文章编号: 1001- 2486(2003) 03- 0062- 04

基于选择准则的参数模型评价方法*

段晓君, 王正明

(国防科技大学人文与管理学院, 湖南 长沙　410073)

摘　要: 为评价参数模型的优劣, 分析了模型信息量与数据被模型拟合后的残差信息量之间的关系, 提出了综合考虑模型拟合残差大小、残差信息量与参数数目的一种模型选择的新方法 RIA。结合 RIA 方法, 定义了时序模型评价的一种准则, 并以航天测量数据处理为例, 说明了不同模型在工程实际中的不同表现和本质区别的意义所在。

关键词: 残差; 信息量; 模型选择; 模型评价

中图分类号: C934　　文献标识码: A

Parametric Model Evaluation Based on the Selection Criterion

DUAN Xiao-jun, WANG Zheng-ming

(College of Humanities and Management, National Univ. of Defense Technology, Changsha 410073, China)

Abstract: There is no uniform frame for evaluating a parametric model. A new criterion named by RIA for the model selection is presented which synthetically considers the approximation precision, sparsity of parameters and the residual information. Here the residual information with Gauss distribution is measured by relativity of the residual. According to a simple thought that the whole system is comprised of model information and residual information after modeling the data. This new criterion is transformed into a measurement form for model information. And its reasonability is demonstrated here by theoretic analysis with the application in space data processing.

Key words: residual; information; model selection; model evaluation

模型是科学研究的基础, 参数模型是一种重要的模型类。对参数模型而言, 模型选择和参数求解是数据处理中非常关键的问题。一个好的模型, 除了能够最大限度地发掘提取先验和测量数据中的信息, 还应该有较为简单的形式。参数模型的选择及评价包括模型类的选择和已知模型的阶数确定。

本文考虑模型:

$$y(t)=f(t)+e(t),\quad t\in I \tag{1}$$

其中 y 为观测数据, f 为真实信号, e 为测量误差(不妨先假定 e 为 Gauss 白噪声)。这里模型的选择即包括描述 f 的基函数类选取与定阶。

关于给定模型类的定阶问题, 已有研究主要是在模型的拟合度和复杂性之间进行折衷。通过对拟合优度(fitness)和参数稀疏性(sparsity)的平衡, 研究者[1~8]提出了许多模型定阶准则, 如 AIC[1], BIC[2], MDL[3]等。对于 CV, FPE, MAP, CME 等准则也有一些研究成果[4~8]。FPE 和 BIC 均是 AIC 准则的变形[4]; 而 MAP 需要知道参数先验分布, 但先验的选择有一定困难[4]。文献[4~8]对各种方法的对比作了理论和计算上的说明, 特别是文献[7]中给出了基于不同模型类的各类定阶方法比较。概括说来, 除 CV 准则外, 其它的都采用了惩罚项, 这一类中又以 MAP 准则最为灵活(因为其惩罚项是根据模型不同而自适应变化的), 效果也很好[7,8]。另外, BIC 准则与 MDL 准则的估计结果在渐近统计意义上是一致的[5]。关于模型类(基函数类)的选择, 则远远没有模型定阶那么简单。首先要考虑的是先验信息[9~12], 在几类先验模型中, 再根据不同的准则进行选取[9,10]。

本文研究基于以下思想: 在对数据建立模型后, 信源或系统的信息表现为两部分, 一部分就是模型

* 收稿日期: 2002- 11- 24
基金项目: 全国优秀博士论文作者专项基金(200140); 国防科技大学基础研究项目(JC01- 02- 001)
作者简介: 段晓君(1976—), 女, 博士生。

表现的确定性信息, 另一部分是残差体现出来的信息。相应地, 若残差所含信息量越少, 相对应的模型所体现的信息量就越大。

本文考虑了残差白化程度的因素, 将模型的信息量(包括观测数据和先验分布中信息的利用效率)与残差的相关性及模型系数的稀疏性结合起来, 在已有方法的基础上, 基于 Gauss 分布噪声提出了一种模型选择的新准则 RIA, 并将这种准则转化为模型信息的一种度量方式, 还给出了一个航天测量数据处理的应用算例。

