第 20 卷 第 1 期　　自 动 化 学 报　　Vol. 20, No. 1
1994 年 1 月　　ACTA AUTOMATICA SINICA　　Jan., 1994

极小极大拟合准则下的线性模型选择方法[1)]

王书宁　戴建设　胡　萍
(华中理工大学系统工程研究所，武汉，430074)

摘　　要

本文指出了在极小极大拟合准则下进行线性模型选择的基本途径，设计了具体的选择方法，并应用这种方法研究了两个实例. 应用结果可以说明这种方法的有效性.

关键词　建模，模型选择，结构辨识，极小极大准则，时间序列拟合.

1　引言

模型选择是建模过程的必要环节. 在线性模型族内，可将该问题一般性地表述如下.

用 $x_j(t)$, $j=0,1,\cdots,L$ 表示一组时间序列，其中某些 $x_j(t)$ 可以是其它时间序列的滞后序列. 用 F 表示下标集 $J=\{1,2,\cdots,L\}$ 的所有子集组成的集类. 所谓线性模型选择就是在 F 中选择一元素 K，使采用线性模型

$$x_0(t)=\sum_{i\in K}\theta_i x_i(t)+e(t) \tag{1}$$

拟合观测数据 $x_j(t)$, $t=t_1,t_2,\cdots,t_M$, $j=0,1,\cdots,L$ 最为合理. 其中各 θ_i 是待定模型参数，$e(t)$ 是模型误差.

对该问题的研究已有较长的历史[1,2]. 业已提出很多模型选择方法，其中 C_P、S_P、AIC 和 CAT 等比较常用[3-9]. 但是，这些方法仅适用于最小二乘拟合准则. 如果在其它准则下使用，则不仅没有理论依据，由于不能解析表达与 F 的各个元素对应的最小拟合误差，为作出最佳选择所要花费的计算量也令人难以接受.

近十余年来，由于鲁棒性问题日益引起人们的关注，采用极小极大拟合准则的建模方法得到了相当广泛的重视[10-13]. 鉴于上述情况，本文拟探讨如何在该准则下进行线性模型选择.

2　模型选择的基本途径

基于对其它模型选择方法的分析，我们认为，从根本上说，模型选择问题是一个双目

本文于 1992 年 8 月 22 日收到.
1) 本项研究得到国家科委（项目批准文件号为(92)国科高字 102 号)、浙江大学工业控制技术国家重点开放实验室(项目编号为 92001) 和霍英东教育基金会(项目编号为 0301081) 的资助.

标极小化问题,其中一个目标是模型拟合误差,另一个目标是模型中独立参数的数目. 本文将以此认识作为解决问题的基础.

对 F 的任一元素 K, 在极小极大拟合准则下,(1)中线性模型的最小拟合误差为

$$E_{\boldsymbol{w}}(\boldsymbol{\theta}(K))=\min\{E_{\boldsymbol{w}}(\boldsymbol{\theta}),\boldsymbol{\theta}\in R^L\mid s.t.\theta_i=0,\forall i\in J-K\} \tag{2}$$

其中, $\boldsymbol{w}\in R^M$ 是根据观测数据设定的各分量为正数的权向量, $\boldsymbol{\theta}(K)$ 表示(2)中优化问题的一个最优解, θ_i 表示 $\boldsymbol{\theta}$ 的第 i 个分量,而 $E_{\boldsymbol{w}}(\boldsymbol{\theta})$ 则由下式定义.

$$E_{\boldsymbol{w}}(\boldsymbol{\theta})=\max\left\{w_t\left|x_0(t)-\sum_{i=1}^{L}\theta_i x_i(t)\right|,t=t_1,t_2,\cdots,t_M\right\} \tag{3}$$

其中 w_t 表示 $\boldsymbol{w}$ 的第 t 个分量. 在 F 上再定义一个泛函 $n(\cdot)$, 对任意的 $K\in F$, $n(K)$ 表示 K 中元素的数目. 根据上述认识, 所求最佳子集应是下述双目标优化问题 P 的解.

