left[\frac{\lambda}{2\pi} \right]^{\frac{k}{2}} \exp\left[-\frac{\lambda}{2} w^T w - \sum_{i=1}^l L(y_i, f(x_i))\right] dw. \end{aligned} \quad (10)$$

令 $E_w = \frac{1}{2} w^T w, E_D = \sum_{i=1}^l L(y_i, f(x_i))$, 则

$$\begin{aligned} p(\lambda/D, H) &= \lambda^{\frac{k}{2}} \exp(-\lambda E_w^{MP} - E_D^{MP}) \\ &= \frac{1}{2} (w - w_{MP})^T A (w - w_{MP}) dw = \lambda^{\frac{k}{2}} \exp(-\lambda E_w^{MP} - E_D^{MP}) (2\pi)^{\frac{k}{2}} \det^{-\frac{1}{2}} A. \end{aligned} \quad (11)$$

其中 w_{MP} 表示最优的 w ,

$$\begin{aligned} A &= \frac{\partial^2 (\lambda E_w + E_D)}{\partial w^2} = \\ &= \nabla^2 \left[\lambda E_w + \sum_{i=1}^l L(y_i, f(x_i)) \right]. \end{aligned}$$

式(11) 两边取对数得

$$\ln p(\lambda/D, H) = -\lambda E_W^{MP} - E_D^{MP} + \frac{k}{2} \ln \lambda - \frac{1}{2} \ln(\det A) + \text{常数}, \quad (12)$$

对应于 λ 最大化 $\ln(\lambda/D, H)$, λ 的最佳值 λ_{MP} 可用下式求得:

$$2\lambda_{MP} E_W^{MP} = \mathcal{Y}, \quad (13)$$

其中 $\mathcal{Y} = k - \lambda \text{trace} A^{-1}$ 称为参数的有效数^[3].

在标准支持向量机估计算法中, 损失函数为 $L(y_i - f(x_i) - \theta) = \xi_i, L(f(x_i) - y_i - \theta) = \xi_i^*$. 由于 ξ_i 和 ξ_i^* 没有二阶微分, 则由下面函数代替: $\xi_i = (y_i - f(x_i) - \theta) s(y_i - f(x_i) - \theta), \xi_i^* = (f(x_i) - y_i - \theta) s(f(x_i) - y_i - \theta)$, 其中 $s(u) = \frac{1}{1 + e^{-u}}$ ^[4]. 于是可得

$$A = \nabla^2 \left[\lambda E_w + \sum_{i=1}^l (\xi_i + \xi_i^*) \right] = N + B, \\ B = \sum_{i=1}^l r_i \mathcal{Q}(x_i) \mathcal{Q}(x_i)^T. \quad (14)$$

其中

$$r_i = (y_i - f(x_i) - \theta) s(y_i - f(x_i) - \theta) + 2s(y_i - f(x_i) - \theta) + (f(x_i) - y_i - \theta) s(f(x_i) - y_i - \theta) + 2s(f(x_i) - y_i - \theta).$$

在最小二乘支持向量机估计算法中, 损失函数为 $L(y_i, f(x_i)) = \xi_i^2 = (y_i - w \mathcal{Q}(x_i) - b)^2$, 则有

$$A = \nabla^2 \left[\lambda E_w + \sum_{i=1}^l L(y_i, f(x_i)) \right] = N + B, \\ B = \sum_{i=1}^l 2\mathcal{Q}(x_i) \mathcal{Q}(x_i). \quad (15)$$

用 ρ_i 表示 B 的特征值 在标准支持向量机中, ρ_i 是元素为 $K(x_i, x_j) = r_i \mathcal{Q}(x_i) \mathcal{Q}(x_j)$ 的 $l \times l$ 矩阵 \bar{K} 的特征值; 在最小二乘支持向量机中, ρ_i 是元素为 $K(x_i, x_j) = \mathcal{Q}(x_i) \mathcal{Q}(x_j)$ 的 $l \times l$ 矩阵 K 的特征值 则支持向量机的有效数^[4] 为

$$\mathcal{Y} = k - \lambda \text{trace} A^{-1} = \sum_{i=1}^N \frac{\rho_i}{\lambda + \rho_i}, \quad (16)$$