1 同一模型类下基于残差信息量的模型定阶准则

1.1 残差的分布假设

针对线性模型, 分析不同模型处理后的残差序列。首先假设真实模型为:

$$Y = H_{k0}\beta_{k0} + \varepsilon \tag{2}$$

$\varepsilon \sim N(0, \sigma^2 I)$, 为 Gauss 白噪声。假设用 k 阶模型去拟合, 有:

$$Y = H_k\beta_k + \varepsilon_k \tag{3}$$

则拟合残差为:

$$e_k = Y - H_k\hat{\beta}_k = [I - H_k(H_k^{\mathrm{T}} H_k)^{-1}H_k^{\mathrm{T}}]Y \triangleq P_kY = P_k(H_{k0}\beta_{k0} + \varepsilon) \tag{4}$$

由于正态分布的线性组合仍为正态分布, 易知 e_k 也是 Gauss 类噪声。且只有当 $k = k_0$ 时, 残差才是零均值的。以下分析均在残差为 Gauss 分布的假设下进行。

1.2 基于残差信息量的模型定阶准则

几种常用准则均是综合考虑模型拟合精度和参数个数两方面的因素, 即权衡模型适用性和复杂性而提出的。但是, 对于信息量而言, 同样的拟合精度可能有着不同的表现。

在 N 维零均值 Gauss 分布的场合, 假定分布的协方差矩阵为 $\boldsymbol{B}$, 则

$$p(x) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{B}|^{1/2}}\exp\left\{-\frac{1}{2}x\boldsymbol{B}^{-1}x^{\mathrm{T}}\right\}, \quad x \in R^N \tag{5}$$

达到最大熵 $\ln(2\pi e)^{N/2}|\boldsymbol{B}|^{1/2}$, 这里 $|\cdot|$ 为行列式值。从(4)式可知, 欠拟合或过拟合时, 模型不匹配会引起残差的均值不为零。但熵只与协方差有关, 与均值无关, 因为只需在连续型积分中作个变量替换即可。

$$H(x) = -\int_{-\infty}^{\infty} p(x+a)\ln[p(x+a)]\,\mathrm{d}x = -\int_{-\infty}^{\infty} p(u)\ln p(u)\,\mathrm{d}u \tag{6}$$

对于由残差组成的平稳序列$(e_1, e_2, \cdots, e_N)$, 记其自相关矩阵为 Γ, 方差为 σ^2, 我们可将自相关矩阵 Γ 代替自协方差矩阵 $\boldsymbol{B}$, 即

$$\boldsymbol{B} = \begin{bmatrix} r_0 & r_1 & \cdots & r_{N-1} \\ r_1 & r_0 & \cdots & r_{N-2} \\ \cdots & \cdots & \cdots & \cdots \\ r_{N-1} & r_{N-2} & \cdots & r_0 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{N-1} \\ \rho_1 & 1 & \cdots & \rho_{N-2} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{N-1} & \rho_{N-2} & \cdots & 1 \end{bmatrix}\sigma^{2N} = \sigma^{2N}\Gamma \tag{7}$$

残差的方差 σ^{2N} 和自相关矩阵 Γ 分别衡量了模型拟合精度和残差的白化程度。显然, 残差的白化程度越高, 则模型拟合的优度应该越好。

基于上述直观描述, 提出考虑拟合精度、残差信息量和参数稀疏性的折衷方案 RIA(Residual Information Amount), 通过取极小值得到最优结果:

$$\mathrm{RIA}: M_s = \arg\min_{(M_k:\, k\in Z_q)}\{N\ln\hat{\sigma}_k^2 - \ln|\Gamma| + Cd_k\ln N\} \tag{8}$$

这里 $C = 2$, 为经验值。RIA 体现了残差相关性不同的影响, 不仅限于用方差衡量拟合的优度。

这种模型选阶准则所依据的原理如下: 如果模型和数据相匹配, 则残差应是独立噪声序列, 且一般为 Gauss 噪声, 这很直观, 也是残差诊断的基本假设; 如果模型和数据不匹配, 则残差不是独立 Gauss 噪声, 若用此信息准则去度量, 则残差的信息度量结果必将偏大。

从(8)式看来,如果不采用分布信息量的方法度量,而采用似然函数来度量(类似于 AIC 和 MDL 等方法),则不能体现残差序列的相关性,相当于 $\ln|\Gamma|$ 这一项没有考虑。