$$P:\qquad \min\{E_{\boldsymbol{w}}(\boldsymbol{\theta}(K)),n(K),K\in F\}$$

求解 P 的途径很多[14]. 其中比较常用的是构造一恰当的标量函数, 通过对其求极小确定 P 的最佳调和解. 实际上,前面提到的模型选择方法均可归于这一途径. 但是,由于下面将要指出的理由,本文采用 ε-约束方法,即通过设置拟合误差的满意水平 ε, 求解下述 P^ε 问题获得 P 的解.

$$P^\varepsilon:\qquad \min\{n(K),K\in F\mid s.t.E_{\boldsymbol{w}}(\boldsymbol{\theta}(K))\leqslant\varepsilon\}$$

用 A^ε 表示 P^ε 的最优解集. 如果 A^ε 仅含有一个元素, 则该元素就是 ε 满意水平下的最佳子集, 将其记为 $K(\varepsilon)$, 此时它必是 P 的非劣解. 如果 A^ε 中有多个元素, 可选择 $K(\varepsilon)$ 满足

$$E_{\boldsymbol{w}}(\theta(K(\varepsilon)))=\min\{E_w(\theta(K)),K\in A^\varepsilon\}$$

该 $K(\varepsilon)$ 也必是 P 的非劣解.

采取以上途径求解 P 的主要理由是: 第一,对(3)式定义的最大拟合误差, 容易根据建模目的设置能够满足应用要求的满意水平;第二,不难证明,

定理 1. 对任意的 $\varepsilon^\Delta\in[E_{\boldsymbol{w}}(\boldsymbol{\theta}(K(\varepsilon))),\varepsilon]$, $K(\varepsilon)$ 均是 ε^Δ 满意水平下的最佳子集.

由此可知, 某个满意水平下的最佳子集实际上是某个区间内任意的满意水平下的最佳子集. 这意味着,采取上述途径进行模型选择,所得结果的质量虽然与满意水平的设置有关,但对满意水平的设置有一定的鲁棒性,这就进一步降低了合理设置满意水平的困难性.

3 求解 P^ε 的理论准备

考虑优化问题

$$P_\lambda^\varepsilon:\qquad \min\{\boldsymbol{\lambda}^T(\boldsymbol{\alpha}+\boldsymbol{\beta}),\boldsymbol{\alpha}\in R^L,\boldsymbol{\beta}\in R^L\mid s.t.E_{\boldsymbol{w}}(\boldsymbol{\alpha}-\boldsymbol{\beta})\leqslant\varepsilon,\boldsymbol{\alpha}\geqslant 0,\boldsymbol{\beta}\geqslant 0\}$$

其中 $\boldsymbol{\lambda}\in R^L$ 为给定向量,我们将其称为调节向量.

容易看出, P_λ^ε 等价于标准线性规划问题

$$LP_\lambda^\varepsilon:\begin{cases}\min \sum_{i=1}^{L}\lambda_i(\alpha_i+\beta_i)\\ s.t.\ -\sum_{i=1}^{L}x_i(t)\alpha_i+\sum_{i=1}^{L}x_i(t)\beta_i+\rho_t=w_t^{-1}\varepsilon-x_0(t),\ t=t_1,t_2,\cdots,t_M\\ \qquad\sum_{i=1}^{L}x_i(t)\alpha_i-\sum_{i=1}^{L}x_i(t)\beta_i+\mu_t=w_t^{-1}\varepsilon+x_0(t),\ t=t_1,t_2,\cdots,t_M\\ \qquad\alpha_i\geqslant 0,\ \beta_i\geqslant 0,\ i=1,2,\cdots,L\\ \qquad\rho_t\geqslant 0,\ \mu_t\geqslant 0,\ t=t_1,t_2,\cdots,t_M\end{cases}$$

其中 ρ_t 和 μ_t 是剩余变量或松弛变量.