其中 N ($N = l$) 表示非零特征值的数目

3.3 准则 3 推断

准则 3 通过最大化后验概率 $p(H/D)$ $p(D/H)p(H)$ 选择最优核参数 假设先验 $p(H)$ 在样本上是平坦分布, 可得^[3]

$$p(H/D) \propto p(D/H)$$

$$p(D/\lambda, H) \propto p(\lambda/H) d\lambda \\ p(D/\lambda_{MP}, H) / \sqrt{\mathcal{Y}}, \quad (17)$$

$$\ln p(H/D) = -\lambda_{MP} E_W^{MP} - E_D^{MP} + \frac{k}{2} \ln \lambda_{MP} - \frac{1}{2} \ln(\det A) - \frac{1}{2} \ln(k - \lambda_{MP} \text{trace} A^{-1}) + \text{常数} \quad (18)$$

对核参数最大化 $\ln p(H/D)$, 可得最优的核参数 对于高斯核参数的确定方法, 令

$$\partial \ln p(H/D) / \partial \sigma = 0 \quad (19)$$

由于标准支持向量机中

$$\frac{\partial (\lambda_{MP} E_W^{MP})}{\partial \sigma} = -\lambda_{MP} (a_i - a_j) (a_i^* - a_j^*) \frac{\partial K}{\partial \sigma} = -\lambda_{MP} \sum_{i,j=1}^l (a_i - a_j) (a_i^* - a_j^*) \times \exp \left[-\frac{(x_i - x_j)^2}{2\sigma^2} \right] (x_i - x_j)^2 \sigma^{-3}, \quad (20)$$

$$\frac{\partial \ln(\det A)}{\partial \sigma} = \text{trace} \left[A^{-1} \left(\frac{\partial A}{\partial \sigma} \right) \right] = \text{trace} \left[A^{-1} \left(\frac{\partial K}{\partial \sigma} \right) \right], \quad (21)$$

$$\frac{\partial \ln(k - \lambda_{MP} \text{trace} A^{-1})}{\partial \sigma} = \frac{-\lambda_{MP}}{k - \lambda_{MP} \text{trace} A^{-1}} \text{trace} \left[A^{-2} \left(\frac{\partial K}{\partial \sigma} \right) \right], \quad (22)$$

将式(18), (20) ~ (21) 代入(19), 可得标准支持向量机的最佳核参数为

$$\sigma = \left[\frac{\sum_{i,j=1}^l (a_i - a_j) (a_i^* - a_j^*) \times \lambda_{MP}}{k - \lambda_{MP} \text{trace} A^{-1} \text{trace} \left[A^{-2} \left(\frac{\partial K}{\partial \sigma} \right) \right]} + \frac{\exp \left[-\frac{(x_i - x_j)^2}{2\sigma^2} \right] (x_i - x_j)^2}{\text{trace} \left[A^{-1} \left(\frac{\partial K}{\partial \sigma} \right) \right]} \right]^{\frac{1}{3}}. \quad (23)$$

用同样的方法, 可得最小二乘支持向量机的核参数为

$$\sigma = \left[\frac{\sum_{i,j=1}^l a_i a_j \times \lambda_{MP}}{k - \lambda_{MP} \text{trace} A^{-1} \text{trace} \left[A^{-2} \left(\frac{\partial K}{\partial \sigma} \right) \right]} + \frac{\exp \left[-\frac{(x_i - x_j)^2}{2\sigma^2} \right] (x_i - x_j)^2}{\text{trace} \left[A^{-1} \left(\frac{\partial K}{\partial \sigma} \right) \right]} \right]^{\frac{1}{3}}. \quad (24)$$

在上式中,核宽度 σ 不可能为负值,因此对核参数取了绝对值

4 基于支持向量机的建模方法

基于支持向量机的模型属于黑盒模型,其模型输入输出之间的非线性函数关系由支持向量机实现。贝叶斯证据框架下支持向量机的建模步骤如下:

Step 1: 确定输入输出变量;

Step 2: 对样本数据进行归一化预处理;

Step 3: 选择支持向量机估计算法;

Step 4: 设定初始参数值,对支持向量机进行训练,获取支持向量模型中的系数 a_i ;

Step 5: 用准则2推断迭代求解正规化参数 c ;

Step 6: 用准则3推断迭代求解核参数 α ;

Step 7: 用所求的正规化参数和核参数重新训练支持向量机,返回Step 4多次,选出最优模型;

