

一类生理信号的采集、建模和辨识

李恬鉴 柳英芝

(同济大学电气工程系·上海, 200092)

摘要: 本文对生理信号的采集、建模和辨识作了较系统的探讨. 在生理信号的采集方面, 介绍了实时和离线信号的采集的设计思想和实现方案. 在生理信号的数据建模方面, 引入了ARMA建模方案, 着重研讨了模型的实现途径、参数辨识和平稳化方法, 编制建模程序及实现框图. 在心功能病类辨识方面, 对典型心功能信号——ECG、CAG和ACG, 根据类比原则, 利用模型分析结果, 对三类心血管病——高血压、冠心病和心肌炎作了聚类辨识, 取得用参数域值辨识病理的结果.

关键词: 生理信号; 时间序列分析; 建模; 参数辨识

1 引言

生物信息处理和生物控制是借助计算机和电子技术, 应用控制论和信息论, 新开拓出来的控制学科的一个分支——信号获取与数据处理.

众所周知, 心脑血管疾病是威胁人类生命最严重(约占总死亡率的60%多)的病种, 因此心血管类生理信号的采集和分析, 为病类辨识和病情预报作出贡献, 具有重大的社会效益和经济效益. 本文则就心功能信号的采集、建模, 并依据模型的参数域值作病类辨识方面进行新的探索.

2 心功能信号的采集^[1,2]

心功能信号的采集和辨识是以下面原则作为前提: (1) 同一病员各种心功能信号之间必然在病理上存在内在的密切联系; (2) 在充分估计病员个体差异性的前提下, 不同病员的同一病例必在各种心功能信号上体现各自的特异性. 据此, 生理信号的无失真采集是医学的物质化、科学化和量化研究的基础. 生理信号的采集分为有创伤与无创伤两种, 无创伤采集方法尤其适合老年人和幼儿.

生理信号的特点是: (1) 信号微弱; (2) 由于人体感应(生物电)和人体旁路, 极易引入外界干扰. 因此信号采集必须适应这些特点对症下药. 用磁带机采集, 先录后放分两步实现, 效率低且增加失真的可能性. 为此我们研制了实时信号采集仪和离线信号采集机.

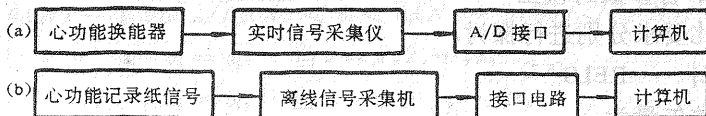


图1 心功能信号采集过程框图

心功能信号包括心电图(ECG)、颈动脉搏动图(CAG)、心尖搏动图(ACG)和心音图(PCG), 前述的实时和离线信号采集设备都具有四通道采集功能, 它们的采集过程框图分

别如图 1 的 (a) 和 (b) 所示。

ECG、CAG 和 ACG 三种心功能信号 (PCG, 因频率成分互异需另作处理) 采集过程, 根据 Shannon 采样定理, 结合三种信号具体情况, 我们采用具有 8 位分辨精度的 250 点/秒的采样率。采样软件设计上, 根据信号特点和数据处理要求, 主程序采用 APPLESOFT 解释语言编制, 而采样子程序则采用 6502 汇编语言编制, 这样既满足各种初始参数设定, 便于人机对话的要求, 又满足对外部输入信号进行 A/D 转换时的快速性要求。

由多导同步生理信号记录仪描记的心电图和该信号经计算机采样结果的离散图象分别如图 2 中的 (a) 和 (b) 所示。

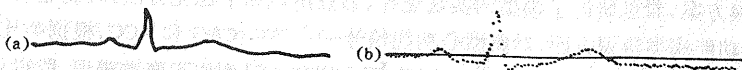


图 2 ECG 及其计算机采样图象

3 生理信号的建模方案及其实现

可以证明, 在离散、连续、标量或矢量情况下, 任何平稳随机过程都可用 $ARMA(n, n-1)$ 模型来逼近。下面介绍一种改进型的时序建模方案, 来对典型心功能信号拟合 $ARMA$ 模型, 模型的通用算子形式为