对任意一对 L 维向量 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$，如果它们与某些 ρ_t 和 μ_t 一起构成 LP_λ^ε 的基本可行解，我们就将其简称为 P_λ^ε 的基本可行解；如果它们与某些 ρ_t 和 μ_t 一起构成 LP_λ^ε 的基本最优解，我们就将其简称为 P_λ^ε 的基本最优解．用 B^ε 和 B_λ^ε 分别表示 P_λ^ε 的全体基本可行解和全体基本最优解的集合．再定义 R^L 到 F 上的一个映射 $I(\cdot)$，其含义是，对任意的 $\boldsymbol{\theta}\in R^L$，$I(\boldsymbol{\theta})$ 表示 $\boldsymbol{\theta}$ 的所有非零分量的下标的集合．可以证明，在 P_λ^ε 和 P^ε 之间存在以下关系．

定理 2. 对任意的 $K\in A^\varepsilon$，存在 $(\boldsymbol{\alpha},\boldsymbol{\beta})\in B^\varepsilon$，满足 $K=I(\boldsymbol{\alpha}+\boldsymbol{\beta})$.

定理 3.[1] 对任意的 $K\in A^\varepsilon$，存在正数 δ，使对任意的 $\boldsymbol{\lambda}\in\Lambda^\delta$，$(\boldsymbol{\alpha},\boldsymbol{\beta})\in B_\lambda^\varepsilon$，满足 $K=I(\boldsymbol{\alpha}+\boldsymbol{\beta})$，其中，

$$\Lambda^\delta=\left\{\boldsymbol{\lambda},\ \boldsymbol{\lambda}\in R^L\,|\,\lambda_i>0,\ \forall i;\ \sum_{i=1}^{L}\lambda_i=1;\ \lambda_i\leqslant\delta\lambda_j,\ \forall i\in K,\ j\in J-K\right\}^{1)}.$$

上述定理 2 表明，通过搜索 P_λ^ε 的基本可行解集可以完全确定 A^ε．而定理 3 则进一步指出，对每个 $K\in A^\varepsilon$，只要指标属于K的每个调节系数与任意一个指标不属于K的调节系数之比小到一定程度，求解 P_λ^ε 即可得到该元素．利用这些事实可以设计求解 P^ε 的有效算法.

4　模型选择方法

基于以上两节的研究结果，我们提出下述完整的线性模型选择方法.

第一步：设置满意水平 ε.

第二步：对各 $x_i(t), i=1,2,\cdots,L$ 进行规范化.

首先求出 $x_i^{\max}=\max\{|x_i(t)|, t=t_1,t_2,\cdots,t_M\}, i=1,2,\cdots,L$；然后令 $\bar{x}_i(t)=x_i(t)/x_i^{\max}, t=t_1,t_2,\cdots,t_M$，$i=1,2,\cdots,L$．对所有的 i 和 t，用 $\bar{x}_i(t)$ 代替 $x_i(t)$ 进行以后各步运算.

第三步：设置初始条件.

令 $\boldsymbol{\lambda}=[1/L\cdots 1/L]^T\in R^L$，$\Lambda=\{\boldsymbol{\lambda}\}$，$\hat{n}=L$，$\hat{A}=\{J\}$.

第四步：改进 $\hat{n}$ 和 $\hat{A}$.

1）限于篇幅，本文所有定理均未给出证明．对定理证明感兴趣的读者可直接与作者联系.

用单纯形法求解 P_λ^ε，对迭代过程中产生的每个基本可行解 $(\boldsymbol{\alpha},\boldsymbol{\beta})$，计算 $n(I(\boldsymbol{\alpha}+\boldsymbol{\beta}))$. 如果 $n(I(\boldsymbol{\alpha}+\boldsymbol{\beta}))<\hat{n}$，令 $\hat{n}=n(I(\boldsymbol{\alpha}+\boldsymbol{\beta}))$，$\hat{A}=\{I(\boldsymbol{\alpha}+\boldsymbol{\beta})\}$；如果 $n(I(\boldsymbol{\alpha}+\boldsymbol{\beta}))=\hat{n}$，仅将 $I(\boldsymbol{\alpha}+\boldsymbol{\beta})$ 加入集类 $\hat{A}$；如果 $n(I(\boldsymbol{\alpha}+\boldsymbol{\beta}))>\hat{n}$，不改变 $\hat{n}$ 和 $\hat{A}$.