Step 8: 用建立好的模型进行估计和预测

5 仿真实验

将贝叶斯证据框架下支持向量机的建模方法应用到如下的非线性系统^[6]:

$$y = \sin(\pi x_1 x_2) + (x_3 - 0.5)^2 + x_4 x_5 \quad (25)$$

仿真建模的目的是通过输入变量 $\{x_i | i = 1, 2, \dots, 5\}$ 来估计输出 y ,其中5个输入变量 $\{x_i | i = 1, 2, \dots, 5\}$ 取 $[0, 1]$ 之间的随机数。时间序列训练集由100组样本组成;另外产生100组时间序列测试样本用来校验模型。对数据进行归一化^[7],选择高斯核函数。用泛化均方根误差(RMSE)和最大的泛化绝对误差(MAXE)来评价模型。

图1为标准支持向量机模型的估计值与实际值曲线(其中: $\sigma = 1.2591, c = 0.6366, \epsilon = 0.1$),图2为最小二乘支持向量机模型的估计值与实际值曲线(其中: $\sigma = 0.0122, c = 53.4759$)。表1给出了本文结果(Standard SVM和LS SVM)与其他方法^[6]的结果比较。从图1、图2可以看出,支持向量机模型和最小二乘支持向量机模型的预测值很好地逼近了非线性系统输出的实际值。从表1可以看出,本文方法不论从泛化均方根误差还是最大的泛化绝对误差方面,均比现有的几种方法好。

通过应用结构风险最小化原则,支持向量机对于给定的样本可以实现最小化误差和最小化模型复杂性的某种折衷,从而可以通过控制模型的复杂性来获得好的泛化能力。而优化的正规化参数和核参数则很好地实现了对模型复杂性的控制,因而使得支持向量机模型具有优良的泛化能力。贝叶斯证据框架下支持向量机的参数选择方法是非常有效的,

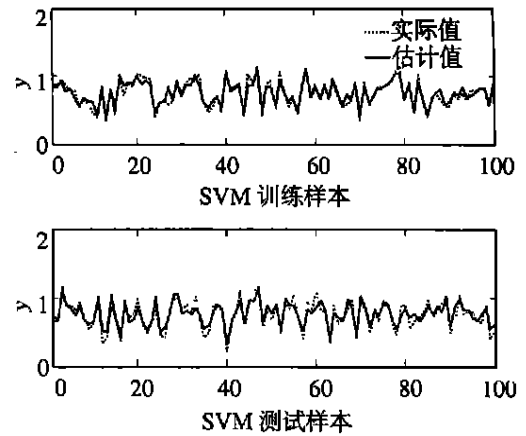


图1 标准支持向量机模型的估计值与实际值曲线

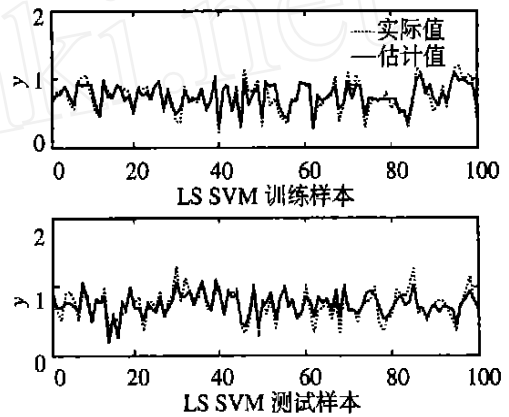


图2 最小二乘支持向量机模型的估计值与实际值曲线

表1 各种模型的结果比较

建模方法	RMSE	MAXE
RBF-OLS	0.1946	0.6259
PLS	0.2534	0.7483
FMM-PLS	0.1695	0.4420
FMM-RBF	0.1901	0.6242
FMM	0.1590	0.4391
LPF	0.1634	0.5791
Standard SVM	0.0952	0.2570
LS SVM	0.1235	0.2837

为支持向量机的应用提供了可靠的支持

6 结论

本文提出了一种贝叶斯证据框架下标准支持向量机和最小二乘支持向量机估计算法的建模方法,给出了一种高斯核支持向量机估计算法的参数选择和调整方法。将贝叶斯证据框架下支持向量机的建模方法应用于非线性系统的辨识,支持向量机模型的预测值很好地逼近了非线性函数输出的实际值。仿真结果表明,贝叶斯证据框架下支持向量机的建模方法是工业过程建模的一种有效的新方法。

(下转第533页)

弦干扰下的最优减振控制问题, 提出了一种前馈反馈次优减振控制器设计的无滞后转换法。仿真结果表明, 与经典反馈最优控制相比, 本文的结果对正弦干扰具有更好的鲁棒性。本文方法可以方便地推广到含有其他类型干扰的时滞系统的次优控制研究中。