$$A(q^{-1})y(t) = B(q^{-1})e(t), \quad (1)$$

式中, $y(t)$ 是指具有时序相关性的信号系统输出序列, $e(t)$ 是均值为 0、方差为 σ^2 的高斯白噪声, q^{-1} 是单位滞后算子, 而 A 和 B 是指参数分别为 a_i 和 b_i 的算子多项式

$$A(q^{-1}) = 1 - a_1q^{-1} - a_2q^{-2} - \dots - a_nq^{-n}, \quad (2)$$

$$B(q^{-1}) = 1 - b_1q^{-1} - b_2q^{-2} - \dots - b_nq^{-n}. \quad (3)$$

对模型 (1) 的结构辨识包括^[3,4]:

1) 模型阶和子阶的确定——子阶是指 $ARMA$ 模型的自回归部分的阶 n_a 和滑动平均部分的阶 n_b , 模型的阶则指各子阶的最大值, 即 $n = \max(n_a, n_b)$ 。

2) 模型参数的确定——将不显著异于 0 的参数用 F 检验法删除, 从而得到节省参数的 $ARMA$ 模型, 便于实用。

改进型建模方案具有下述特点:

1° 采用递推增广最小二乘法 (RELS) 实现参数估计——计算和存贮量都小, 便于微机实现, 可得到未知参数的一致估计。

2° 采用 F 检验判决器确定模型的阶和子阶。

3° 可得到节省参数的模型。

下面即按此要求分别进行探讨。

1° 参数估计——RELS^[6,7]

记未知参数矢量为

$$\theta^T \triangleq (a_1, a_2, \dots, a_n; b_1, b_2, \dots, b_n) = (\theta_1, \theta_2, \dots, \theta_{2n}), \quad (4)$$

$$\Phi^T \triangleq [y(t-1), \dots, y(t-n); \hat{e}(t-1), \dots, \hat{e}(t-n)], \quad (5)$$

则在时刻 t 的参数矢量的递推增广最小二乘估计值为

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\hat{e}(t). \quad (6)$$

式中

$$\hat{e}(t) = y(t) - \Phi^T(t)\hat{\theta}(t-1), \quad (7)$$

$$K(t) = \frac{P(t-1)\Phi(t)}{1 + \Phi^T(t)P(t-1)\Phi(t)}. \quad (8)$$

$$P(t) = [I - K(t)]\Phi^T(t)P(t-1). \quad (9)$$

2° 模型阶及子阶的确定—— F 检验判决器^[4,6,7]

此法归结为用 F 检验判决模型中的某些参数是否显著异于 0 的统计假设问题,统计假设准则的判定公式为

$$F = \frac{A_0 - A_N}{s} \div \frac{A_N}{N-r} \sim F_{\alpha}(s, N-r). \quad (10)$$

其中 A_0 : 参数受限制(老模型)的残差平方和(较大),

A_N : 参数不受限制(新模型)的残差平方和(较小),

$F_{\alpha}(S, N-r)$: 表示自由度为 S 和 $N-r$ 的 F 分布, α 为显著性水平.

从条件或静态方面看, $\text{ARMA}(n, m)$ 是线性回归模型, 上述准则可用于 $(n+m)=r$ 个参数中有 S 个参数为 0 的假设检验, 这时 A_N 即为 $\text{ARMA}(n, m)$ 模型的残差平方和, 而 A_0 是去掉 S 个参数后的模型的残差平方和.

为了确定模型的阶 n , 取定显著性水平 α 的值, 即可查得 F 分布临界值 F_{α} .

若 $F < F_{\alpha}$, 则 F 检验不显著, 原模型适用; 若 $F \geq F_{\alpha}$, 则 F 检验显著, 原模型不适用, 需要继续检验新模型的适用性. 设最后确定 ARMA 模型的阶为 n , 则 F 检验只需相继建立 $n+1$ 个模型, 并在相邻模型之间共进行 n 次检验, 即可确定合适的阶, 这一检验过程可由计算机自动完成.

3° 确定子阶的 F 检验判决器^[6,7]

经上述 F 检验虽可得到合适的 $\text{ARMA}(n, n)$ 模型, 但其中某些自回归参数 a_i 和某些滑动平均参数 b_i 的估计值可能很小或接近于 0, 即预示子阶可能比模型的阶要小. 这就归结为判定 $\text{ARMA}(n, n)$ 模型中某些参数是否为 0 的统计假设检验问题.