第五步：修改调节向量.

设第四步得到的最优解为 $(\bar{\boldsymbol{\alpha}},\bar{\boldsymbol{\beta}})$. 首先求出 $r=\min\{\bar{\alpha}_i+\bar{\beta}_i, i=1,2,\cdots,L \mid s.t.\ \bar{\alpha}_i+\bar{\beta}_i>0\}$；然后令

$$\lambda_i^0=\begin{cases}1/(\bar{\alpha}_i+\bar{\beta}_i) & \text{若 } \bar{\alpha}_i+\bar{\beta}_i>0\\ 1/(0.5r) & \text{若 } \bar{\alpha}_i+\bar{\beta}_i=0\end{cases}\qquad i=1,2,\cdots,L;$$

最后取 $\lambda_i=\lambda_i^0\Big/\Big(\sum\limits_{i=1}^{L}\lambda_i^0\Big)$，$i=1,2,\cdots,L$.

第六步：判断是否停止迭代.

如果由第五步确定的 $\boldsymbol{\lambda}$ 属于 Λ，停止迭代，进到第七步；否则将 $\boldsymbol{\lambda}$ 加进集合 Λ，回到第四步.

第七步：确定最佳模型.

如果 $\hat{A}$ 中仅有一个元素，取该元素为最终选择结果；否则，对所有的 $K\in\hat{A}$，求出 $E_\infty(\boldsymbol{\theta}(K))$，取 $\hat{A}$ 中使 $E_\infty(\boldsymbol{\theta}(K))$ 最小的 K 为最终选择结果.

关于以上模型选择方法，我们给出以下几点说明.

第一，对所有的 $x_i(t)$ 进行规范化是为了使各 $\alpha_i+\beta_i$ 的数值大小对拟合 $x_0(t)$ 具有同等作用. 容易看出，用 $\bar{x}_i(t)$ 代替 $x_i(t)$ 不会改变(2)式确定的最小拟合误差值.

第二，在首次进入第四步时，需采用两阶段单纯形法先确定一个初始可行解，再求最优解. 此时如果 P_λ^ε 没有可行解，即可停止迭代. 在这种情况下继续建模，需要或者增加 ε（这意味着降低对拟合精度的要求），或者增加 L（这意味着要引入其它与 $x_0(t)$ 相关的时间序列）. 在第二次及以后进入第四步时，可用前次求解 P_λ^ε 得到的最优解作为当前的初始可行解. 此时仅需对上述最优解对应的单纯形表修改检验数，就可得到当前的单纯形表.

第三，第五步对调节向量的修改规则具有正反馈的作用，可以促使能够成为 0 的 $\bar{\alpha}_i$ 和 $\bar{\beta}_i$ 在迭代过程中尽快成为 0 .

第四，第六步的停止条件必在有限步迭代内得到满足. 这是因为第四步产生的 $\boldsymbol{\lambda}$ 完全由 $(\bar{\boldsymbol{\alpha}},\bar{\boldsymbol{\beta}})$ 所决定，而 $(\bar{\boldsymbol{\alpha}},\bar{\boldsymbol{\beta}})$ 属于有限点集 B^ε，必在若干次迭代后重复.

第五，有必要指出，上述算法在理论上并不能保证给出真正的 $K(\varepsilon)$. 实际上，只有彻底检查 B^ε 的每个元素才可能做到这一点. 然而，由于定理 3 指出的事实，有理由相信，由此获得的结果，即使不是真正的 $K(\varepsilon)$，也不会和它相差很多.

5 案例研究

5.1 建立我国国民收入和主要物质生产部门产值间的回归方程.

