

用于系统辨识的神经模糊推理系统的研究*

丛爽,高雪鹏

(中国科学技术大学自动化系,合肥 230027)

摘要:通过具有代表性的、针对非线性系统辨识的例子,对自适应神经模糊推理系统在实际应用中采用不同方式的实现过程进行了详细分析与综合。首先,采用搜索树对模型输入变量的选取原则进行了分析;然后,分别采用聚类分割法与平均分割法对模糊模型结构辨识方面的各自特点进行了比较。在对辨识模型的验证中,重点分析了训练样本数目、模糊模型参数数目与模糊模型泛化能力之间的关系,并通过不同模型测试性能结果的对比,说明并联模型泛化能力的复杂性。

关键词:神经模糊推理系统;搜索树;模糊聚类;泛化能力;串联并列模型;并联模型
中图分类号:TP273⁺.4 **文献标识码:**A

0 引言

自从20世纪80年代末期以来,随着模糊系统与神经网络的融合研究的不断深入,通过借鉴神经网络强大的学习能力,使得基于模糊推理系统的非线性系统辨识已成为一种有效的方法^[1]。虽然对模糊推理系统的逼近能力与模糊模型的参数辨识算法的研究已经相当透彻,但是在实际系统辨识的应用中仍然存在若干问题影响系统辨识的质量。

本文将以一个非线性电机系统^[2]为例,详细论述非线性系统辨识的全过程。其中,将重点分析几个重要问题,其中包括:模糊模型输入变量选取、模糊模型的结构辨识以及采用并联结构的模糊模型泛化能力的复杂性。

1 非线性系统模型的表示

通常人们用与线性系统中常用的自回归滑动平均模型(ARMA模型)相对应的非线性自回归滑动平均模型(NARMA模型)表示非线性系统,如(1)式所示^[3]。

$$y_k = f(y_{k-1}, y_{k-2}, \dots, y_{k-n}, u_{k-1}, \dots, u_{k-m}) \quad (1)$$

* 收稿日期:2002-04-06

基金项目:安徽省自然科学基金(03042301)资助项目

作者简介:丛爽,女,1961年4月生,教授,博导。研究方向:人工神经网络,模糊控制,运动控制,遗传算法等。

E-mail: scong@ustc.edu.cn

根据第 k 次以前的输出信号是用系统的实际输出还是模型输出,可以得到两类不同的辨识模型——串联并列模型和并联模型,见图 1. 串联并列模型实际上是一个普通的多层前向网络, 并联模型则由一个递归神经网络实现.

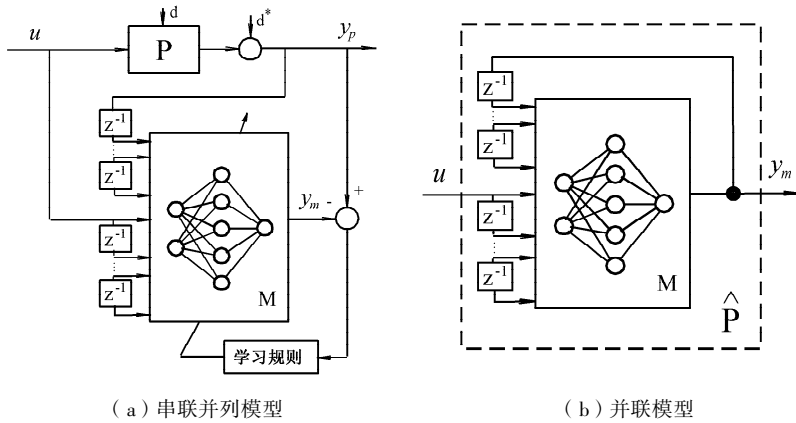


图 1 建立系统模型的神经网络结构图

Fig. 1 The structure of neural network in system identification

并联模型需要采用递归算法进行学习,一旦学习收敛,所得网络模型将与实际系统等价.然而,在实际训练中并联模型往往不能保证网络训练的收敛性.因此人们通常采用串联并列模型,这样不但可以使辨识过程简化,而且能保证算法的全局稳定性.对于串联并列模型,网络的输入/输出关系式可以表示为:

$$\hat{y}_k = \tilde{f}(y_{k-1}, y_{k-2}, \dots, y_{k-n}, u_{k-1}, \dots, u_{k-m}) \quad (2)$$

其中,整数 n 和 m 应在满足模型精度的要求下,作为辨识设计的一部分适当的选取.