根据 Åström 和 Wittenmark 的结果^[5], 在已知 N 组观测数据 $y(i) (i=1, 2, \dots, N)$ 的条件下, 未知参数矢量 θ 的条件分布渐近正态分布: 均值 $\hat{\theta}(N)$ 和协方差 $\sigma^2 P(N)$, 其中 $\hat{\theta}(N)$ 和 $P(N)$ 由 (6)~(9) 式递推计算.

记矩阵 $P(N) = [p_{ij}(N)]$, $(i, j=1, 2, \dots, 2n)$, 则未知参数矢量 θ 的第 i 分量 θ_i 的条件分布也是正态的: 均值 $\hat{\theta}_i(N)$ 和方差 $\sigma^2 P_{ii}(N)$. 当 σ 的估计值可由样本计算时, 由正态分布性质得 θ 的 95% 置信区间为

$$\hat{\theta}_i(N) - 1.96\hat{\sigma}\sqrt{P_{ii}(N)} < \theta_i < \hat{\theta}_i(N) + 1.96\hat{\sigma}\sqrt{P_{ii}(N)}. \quad (11)$$

其中 θ_i 指

$$\theta_i = a_i, \quad \text{当 } i = 1, 2, \dots, n,$$

$$\theta_i = b_i, \quad \text{当 } i = n+1, \dots, 2n.$$

显然, 如果估计值 $\hat{\theta}_i(N)$ 很小(近于 0), 则由 (11) 式可知 θ_i 的 95% 置信区间将包含 0, 具有这种估计的参数实际上可能为 0 (即不显著异于 0), 因此可得如下的子阶判决器:

(1) 判定哪些参数的 95% 置信区间已包含 0;

(2) 从所得的 $\text{ARMA}(n, n)$ 模型中删去这些 95% 置信区间包含 0 的参数, 重新用 RELS 方法建立节省参数的模型, 记之为 $\text{ARMA}^-(n, n)$, 它不包含所删去的参数. 继而对

ARMA⁻(n, n)模型进行适用性检验(所删去的参数是否不显著异于 0),即检验 F 统计分布:

$$F = \frac{s^-(n) - s(n)}{s(n)} \cdot \frac{N - 2n}{M} \sim F_\alpha(M, N - 2n). \quad (12)$$

其中 $s^-(n)$ 为 ARMA⁻(n, n) 的残差平方和, $s(n)$ 为原模型的残差平方和, M 为拟删去参数的个数, 取显著性水平 $\alpha = 5\%$ 或 1% , 查 F 分布表可得临界 F_α 值, 如果

1) $F < F_\alpha$, 则 ARMA⁻(n, n) 被接受为适用模型, 即所删去的参数不显著异于 0 (可信为接近于 0);

2) $F \geq F_\alpha$, 则需作进一步分析.

在情况 1), 将所得 ARMA⁻(n, n) 模型与原模型 ARMA(n, n) 比较, 若参数 $a_n, a_{n-1}, \dots, a_{r+1}$ 被删去, 则自回归的子阶为 p ; 如果参数 $b_n, b_{n-1}, \dots, b_{q+1}$ 被删去, 则滑动平均的子阶为 q , 故终审结果得节省参数模型为 ARMA(p, q), 这是多数情况. 情况 2) 可能出现在 n 较小, 因而 $p_n(N)$ 较大和/或 σ^2 较大的情形, 此时置信区间将增大, 这表明有些置信区间包含 0 的参数实际上不为 0 (不能删去). 此时可对置信区间包含 0 的参数运用前述 RELS 方法逐个进行 F 检验, 以便最终拟合适用的模型