用 $x_j(t)$, $j=0,1,\cdots,6$ 分别表示第 t 年国民收入、农业产值、轻工业产值、重工业产值、建筑业产值、运输业产值和商业产值相对于 1952 年的增长指数. 在文献[15]中有

上述各时间序列自 1952 年到 1990 年的数据．定义 $E_w(\theta)$ 如下．

$$E_w(\theta) = \max\left\{(x_0(t))^{-1}\left|x_0(t) - \sum_{i=1}^{6}\theta_i x_i(t)\right|, t = 1952, 1953, \cdots, 1990\right\}$$

可以看出，$E_w(\theta)$ 表示历年最大相对拟合误差．

我们首先求得 $E_w(\theta(J)) = 0.050$，然后依次取 ε 为 0.055、0.065、0.075 和 0.085，所得结果见表 1．

表 1

ε	$K(\varepsilon)$	$\theta(K(\varepsilon))$	$E_w(\theta(K(\varepsilon)))$
0.050	{1,2,3,4,5,6}	$[0.8436\ -0.0164\ 0.0681\ 0.1238\ 0.0200\ -0.0431]^T$	0.050
0.055 0.065 0.075	{1,3,4}	$[0.8464\ 0.0\ 0.0602\ 0.1225\ 0.0\ 0.0]^T$	0.053
0.085	{1,3}	$[0.9829\ 0.0\ 0.0946\ 0.0\ 0.0\ 0.0]^T$	0.078

5.2　建立我国国民收入和全民所有制单位固定资产投资之间的 ARMA 模型．

用 $x_0(t)$ 和 $x_4(t)$ 分别表示第 t 年国民收入和全民所有制单位固定资产投资相对于 1953 年的增长指数，再令 $x_1(t) = x_0(t-1)$，$x_2(t) = x_0(t-2)$，$x_3(t) = x_0(t-3)$ 以及 $x_5(t) = x_4(t-1)$，$x_6(t) = x_4(t-2)$，$x_7(t) = x_4(t-3)$．在文献[15]中有 1953 年至 1990 年的全民所有制单位固定资产投资数据．利用这些数据和上述国民收入增长指数的数据，能够算出 1956 年至 1990 年的 $x_j(t), j = 0,1,\cdots,7$．再定义 $E_w(\theta)$ 如下．

$$E_w(\theta) = \max\left\{(x_0(t))^{-1}\left|x_0(t) - \sum_{i=1}^{7}\theta_i x_i(t)\right|, t = 1956,1957,\cdots,1990\right\}$$

可以求得 $E_w(\theta(J)) = 0.063$．依次取 ε 为 0.070、0.075、0.080 和 0.085，所得结果见表 2．

表 2

ε	$K(\varepsilon)$	$\theta(K(\varepsilon))$	$E_w(\theta(K(\varepsilon)))$
0.063	{1,2,3,4,5,6,7}	$[1.602\ -.9277\ .3628\ .1948\ -.2872\ .1303\ -.0479]^T$	0.063
0.070	{1,2,3,4,5,7}	$[1.3259\ -.1505\ -.1446\ .1950\ -.2572\ 0.0\ .0575]^T$	0.064
0.075	{1,2,4,5}	$[1.2252\ -0.159\ 0.0\ 0.1495\ -0.1787\ 0.0\ 0.0]^T$	0.072
0.080 0.085	{1,4,5}	$[1.0741\ 0.0\ 0.0\ 0.1709\ -0.1983\ 0.0\ 0.0]^T$	0.079

6　结束语

本文证明了在给定拟合误差的满意水平的前提下，若以独立参数尽可能少为模型选择准则，则可通过搜索某种线性规划问题的基本可行解集确定最佳线性模型．同时还指出了该最佳模型和某种调节向量之间的关系．在此基础上设计了通过修改这种调节向量

找到上述最佳模型的算法．实例计算结果表明这种模型选择方法能够给出令人满意的结果．

参 考 文 献

[1] Thompson, M. L., Selection of Variables in Multiple Regression, Part I and Part II, Int. *Statist. Rev.*, 1978, **46**: 1—19, 129—146.

[2] Linhart, H. and W. Zucchini, Model Selection, John Wiley & Sons, 1986.

[3] Hocking, R. R., Criteria for Selection of a Subset Regression: Which One Should Be Used? *Technometrics*, 1972, **14**: 967—970.

[4] Mallows, C. L., Some Commonts on C_p, *Technometrics*, 1973, **15**: 661—675.

[5] Akaike, H., Fitting Autoregressive Models for Prediction, *Ann. Inst. Statist. Math.*, 1969, **21**: 243—247.