假设经过训练后,串联并列模型的模型输出 $\hat{y}_k \approx y_k$, 在模型工作时,可以将模型输出代替实际输出反馈给模型输入,这样网络模型便完全独立于系统了,即转化为并联模型为:

$$\hat{y}_k = \tilde{f}(\hat{y}_{k-1}, \hat{y}_{k-2}, \dots, \hat{y}_{k-n}, u_{k-1}, \dots, u_{k-m}) \quad (3)$$

需要指出的是,实际系统中往往存在噪声,即使加上滤波器,仍然不能完全消除噪声的影响,使得 $\hat{y}_k \approx y_k$ 不能很好的满足.因此,当串联并列模型处于工作状态时,由于模型输出直接反馈回模型输入,从而产生测试误差.

2 模糊模型输入变量的选取

在采用模糊推理系统对非线性系统的辨识过程中,输入变量数目的选取,直接关系到辨识的精度,但在实际中常常又是一件相当困难的事:变量选的少,意味着模型简单,但精度不高;选的多,模型又复杂,所以本文把如何选取最佳输入变量作为研究的一个重点.

辨识中对于系统的激励信号经常采用的是多频率、多幅值的正弦复合信号,以便使得辨识出的模型具有较好的泛化能力.研究中所选取的输入信号为

$$u(k) = 650 \sin \frac{k\pi}{1000} + 300 \sin \frac{k\pi}{500} + 400 \sin \frac{3k\pi}{1000} + 240 \sin \frac{k\pi}{125} + 430 \sin \frac{3k\pi}{200} + 420 \sin \frac{3k\pi}{50} \quad (4)$$

通常采用一阶 Sugeno 型模糊推理系统,其模糊规则如下:

$$\text{if } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \text{ then } f_1 = a_1x + b_1y + c_1 \quad (5)$$

输入变量的选取,即是从一组候选的输入变量集中选取部分对输出的影响最显著的最优输入变量组合. 对于系统辨识,输入变量的选取就是确定被辨识系统的阶次和纯延时. 比较而言,采用搜索树来选取最优输入变量组合较为合适. 以一个简单的只有四个候选输入变量的系统为例,其搜索树^[4]如图 2 所示. 搜索树是分层次(level)的,其层数等于该层输入变量组合中的变量个数. 搜索树的搜索算法如下:计算当前层中每一个可能节点的性能指标,选择性能指标最优的节点,并沿此最优节点继续搜索. 若性能指标随着层数的增加逐渐变优,则一直搜索到最后一层,否则在性能指标变坏处停止搜索. 用此搜索算法选取最优输入变量组合可以大大减少需要处理的节点数目. 在图 2 中,假定每一层中性能指标最优的节点均为最左边的一个节点,搜索过程中遍历的节点由实线连接,而虚线连接的节点并不被遍历. 当候选输入变量共有 p 个时,搜索树中共有 $2^p - 1$ 个节点,而搜索算法只需计算至多 $p(p + 1)/2$ 个节点的性能指标.