下面说明考虑相关性的意义。对于色噪声的自相关阵,有以下结论成立:

$$0<|\Gamma|=\left|\begin{bmatrix}1 & \rho & \cdots & \rho_{N-1}\\ \rho_1 & 1 & \cdots & \rho_{N-2}\\ \cdots & \cdots & \cdots & \cdots\\ \rho_{N-1} & \rho_{N-2} & \cdots & 1\end{bmatrix}\right|<\left|\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & \cdots & 1\end{bmatrix}\right|=1 \tag{9}$$

其中 $0\leqslant\rho_i<1$,为自相关函数。

事实上,由于 Γ 非负定,则 $\lambda_i>0(i=1,\cdots,N)$,其中 λ_i 为矩阵 Γ 的特征值。于是

$$|\Gamma|=\prod_{i=1}^{N}\lambda_i\leqslant\left(\frac{\sum_{i=1}^{N}\lambda_i}{N}\right)^N=\left[\frac{1}{N}\mathrm{trace}(\Gamma)\right]^N=\left[\frac{1}{N}N\right]^N=1 \tag{10}$$

(10)式当且仅当 $\lambda_i\equiv1$ 时取等号,即当且仅当白噪声时取等号。

因此,根据(9)式知,对于色噪声,有 $\ln|\Gamma|<\ln1=0\Leftrightarrow-\ln|\Gamma|>0$,故(8)式中取极小值就相当于使残差尽量白化。

1.3 仿真算例

例 类似文献[4],选取以下多项式进行仿真:

$$s[t]=0+0.4t+0.1t^2-0.03t^3+\varepsilon_t,\quad t=1,2,\cdots,n$$

故此,真正的阶数为 $p=4$。选取最大阶数为 $M=6$,对于不同的采样点数 n 和不同的随机误差方差 σ^2,取 ε_t 为正态分布,仿真 1000 次,得到的结论如表 1 所示。

表 1 多项式基的仿真结果比较

Tab. 1 Simulation result based on polynomial model

	$n=30,\sigma^2=10$			$n=30,\sigma^2=100$				$n=100,\sigma^2=10$			$n=100,\sigma^2=100$		
order	4	5	6	3	4	5	6	4	5	6	4	5	6
AIC	710	169	121	0	697	162	141	768	137	95	764	144	89
MDL	861	92	47	0	849	97	54	962	32	6	961	33	6
MAP	999	1	0	12	988	0	0	1000	0	0	1000	0	0
RIA	998	2	0	0	997	3	0	1000	0	0	1000	0	0

可以看到,对多项式基而言,MAP 和 RIA 方法相近,比另外两种方法好,AIC 方法最差。

2 不同模型类表示数据的信息评价

本节研究模型对数据的表示能力,即模型的相对信息含量。

2.1 模型评价思想

由于数据包含的信息体现在两个方面:模型信息含量(model information)和残差信息含量。考虑用同一种模型评价准则来定义模型相对信息的大小。基于一个简单的思想,模型选择的准则最优化函数取极小值,说明准则函数越小,则模型越好。因此可利用 RIA 来度量模型之间的差别。若将 RIA 作为残差的信息含量度量,根据

$$\mathrm{MI}_1=-\mathrm{RIA}_1=-\{N\ln\sigma_{k_1}^2-\ln|\Gamma_1|+Cd_{k_1}\ln N\} \tag{11}$$

$$\mathrm{MI}_2=-\mathrm{RIA}_2=-\{N\ln\sigma_{k_2}^2-\ln|\Gamma_2|+Cd_{k_2}\ln N\} \tag{12}$$

则

$$\Delta(2|1)=\mathrm{MI}_2-\mathrm{MI}_1=\mathrm{RIA}_1-\mathrm{RIA}_2 \tag{13}$$

可作为模型相对信息含量的一种度量,视为模型 2 较之模型 1 的信息差。如果 $\Delta(2|1)>0$,意即模型 2

比模型 1 好, 模型 2 信息更多; 如果 $\Delta(2|1) < 0$, 意即模型 2 比模型 1 差, 模型 1 信息更多。

2.2 模型评价在航天测量数据处理中的应用

在航天测量数据处理问题中, $X_1, X_1, \ldots, X_N$ 为独立或相关测元, 给定多个测站数据, 需要解算出一条弹道。目前解算弹道有两种模型: 弹道逐点解算 EMBET 模型和非线性融合模型[13]。我们需要评价哪一种模型更适于数据处理, 更符合工程实际。