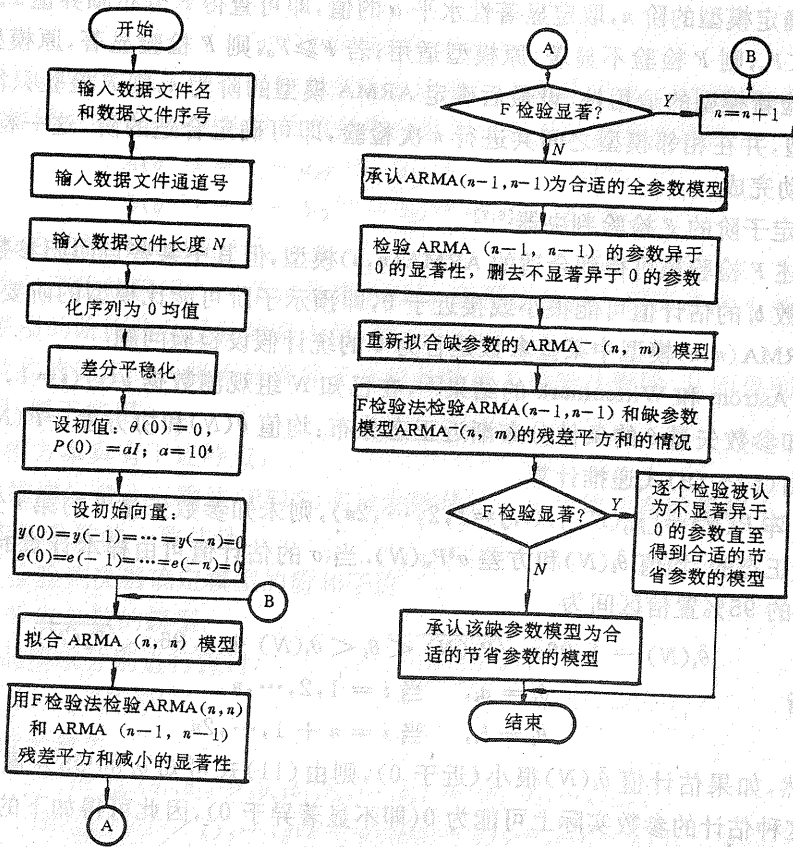


图 3 ARMA 模型实现框图

4 非平稳信号的平稳化问题

心功能信号从总体看是准平稳的,但限于计算机内存,通常只取几个周期的数据来拟合模型,这就产生序列的平稳化问题.

我们采用差分方法化非平稳序列为平稳序列^[3],再对所得序列实施严格的平稳性检验^[6,7].平稳化以后的 ARIMA(n, d, m)模型以及平稳性检验的理论依据从略.

当样本量 N 足够大时,平稳性的统计假设定量检验是有效的.我们在实用中取 $N > 250$.

5 程序框图与建模结果^[6,7]

综合前面第 3 节和第 4 节所述,可得建模实现程序框图如图 3.

图 4(1)(a)、(2)(a)、(3)(a)分别为临床采集的 ECG、CAG、ACG 信号和用 ARMA 模型对相应的心功能信号建模结果重合在一起的图象;(1)(b)、(2)(b)、(3)(b)分别是它们两者间的误差曲线.误差分别为 5.8%、3.4%和 3.6%,可见模型精度是很高的.

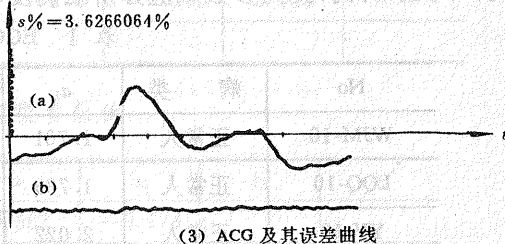
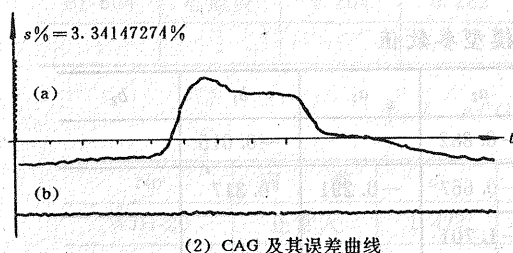
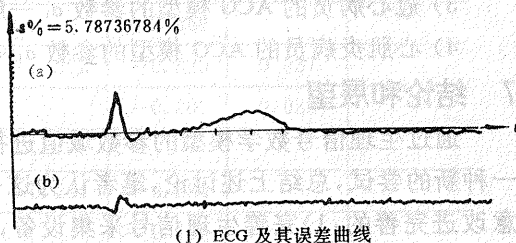