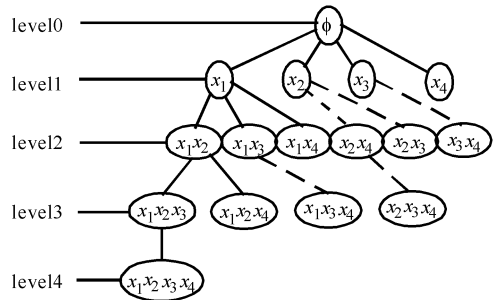


图 2 搜索树

Fig. 2 Searching tree

现以对非线性电机系统的建模为例,选择候选输入变量为 $u(k - 1)$, $u(k - 2)$, $u(k - 3)$, $y(k - 1)$, $y(k - 2)$, $y(k - 3)$. 采用基于 GMDH(Group Method of Data Handling) 方法的正则化准则^[4](Regularity Criterion, RC)作为性能指标,RC 定义如下式:

$$RC = \left[\sum_{i=1}^{n_A} \frac{(y_i^A - y_i^{AB})^2}{n_A} + \sum_{i=1}^{n_B} \frac{(y_i^B - y_i^{BA})^2}{n_B} \right] / 2 \quad (6)$$

其中: n_A 、 n_B 分别表示数据集 A 与 B 中的数据点个数; y_i^A 、 y_i^B 分别表示数据集 A 与 B 中的实际输出值; y_i^{AB} 用数据集 B 中数据训练所得辨识模型测试数据集 A 时,数据集 A 的模型输出; y_i^{BA} 用数据集 A 中数据训练所得辨识模型测试数据集 B 时,数据集 B 的模型输出; y_i^{AB} 、 y_i^{BA} 均为串联并列模型测试.

数据集 A 与 B 是由 (4) 式中的多频率正弦复合信号作为激励信号前后作用两次所产生的. 因此,数据集 A 与 B 中的 $u(k)$ 是相同的,而 $y(k)$ 由于不同时刻的噪声差异以及系统参数可能存在的微小的时变而不尽相同. 采用这种方法选择最优输入变量组合的基本思想是:在有相当程度的噪声影响下,具有正确(真实)结构的模型参数对于辨识参数的观测数据的变化最不敏感. 简言之,具有正确结构的模型的 RC 值最小.

表 1 为由输入变量组合不同情况下的 RC 值组成的搜索树,表中每一层中都有一个局部的最优输入变量组合(搜索树中某层的最优节点),已用黑体表明. 如果单从性能指标 RC 考虑,显然全局最优的输入变量组合是 $y(k - 1)$, $y(k - 2)$, $y(k - 3)$, $u(k - 3)$, $u(k - 2)$. 但是,从表 1 中还可以看出,第 3 层以后 RC 的变化已经不明显,而输入变量的增加会大大提高模型的复杂性. 所以,综合精确性与简单性的考虑,我们认为对此建模系统,选择三个输入变量 $y(k - 1)$, $u(k - 3)$, $y(k - 2)$ 作为最优输入变量组合是较合适的.

表 1 利用搜索树选择最优输入变量组合

Tab. 1 The Selection of Optimal Input Variable Group Using Searching Tree

树中级别	输入变量组合	RC
第 1 层	$u(k-1)$	5.985 9e+004
	$u(k-2)$	5.752 2e+004
	$u(k-3)$	5.527 5e+004
	$y(k-1)$	108.901 0
	$y(k-2)$	249.135 5
	$y(k-3)$	477.700 3
第 2 层	$y(k-1), u(k-1)$	65.576 4
	$y(k-1), u(k-2)$	64.082 8
	$y(k-1), u(k-3)$	64.034 5
	$y(k-1), y(k-2)$	114.505 4
	$y(k-1), y(k-3)$	97.639 6
	$y(k-1), u(k-4)$ ^{注1}	97.099 5
第 3 层	$y(k-1), u(k-3), u(k-1)$	64.946 1
	$y(k-1), u(k-3), u(k-2)$	65.339 3
	$y(k-1), u(k-3), y(k-2)$	51.846 0
	$y(k-1), u(k-3), y(k-3)$	58.380 7
第 4 层	$y(k-1), y(k-2), u(k-3), u(k-1)$	52.480 0
	$y(k-1), y(k-2), u(k-3), u(k-2)$	52.540 1
	$y(k-1), y(k-2), u(k-3), y(k-3)$	51.006 8
第 5 层	$y(k-1), y(k-2), y(k-3), u(k-3), u(k-1)$	49.471 4
	$y(k-1), y(k-2), y(k-3), u(k-3), u(k-2)$	48.989 3
第 6 层	$u(k-1), u(k-2), u(k-3), y(k-1), y(k-2), y(k-3)$	54.809 0