[6] Akaike, H., Information Theory and an Extension of the Maximum likelihood Principle, in Petrov and Czaki eds., Proceedings of the 2nd International Symposium on Information Theory, 1973. 267—281.

[7] Parzen, E., Some Recent Advances in Time Series Modelling, *Trans. Auto. Control*, 1974, **19**: 723—730.

[8] Parzen, E., Multiple Time Series: Determing the Order of Approximating Autoregressive Schemes, in P. R. Krishnaiah ed., Multivariate Analysis IV, North-Holland, Amsterd am, 1977, 283—295.

[9] Linhart, H. and P. Volkers, On a Criterion for the Selection of Models for Stationary Time Series, *Metrika*, 1985, **32**: 181—196.

[10] Launer, R. L and G. N. Wilkinson eds., Robustness in Statistics, Academic Press, 1979.

[11] Poljak, B. T. and J. Z. Tsypkin, Robust Identification, *Automatica*, 1980, **16**:53—63.

[12] A. van den Bos, Nonlinear Least-absolute-values and Minmax Model Fitting, *Automatica*, 1988, **24**: 803—808.

[13] Milanese, M. and A. Vicino, Optimal Estimation Theory for Dynamic Systems with Set Membership Uncertainty: an Overview, *Automatica*, 1991, 27: 997—1009.

[14] Chankong, V. and Y. Y. Haimes, Multiobjective Decision Making: Theory and Methodology, Series Volume 8, North—Holland, 1983.

[15] 国家统计局编，中国统计年鉴 1991，中国统计出版社，1991.

LINEAR MODEL SELECTION WITH MINMAX FITTING CRITERION

Wang Shuning Dai Jianshe Hu Ping

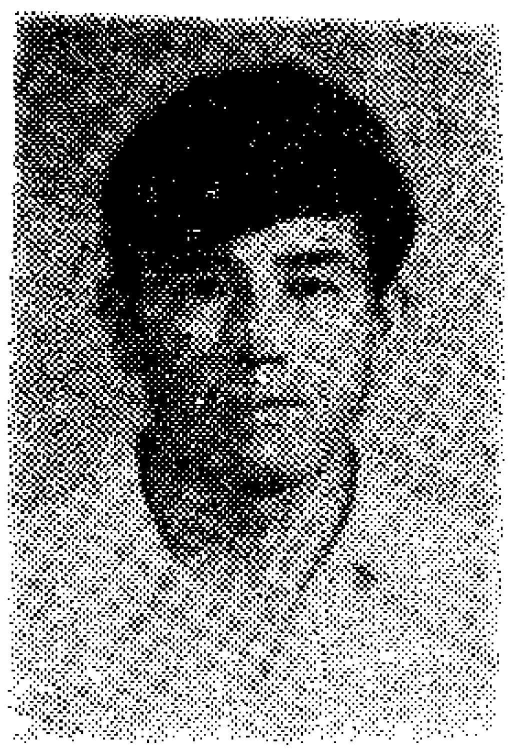
(*Institute of Systems Engineering, Huazhong Univ. of Sci. & Tech., Wuhan*, 430074)

Abstract

The basic principle for linear model selection with minmax fitting criterion is proposed in this paper. With this principle a systematic method is proposed. Two practical examples are studied with the new method. The efficiency of the model selection method can be seen from the results of these applications.

Key words: model building; model selection; structure identification; minmax fitting criterion; time series fitting.

王书宁　1982 年在湖南大学获学士学位，1984 年、1988 年先后在华中理工大学获硕士.博士学位，现为华中理工大学系统工程研究所副教授. 目前主要研究兴趣：系统建模，系统辨识和参数估计以及决策分析. 近年来正式发表学术论文二十多篇.



戴建设　1982 年获上海交通大学自动控制学士学位，1984 年、1990 年先后获华中理工大学系统工程硕士、博士学位，现为该校自控系系统工程研究所副教授. 主要研究兴趣：基于知识的计算机辅助系统、分布式管理与决策信息系统、系统建模以及并行工程. 近年来正式发表学术论文二十多篇.