参考文献(References):

- [1] Tang G Y, Luo Z W. Suboptimal control of linear systems with state time-delay[A]. *Proc of IEEE Conf on SMC*[C]. Tokyo, 1999. 5: 104-109
- [2] Tang G Y, Fu P L. A suboptimal control approach of linear time-delay systems[A]. *Proc of 14th World Congress of IFAC*[C]. Beijing, 1999. D: 99-104
- [3] 唐功友, 王芳. 具有小时滞的线性大系统的次优控制[J]. *控制理论与应用*, 2003, 20(1): 121-124
(Tang G Y, Wang F. Suboptimal control for linear large-scale systems with small time-delay[J]. *Control Theory and Applications*, 2003, 20(1): 121-124)
- [4] Blanchini F, Sznajder M. Persistent disturbance rejection via static-state feedback[J]. *IEEE Trans on Automatic Control*, 1995, 40(6): 1127-1131.
- [5] Dahleh M A, Shamma J S. Rejection of persistent bounded disturbances: Nonlinear controllers[J]. *Systems and Control Letters*, 1992, 18(2): 245-252
- [6] Savkin A V, Petersen IR. Robust control with rejection of harmonic disturbances[J]. *IEEE Trans on Automatic Control*, 1995, 40(11): 1968-1971.
- [7] Elliott S J, Sutton T J. Performance of feedforward and feedback systems for active control[J]. *IEEE Speech and Audio Processing*, 1996, 4(3): 214-223
- [8] Wu W, Chou Y-S. Adaptive feedforward and feedback control of non-linear time-varying uncertain systems[J]. *Int J of Control*, 1999, 72(12): 1127-1138
- [9] Zhang Y, Mehta P G. Feedforward-feedback adaptive control design and stability analysis[A]. *Proc of 14th World Congress of IFAC*[C]. Beijing, 1999. E: 409-414
- [10] Lindquist A, Yakubovich V A. Optimal damping of forced oscillations in discrete-time systems[J]. *IEEE Trans on Automatic Control*, 1997, 42(6): 786-802
- [11] Tang G Y. Feedforward and feedback optimal control for linear systems with sinusoidal disturbances[J]. *High Technology Letters*, 2001, 7(4): 16-19
- [12] Leitmann G, Pandey S. Aircraft control under conditions of windshear[A]. *Proc of 29th IEEE Conf on Decision and Control*[C]. Honolulu, 1990. 747-752
- [13] Miele A. Optimal trajectories and guidance trajectories for aircraft flight through windshears[A]. *Proc of 29th IEEE Conf on Decision and Control*[C]. Honolulu, 1990. 737-746
- [14] Li H, Hu S J, Cheng Z. Multiple-step predictive control for offshore structures[J]. *China Ocean Engineering*, 1999, 13(3): 231-246
- [15] Fragopoulos D, Spathopoulos M P, Zheng Y. A pendulation control system for offshore lifting operations[A]. *Proc of 14th World Congress of IFAC*[C]. Beijing, 1999. C: 465-470
- [16] Ohmori H, Miyamoto H, Sano A. Sinusoidal disturbance rejection by plug-in adaptive controller[A]. *Proc of 14th World Congress of IFAC*[C]. Beijing, 1999. E: 289-294

(上接第 528 页)

参考文献(References):

- [1] Vapnik V. *The Nature of Statistical Learning Theory* [M]. New York: Springer-Verlag, 1999
- [2] 阎威武. 支持向量机理论、方法和应用研究[D]. 上海: 上海交通大学, 2003
- [3] Mackay D J C. Probable network and plausible predictions—A review of practical Bayesian methods for supervised neural networks[J]. *Neural Computation in Neural Systems*, 1995, 6: 469-505
- [4] Kowk J T. The evidence framework applied to support vector machines[J]. *IEEE Trans on Neural Networks*, 2000, 11(5): 1162-1173
- [5] Suykens J A K. Nonlinear modeling and support vector machines[A]. *Proc of the 18th IEEE Conf on Instrumentation and Measurement Technology*[C]. Budapest, 2001. 287-294
- [6] 李柠. 多模型建模与控制的若干问题研究[D]. 上海: 上海交通大学, 2002
- [7] 樊家琨. 应用多元分析[M]. 开封: 河南大学出版社, 1993. 52-53