注 1: 加入候选变量外的输入 $u(k-4)$ 是为了防止候选变量阶次过低, 事实证明取 $u(k-3)$ 已足够

3 模型结构辨识

在模型输入变量确定之后, 模糊模型的结构辨识至关重要. 模型结构辨识, 即在当前辨识数据下, 确定模糊模型的最佳模型结构, 包括模糊规则数目的确定及输入空间的划分.

3.1 基于模糊聚类法的最优模型结构确定

聚类方法^[5]是用来调整各输入变量的聚类半径, 令输入变量的聚类半径为 r_u , 输出变量的聚类半径为 r_y . 首先可以取 $r_u = r_y = 0.5$, 然后使用二分法逐步减小 r_u 和 r_y , 当聚类半径接近最优解时, 取消 $r_u = r_y$ 的约束, 分别对 r_u 和 r_y 进行微调, 直到达到一个近最优的解下的模糊规则数.

仍然采用(4)式中的多频率复合正弦信号作为训练输入信号. 并采用另外两个测试信号, 第一个测试信号仍为多频率复合正弦信号, 这一测试信号可以考察所得模糊模型对不同频率与幅值信号的泛化能力; 第二个测试信号是一个单频率的大幅值的正弦信号, 包含了实

际建模系统所包含的死区和饱和非线性特性. 这两个测试信号分别由式(7)和(8)表示, 称为测试集 1 和测试集 2.

$$u_2(k) = 250\sin\frac{k\pi}{1000} + 400\sin\frac{3k\pi}{500} + 300\sin\frac{3k\pi}{200} + 240\sin\frac{k\pi}{40} + 530\sin\frac{3k\pi}{100} + 220\sin\frac{k\pi}{20} \quad (7)$$

$$u_3(k) = 800\sin\frac{k\pi}{1000} \quad (8)$$

从表 2 列出的各个不同模型对两个测试信号的测试误差中不难看出, 采用串联并列模型的测试误差随着模糊模型复杂性(模糊规则与参数的增加)的提高而减小; 但对于采用并联模型的测试误差, 这一结论不成立. 只有当模型具有适度的复杂性时, 才能获得最佳辨识结果. 当模型的规则数和参数数目过少时, 则不足以表达样本集中大量样本所包含的输入/输出关系, 即“欠拟合”, 使得模型的精度降低; 而当模型的规则数和参数数目过多, 虽然可以提高训练精度, 但会造成“过拟合”, 同样会减弱模型的泛化能力, 从而降低模型的精度. 只有当模型具有适度的复杂性时, 即适量的规则数和参数数目时, 才能获得最佳辨识结果.