首先从定性的角度分析。非线性融合模型用到了弹道是二次连续可微的函数的先验信息, 还考虑到先验工程背景中弹道的特征点问题和弹道的分段光滑特性, 由此将弹道用三次样条表示, 体现了弹道参数之间的相关性, 可大大节省参数; 而逐点解算模型没有用上任何先验信息, 认为弹道参数没有规律, 相当于服从某区域内的均匀分布, 是无先验情况下熵最大的情形, 体现的信息也少。

再从定量的角度来分析。如, 考虑某弹道解算, 给出相应的逐点模型和非线性融合模型。选取 1500 个点, 考虑 x 方向解算弹道与真值的方差, 用两种模型计算得到的结果分别为: $\sigma_1 = 1.47448143$(逐点), $\sigma_2 = 0.14918804$(非线性); 逐点最小二乘模型解算所需的参数为 $k_1 = 1500$, 非线性融合模型对弹道进行最优节点拟合, 采用 100 个节点, 参数 $k_2 = 102$, 则易计算出模型信息差量为:

$$\Delta(\mathrm{NFM}|\mathrm{EMBET}) = 1.7096 \times 10^4$$

非线性样条融合模型较之逐点模型的优势就体现于这个量上。

模型的区别体现在: 不同的模型处理数据后, 其残差大小、残差白化程度和参数数量不同。区别的本质在于: 好的模型结合了符合实际的工程背景, 描述了事件的自然特性, 这可称为模型先验。如果知道弹道系数的范围, 加上约束和附加信息, 得到更加准确的估计, 这体现的是参数先验, 可使参数后验估计的方差减小。可见, 模型先验(建模过程)比参数先验更重要, 是参数先验的前提。

3 结论

本文考虑了残差白化程度的因素, 将模型的拟合度、模型系数的稀疏性和残差的相关性结合起来, 提出了一种模型选择的新准则 RIA, 并将这种准则转化为模型信息的一种度量方式, 其合理性得到了印证。当然, 评价模型的信息还可有其它准则。如果考虑数据处理的实时性, 则可加上参数估计效率这一项。但参数估计效率的评价本身就是一个难点, 模型评价的难度会更大, 各项之间的平衡参数也需要更深入的研究。

参 考 文 献:

[1] Akaike H. A New Look at the Statistical Model Identification[J]. IEEE Trans. Automat. Contr. AC-19, 1974: 716-723.

[2] Schwarz G. Estimating the Dimension of a Model[M]. Ann. Statist., 1978: 461-464.

[3] Rissanen J. A Universal Prior for Integers and Estimation by Minimum Description Length[J]. The Annals of Statistics, 1983, 11(2): 416-431.

[4] Steven K. Conditional Model Order Estimation[J]. IEEE Trans. on Signal Processing. 2001, 49(9): 1910-1917.

[5] Gustafsson F, Hjalmarsson J. Twenty-one ML Estimators for Model Selection[J]. Automatica, 1995, 31: 1377-1392.

[6] Liavas A P, Regalia P A. On the Behavior of Information Theoretic Criteria for Model Order Selection[J]. IEEE Trans. Signal Processing, 1998, 46: 1689-1695.

[7] Djuric P. Asymptotic MAP Criteria for Model Selection[J]. IEEE Trans. On Signal Processing, 1998, 46: 2726-2735.

[8] Kay S M, Nuttall A H, Baggenstoss P M. Multidimensional Probability Density Function Approximations for Detection, Classification, and Model Order Selection[J]. IEEE Trans. on Signal Processing, 2001, 49: 2240-2252.

[9] Feng Y Q, Pan Q S. Automation Selection of Model Based on Integration of Expert System and Neural Network[J]. Journal of Harbin Institute of Technology in China, 2001, 33(1): 24-27.

[10] Sofiane Brahim-Belhouari, et al. Model Selection via Worst-Case Criterion for Nonlinear Bounded-error Estimation[J]. IEEE Trans. on Instrumentation and Measurement, 2000, 49(3): 653-658.

[11] Lee D S, Chia N K. A particle Algorithm for Sequential Bayesian Parameter Estimation and Model Selection[J]. IEEE Trans. on Signal Processing, 2001, 50(2): 326-336.

[12] Piet M T, Broersen. How to Select Polynomial Models with an Accurate Derivative[J]. IEEE Trans. On Instrumentation and Measurement, 2000, 49(5): 910-914.

[13] 王正明, 易东云, 周海银, 等. 弹道跟踪数据的校准与评估[M]. 长沙: 国防科技大学出版社, 1999: 367-374.