图 4 心功能信号及其模型误差

6 模型参数和病类辨识^[8]

根据上海新华医院对 48 例心血管病员的临床心功能信号记录,我们运用前述时间序列的系统建模方法,将各种心功能信号经计算机实现的各类 ARMA 模型,其模型的特征参数摘录如表 1~3 所示.表中列出的 a_1, a_2, a_3, a_4 是模型的自回归参数, b_1, b_2, b_3, b_4 是同一模型的自回归参数.经过对各类 ARMA 模型参数值的综合比较,我们发现各种心血管疾病的模型参数域值分布都存在一定的特异规律.现以高血压、冠心病和心肌炎为例,在与正常人作比较的基础上,借助参数域值对各种心功能信号进行病类辨识.

6.1 ECG 模型参数域值分析

- 1) 正常人 ECG 模型的参数 a_1 一般较大,大约在 1.7~2.0 之间.
- 2) 高血压病员的 ECG 模型的参数 a_1 一般偏小: $a_1 < 1.5$, $|a_2| > 0.5$.
- 3) 冠心病病员的 ECG 模型的参数 a_1 一般在 1.5~1.7 之间.
- 4) 心肌炎病员的 ECG 模型的参数 a_1 变化范围较宽,一般在 1.5~2.0 之间, $|a_2| > 0.7$,且大多出现 a_3 项.

6.2 CAG 模型参数域值分析

- 1) 正常人 CAG 模型的参数 a_1 一般较大, 大约在 1.4~1.7 之间.
- 2) 高血压病员的 CAG 模型的参数 a_1 较小, 一般在 1.1~1.3 之间, 参数 a_3 因不显著异于 0 而被删除, 而且参数 $a_2, a_4 \neq 0$.
- 3) 冠心病员的 CAG 模型的参数 $a_1 < 1.2$.
- 4) 心肌炎病员的 CAG 模型的参数 a_1 一般在 1.2~1.4 之间, 而 $a_2 \approx 0.2 \sim 0.3$, 且 $a_3, a_4 \neq 0$.

6.3 ACG 模型参数域值分析

- 1) 正常人 ACG 模型的参数 $a_1 > 1.3$, 且 $a_2, a_3 \neq 0$.
- 2) 高血压病员的 ACG 模型的参数 $a_1 < 1.2$.
- 3) 冠心病员的 ACG 模型的参数 a_1 一般在 1 上下.
- 4) 心肌炎病员的 ACG 模型的参数 $a_1 \approx 1.0 \sim 1.3$, 而 $a_2 \approx 0.2 \sim 0.35$.

7 结论和展望

通过生理信号数学模型的参数域值进行病类辨识是将现代科技引用到医疗诊断上的一种新的尝试, 总结上述讨论, 笔者认为这一设想是可以实现和值得推广的. 为此需要注意改进完善的: 1) 完善生理信号采集设备, 提高采集精度; 2) 改进建模软件, 提高运算速度; 3) 积累病例, 提高参数域值分辨准确度.

表 1 ECG 模型参数值

No	病 类	a_1	a_2	a_3	b_1	b_2
WJM-10	正常人	1.701	-0.882		-0.015	
LQQ-10	正常人	1.789	-0.667	-0.291	0.317	
YV-10	正常人	2.022	-1.701			
⋮	⋮					
B1-521	高血压	1.334	-0.502	0.016		0.077
Q1-521	高血压	1.494	-0.479		0.017	
F1-528	高血压	1.397	-0.456	0.366	0.091	
⋮	⋮					
K1-521	冠心病	1.504	-0.521			
H1-528	冠心病	1.668	-0.668			
D1-604	冠心病	1.598	-0.251	-0.773	0.032	
⋮	⋮					
G1-521	心肌炎	1.508	-0.842	0.250	0.076	-0.012
G1-528	心肌炎	1.857	-0.983			0.005
A1-604	心肌炎	1.998	-0.983	0.361		
⋮	⋮					