表 2 采用聚类方法对最优模型结构的确定过程

Tab. 2 Procedure of determining optimal model structure with clustering

聚类半径 (r_u, r_y)	规则 总数	参数 总数	训练集 误差	测试集 1 误差		测试集 2 误差	
				串并模型	并联模型	串并模型	并联模型
0.5, 0.5	4	40	12.160 7	32.792 8	2.363 7e+003	10.981 7	3.327 8e+003
0.375, 0.375	5	50	11.288 6	30.347 5	793.389 8	10.373 4	4.877 4e+003
0.25, 0.25	9	90	9.976 3	28.804 7	537.921 2	9.031 3	1.071 6e+003
0.187 5, 0.187 5	11	110	9.969 4	28.470 3	614.371 1	9.169 5	1.680 5e+003
0.125, 0.125	22	220	9.134 3	27.174 8	255.548 5	8.426 2	787.639 7
0.093 25, 0.093 25	28	280	9.099 4	27.940 8	224.681 4	8.402 6	426.984 9
0.062 5, 0.062 5	46	460	9.073 0	33.983 0	334.787 2	8.370 7	772.003 4
0.031 25, 0.031 25	117	1 170	7.143 2	38.860 0	643.913 9	19.889 3	2.578 5e+004
0.08, 0.08	35	350	8.948 3	27.828 8	290.025 2	8.313 1	650.347 1
0.08, 0.085	34	340	9.012 4	33.798 3	265.357 9	8.384 7	399.191 0
0.08, 0.09	34	340	8.981 6	34.005 5	294.305 4	8.398 2	349.914 7
0.08, 0.10	29	290	9.057 8	27.733 9	276.517 3	8.467 7	613.817 6
0.085, 0.10	29	290	9.065 2	27.672 0	272.252 5	8.427 3	532.040 5

一般来说, 验证模型的精确性主要以并联模型为准, 从表 2 可知, 并联模型对聚类半径的变化较敏感, 因此, 在本文所给出的例子中, 在[0.093 25, 0.093 25]附近对多组不同的聚类半径进行测试, 同时选择最好的模型, 用黑体标明, 并在下面的参数辨识中, 将采用聚类半径分别为[0.093 25, 0.093 25]及[0.08, 0.09]的模型作为最优模型结构的后选模型, 分别称为模型 1 与模型 2.

3.2 基于平均分割法的最优模型结构确定

平均分割法将输入空间的每一维按比例分割成宽度相等的区域. 一方面对此非线性电机系统, 显然 n_i 至少应当划分为 3 个区域; 另一方面当 $n_i = 5$ 时, 则模糊模型具有 125 条模糊规则, 530 个参数. 这个模型已经足够复杂了, 如果再增加模糊规则, 则计算的复杂性过高, 所需的存储量也很大, 而且参数过多, 也会造成“过拟合”. 所以对于本文所给系统, 限制搜索空间为 $3 \leq n_i \leq 5$. 表 3 给出了采用平均分割方法对最优模型结构的确定过程.

在实际应用中, 应以测试模型对于典型测试信号在并联模型中的均方误差来选择较好的模型结构. 表 3 中用黑体标明的两个模型是最优模型结构的后选者, 其输入空间划分为 $3 * 4 * 4$ 和 $3 * 5 * 4$, 分别称为模型 3 与模型 4.

表 3 采用平均分割方法对最优模型结构的确定过程

Tab. 3 Procedure of determining optimal model structure with means-clustering

输入空间划分	规则 总数	参数 总数	训练集 误差	测试集 1 误差		测试集 2 误差	
				串并模型	并联模型	串并模型	并联模型
3 * 3 * 3	27	126	9.156 3	29.012 0	1.773 6e+003	8.426 5	410.110 7
3 * 3 * 4	36	164	8.914 2	29.132 0	589.447 1	8.326 1	256.226 5
3 * 4 * 3	36	164	8.898 8	29.995 2	1 398.6	8.318 4	225.005 4
4 * 3 * 3	36	164	9.018 1	32.229 5	3.761 2e+003	8.355 6	327.827 4
3 * 4 * 4	48	214	8.832 4	30.065 4	525.613 1	8.356 6	223.399 0
3 * 4 * 5	60	264	8.661 0	33.875 9	1.375 6e+003	8.285 8	228.019 0
3 * 5 * 4	60	264	8.632 6	33.663 4	493.706 9	8.358 4	414.560 5
4 * 4 * 4	64	280	8.626 8	48.523 5	4.178 5e+003	8.411 6	328.042 5
3 * 5 * 5	75	326	8.563 1	37.175 1	4.874 6e+003	8.298 5	207.757 6
4 * 5 * 5	100	428	8.332 3	49.339 5	5.924 0e+003	8.405 0	877.094 1
5 * 5 * 5	125	530	8.171 8	155.553 3	8.147 8e+003	8.420 8	898.999 2

3.3 聚类分割法和平均分割法的比较

聚类法根据训练集中样本在输入/输出空间中的分布, 产生相应的分类, 使得输入/输出空间中距离较近的点归为一类, 从而产生一条规则, 这种方法在确定模型的初始结构时实际上用到了输入/输出样本集所包含的先验知识. 而平均分割法完全不考虑样本的分布, 直接确定模型的初始结构. 聚类法的最大优势在于其可以避免“维数灾难”, 当模型输入变量较多时, 比平均分割法所需的规则数少的多.

比较表 2 和表 3 可以知道, 在相近的辨识精度下, 采用平均分割法产生的模型的复杂性要低于采用聚类分割法产生的模型. 这是因为, 采用聚类算法, 各输入变量的隶属函数数目都等于规则的数目, 所以聚类算法产生的模型输入的每一维上的隶属函数数目远大于采用平均分割所产生的隶属函数数目, 相应的聚类算法产生的模型的每一维上的各个隶属函数的宽度远小于平均分割所产生的模型的每一维上的各个隶属函数的宽度.

一般来说, 平均分割法对输入空间的每一维的划分数目不会多于 5 个, 所以当输入空间维数较低时, 平均分割法的参数数目远小于聚类分割法; 当输入空间维数较高时, 由于平均分割法的“维数爆炸”, 会产生非常多的规则数, 从而参数数目也会急剧膨胀. 一般而言, 当模型输入变量个数少于 3 个, 使用平均分割法较好; 模型输入变量个数多于 3 个, 使用聚类

分割法较好;模型输入变量个数等于3个,两种方法均可使用。

3.4 训练样本数目、模糊模型参数数目与模型泛化能力的经验关系

训练样本数目、模糊模型参数数目与模型泛化能力三者之间存在必然联系。由于关系复杂,在此我们只能进行定性的关系分析。

当模型参数数目不足,则不能够完全表达训练样本中包含的系统输入/输出关系,此时对测试集样本会产生较大的误差,即“欠拟合”,减弱模型的“泛化”能力;模型参数过多,过分地学习了训练集中的样本,使得模型描述输入/输出关系过于复杂,使得样本点之间的曲面非线性过强,从而造成对测试集样本的误差过大,及“过拟合”,同样会使“泛化”能力减弱。一般而言,在训练样本中存在噪声的情况下,训练样本数目应选在模糊模型参数总数目的5倍以上和10倍以下;当不存在噪声时,训练样本数目应选在模糊模型参数总数目的3倍以上和6倍以下。

对于本文所采用的建模系统,训练集与测试集样本数均为2000个,且属于有噪声情况,从表2和表3可以看出四个模型的参数总数分别为:214、264、280和340,其模型结构均符合以上描述的关系。

4 模型参数辨识

在以上分析中,利用搜索树算法获得了最佳的输入变量组合,又通过结构辨识获得了几个最佳的初始结构,下面分别针对这几个结构的模型分别进行参数调整,以获得优化的模型。本文用来确定所有参数以及最终训练神经模糊网络所采用的算法全部采用的是混合学习算法^[6],即对模糊规则的条件参数采用反向传播算法,结论参数采用最小二乘法反复进行迭代,以获得优化的模型参数。一般而言,用模糊模型对实际系统进行辨识时,只需训练很少的几次,即可获得较好的效果。这一特点是模糊模型辨识速度快于神经网络的重要原因。

对前面通过两种结构辨识法获得的四个次最优的初始模型进行多次参数辨识,分别训练100次,以获得优化的模型参数。其结果如表4所示,其中的误差指标均为均方误差,测试误差均采用并联模型。