表 2 CAG 模型参数值

No	病 类	a_1	a_2	a_3	a_4	b_1	b_2
WJH-10	正常人	1.703	0.305		-0.178		
WJM-10	正常人	1.405	0.402	-0.228	-0.125	0.021	
YV-10	正常人	1.669	0.415	-0.256	-0.201		0.031
⋮	⋮						
B1-521	高血压	1.106	0.260		-0.150		
F1-521	高血压	1.311	0.267		-0.193		0.016
O1-521	高血压	1.251	0.265	-0.178	-0.230		
⋮	⋮						
J1-521	冠心病	0.949	0.343		-0.266		
K1-521	冠心病	1.069	0.257	-0.202	-0.325	0.031	
N1-521	冠心病	1.211	0.241		-0.201	0.089	
⋮	⋮						
L1-521	心肌炎	1.405	0.267	-0.291	-0.284		
M1-528	心肌炎	1.302	0.240		-0.186		
B1-604	心肌炎	1.204	0.282	-0.275	-0.211	0.025	
⋮	⋮						

表 3 ACG 模型参数值

No	病 类	a_1	a_2	a_3	b_1	b_2
WJH-20	正常人	1.477	0.367	-0.774		
WJM-20	正常人	1.633	0.324	-0.607		
C2-528	正常人	1.396	0.316	-0.694		
⋮	⋮					
B1-521	高血压	0.998	0.347			
F2-521	高血压	1.162	0.193			
O2-521	高血压	1.172	0.123		0.043	
⋮	⋮					
K2-521	冠心病	0.993	0.238			
H2-528	冠心病	0.982	0.244			
B2-528	冠心病	1.092	0.231			
⋮	⋮					
G2-528	心肌炎	1.248	0.237			
N2-528	心肌炎	1.172	0.353			
D2-528	心肌炎	0.996	0.296			
⋮	⋮					

参 考 文 献

- [1] Heuningen, R. Van, Goovaers H. D. and De vries F. R. . A low noise isolated amplifier system for electrophysiological measurements; basis considerations and design. Medical and Biological Engineering and Computing, Jan, 1984, 22(1): 77—85
- [2] Cochran Wilian G. . Sampling techniques, 3rd edition, chapter 2.
- [3] Åström K. J. and Wittenmark B. . Problem of identification and control, J. of Math. Analysis and Applic. , 1971, 34: 90—113
- [4] Box, G. E. P. and Jenkins G. M. . Time Series Analysis, forecasting and control. Wisconsin-Madison, 1973
- [5] Wu, S. M. and Pandit S. M. . Time series and System Analysis, model building and application.
- [6] Robinson Enders A. . Multichannel Time Series Analysis, with digital computer programs. second edition Goose Pond Press, 1983
- [7] 邓自立, 郭一新. 动态系统分析及其应用. 沈阳: 辽宁科学技术出版社, 1985.
- [8] Leaning, M. S. , Pullen, H. E. and Carson, E. R. etc. . Modelling a Complex Biological System; the human cardiovascular system. I. Methodology and Model Description, II. Model Validation, Reduction and Development. Trans. Inst. Meas. and Control (GB) 5(2): 71—98

Sampling, Model Building and Identification of a Category of Physiological Signals

LI Tianjian and LIU Yingzhi

(Department of Electrical Engineering, Tongji University • Shanghai, 200092, PRC)

Abstract: This paper intends to give a systematic investigation to the sampling, model building and identification of a category of physiological signals. As to model fitting of physiological signals, by applying the ARMA model scheme to the sampled cardiac function data series, we stressed on the realistic methodology, the parameter identification technique and stationarized method. By virtue of typical cardiac function signals-ECG, ACG, CAG, combining with microscopic analysis results, we have made characteristic identification microscopically to the typical diseases-hypertension, coronary heart disease and myocarditis, according to the parameter's domain values of ARMA models.

Key words: physiological signal; time series analysis; model building; parameter identification

本文作者简介

李恬鉴 1947年就读于复旦大学。1949年参军, 1954年复员。1958年毕业于浙江大学工业自动化专业。后到同济大学任教至今。现任电气工程系教授。主要从事建模理论及系统辨识的研究工作, 也涉足生物医学信息的获取及数据处理领域的研究。

柳英芝 女。1985年及1988年在同济大学电气工程系分别获得学士和硕士学位。现在主要从事工业控制机的开发及数据处理方面的研究。