表4 参数优化前后的结果对比

Tab.4 Comparison of results before and after optimization parameters

模型结构	训练时间/s	参数优化前的训练误差	参数优化后的训练误差	参数优化前的测试误差		参数优化后的测试误差	
				测试集1	测试集2	测试集1	测试集2
模型1	1 357.7	9.099 4	9.090 3	224. 681 4	426.984 9	226.651 8	426. 485 1
模型2	1 982.4	8.981 6	8.969 9	294.305 4	349.914 7	293. 122 6	340. 893 9
模型3	2 988.2	8.832 4	8.807 1	525. 613 1	223. 399 0	1 458.3	229.088 0
模型4	4 940.7	8.632 6	8.626 8	493.706 9	414.560 5	461. 968 0	408. 580 5

表4中可以看出,对绝大多数情况,在参数优化后,并联模型的测试误差均会有少许降低,这正说明需要通过多个模型来确定最终模型。综合考虑测试集1和测试集2的误差,从表4中可知模型2在参数优化后具有最优的辨识结果。

综上所述,在采用混合学习算法来设计训练自适应神经模糊模型辨识系统时,参数优化是一个相对次要的步骤,且未必可以进一步提高辨识精度。一般只需训练很少的几次,即可

获得较好的效果. 而在对其模型输入变量数目辨识以及模糊模型的结构辨识方面应花费较多的精力并采用一定的手段来保证模型的准确性与辨识的精度.

5 总结

总之, 实际非线性系统的辨识不是一个简单的问题. 一个好的辨识效果取决于模型输入变量的选择、模型结构辨识与参数优化. 对于采用混合学习算法来设计训练的自适应神经模糊系统辨识非线性系统, 这些步骤中的每一步都不能只用简单的理论公式描述, 对一个具体的系统, 只有通过以上介绍的方法反复做实验, 可获得一个最佳的模型.

参 考 文 献

- [1] Shin-ichi Horikawa, Takeshi Furuhashi & Yoshiki Uchikawa. On fuzzy modeling using fuzzy neural networks with the back-propagation algorithms[J]. IEEE Trans. on Neural Networks, 1992, 3(5):801-806.
- [2] 丛爽, 高雪鹏, 魏衡华. 非线性直流电机仿真模型系统的建立[J]. 系统仿真学报, 2001, 13(1):25-27.
- [3] 丛爽. 典型人工神经网络的结构、功能及其在智能系统中的应用[J]. 信息与控制, 2001, 30(2):97-103.
- [4] Michio Sugeno, Takahiro Yasukawa. A fuzzy-logic-based approach to qualitative modeling [J]. IEEE Trans. on Fuzzy Systems, 1993, 1(1).
- [5] Suleiman Barada, Harpreet Singh. Generating optimal adaptive fuzzy-neural models of dynamical systems with applications to control[J]. IEEE Trans on System, Man, and Cybernetics, 1998, 28(3):371-391.
- [6] Jyh-Shing Roger Jang and Chuen-Tsai Sun. Neuro-fuzzy modeling and control[J]. Proceedings of the IEEE, 1995, 83(3):378-406.

Study on Neural-fuzzy Inference System Used in the System Identification

CONG Shuang, GAO Xue-peng

(Dept. of Automation, USTC, Hefei 230027, China)

Abstract: This paper analyses in detail and synthesizes the different realization methods of adaptive neural-fuzzy inference system by means of the typical example of nonlinear system identification. First, the selection principle for the fuzzy model's number of input variables is analyzed by making use of the searching tree. Then, different features in the aspects of structure identification of fuzzy model using the clustering and means-clustering are compared. The relationships among the numbers of samples, fuzzy model parameters and the ability of generalization of the fuzzy model are analyzed, and the comparison of the test results of models shows the complexity of generalization ability of parallel model.

Key words: neural fuzzy inference system; searching tree; fuzzy clustering; generalization ability; serial-parallel model; parallel model