

# 一种数据驱动的Ⅱ型T-S模糊建模方法

廖倩芳, 李 柠, 李少远

(上海交通大学 电子信息与电气工程学院, 上海 200240)

**摘 要:** 现场采集的数据不可避免地包含一些诸如噪声干扰之类的不确定性, 由数据驱动建立的模型需要具备较强的处理不确定因素影响的能力. 在以往文献的Ⅰ型T-S模糊建模方法的基础上, 提出了一种基于数据驱动的Ⅱ型T-S模糊建模方法. 其过程是通过分析采集的数据样本计算得到不确定因素的影响程度, 在Ⅰ型T-S模糊模型的基础上, 前件参数上采用Ⅱ型的模糊集来代替Ⅰ型的模糊集, 后件参数上则采用Ⅰ型模糊集来代替数值, 由此拓展得到Ⅱ型T-S模糊模型. 最后通过pH中和反应过程对所提出的方法进行仿真验证. 仿真结果表明, 该方法建立的模型能更好地处理不确定因素的影响, 取得更高的准确度.

**关键词:** Ⅱ型模糊; T-S模糊模型; 数据驱动; pH中和反应过程

**中图分类号:** TP273 **文献标识码:** A **文章编号:** 1673-4785(2009)04-0303-06

## A Type-Ⅱ T-S fuzzy modeling method for data-driven approaches

LIAO Qian-fang, LI Ning, LI Shao-yuan

(School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

**Abstract:** Data collected from the field inevitably contains uncertainties such as noise or other disturbances; mathematical models established with data-driven approach must possess strong capability to deal with the influence of uncertainties. Following analysis of current methods for type-I Takagi-Sugeno (T-S) fuzzy modeling, a method suitable for type-Ⅱ T-S fuzzy modeling was proposed. In the data driven modeling process, the influence of the degree of uncertainty was determined by analysis of the collected data. On the basis of the type-I fuzzy model, for antecedent parameters, we employed the fuzzy set of the type-Ⅱ fuzzy model to replace the counterpart from the type-I model. But for consequent parameters, we took type-I fuzzy sets to replace crisp numbers. This produced an improved type-Ⅱ T-S fuzzy model. Finally, a pH neutralization process was taken as an example to verify the proposed mathematical model. Simulation results showed that this method can handle the influence of uncertainties better and achieves higher accuracy.

**Keywords:** Type-Ⅱ fuzzy; T-S fuzzy model; data-driven; pH neutralization process

目前许多工业过程中,不可避免地存在不确定性因素的影响. 从工业过程采集到的数据中,也不可避免地包括各种噪声、干扰甚至测量时产生的误差,导致与真实值不相符. 如何从这些包含了不确定性的数据出发,建立基于数据驱动并且能完整体现实际过程特性的动态模型,是目前研究的一个热点问题.

模糊模型常用于对非线性较强的工业过程建模,目前Ⅰ型模糊模型使用比较普遍. Zadeh于1975年提出了Ⅱ型模糊集合论<sup>[1]</sup>,与之前提出的Ⅰ型的不同之处在于,Ⅱ型模糊理论中元素的隶属度由一阶隶属度(primary membership)和二阶隶属度(secondary membership)组成,一阶隶属度表示元素的隶属度,二阶隶属度就表示一阶隶属度的隶属度. 采用Ⅱ型模糊理论来描述带噪声的数据,得到的隶属度不是一个确定的值,而是一个Ⅰ型模糊集,它可以包含可能存在的不确定性,能提高模型处理干扰的能

收稿日期:2008-10-24.

基金项目:国家自然科学基金资助项目(60604018);上海自然科学基金资助项目(06ZR14044).

通信作者:李 柠. E-mail: ning\_li@sjtu.edu.cn.



果干扰程度小,  $\delta$  就相应取小一些的值. 如果  $\delta$  取得过小, 那么模型的精度就接近于Ⅰ型模糊模型的精度, 无法体现Ⅱ型模糊模型的优势; 取得过大, 无法反映实际的干扰程度, 会致使模型的精度下降. 选完样本后, 找出每组样本中隶属度最大的一个和隶属度最小的一个, 分别记为  $\max_l(\mu)$  和  $\min_l(\mu)$ ,  $l = 1, \dots, k$ , 令

$$|\Delta\mu^i| = \frac{1}{k} \sum_{l=1}^k (\max_l(\mu) - \min_l(\mu))/2. \quad (1)$$

由式(1)得到每个类的隶属度偏差. 最后令  $\mu_l^i = \mu^i - |\Delta\mu^i|$ ,  $\mu_r^i = \mu^i + |\Delta\mu^i|$  (当  $\mu_l^i = \mu^i - |\Delta\mu^i| < 0$  时, 令  $\mu_l^i = 0$ ; 当  $\mu_r^i = \mu^i + |\Delta\mu^i| > 1$  时, 令  $\mu_r^i = 1$ ). 由此得到一阶隶属度区间集两边的端点, 表示为  $[\mu_l^i, \mu_r^i]$ . 这样就得到了区间Ⅱ型模糊集, 表示为  $\tilde{Z}_i$ ,  $i = 1, 2, \dots, c$ . Ⅱ型T-S模糊模型规则的前件参数就得到了确定.

### 1.2 Ⅱ型T-S模糊模型后件的辨识

后件参数辨识是确定后件多项式系数, 本文中的Ⅱ型T-S模糊模型后件中每个系数都是一个区间Ⅰ型集, 需要确定每个系数两端的端点. 先借助1.1节中G-K算法产生的隶属度矩阵, 把每个数据  $z_j = [\varphi_j \ y_j]$  归属到它的隶属度最大的那个类里面, 将样本集划分成  $c$  个类. 通过对数据样本的分析, 来估算得到在噪声干扰的影响下系统输出值的变化范围. 首先, 在每个类中选取  $k$  组数据样本, 每组的样本数量大于1, 选取的原则是令每组里的样本  $z_j = [\varphi_j \ y_j]$  中的输入部分  $\varphi_j$  的之间的差距在  $\delta$  以内, 即  $\|\varphi_i - \varphi_j\| \leq \delta$ ,  $\varphi_i$  和  $\varphi_j$  为同一组中不同的数据样本的输入部分,  $\delta$  的值可根据实际情况凭经验来选取, 选取的原则类似于前件的辨识中的选取原则. 然后找出每组样本中输出部分的最大的一个和最小的一个  $y_j$  值, 分别记为  $\max_l(y)$  和  $\min_l(y)$ ,  $l = 1, \dots, k$ , 令

$$|\Delta y^i| = \max\{(\max_l(y) - \min_l(y))/2\}, \quad l = 1, \dots, k, \quad (2)$$

然后令

$$|\Delta y| = \max\{|\Delta y^i|, i = 1, \dots, c\}. \quad (3)$$

通过式(2)和(3)来得到输出部分的最大偏差  $|\Delta y|$ , 然后用随机数发生器在  $[0, |\Delta y|]$  之间取2个数值  $\Delta y1_j$  和  $\Delta y2_j$ , 令  $y_{ij} = y_j + \Delta y1_j$ ,  $y_{ij} = y_j - \Delta y2_j$ ,  $j = 1, 2, \dots, N$ . 这样便由每个数据  $z_j = [\varphi_j \ y_j]$  拓展得到另外2组数据:  $z_{ij} = [\varphi_j \ y_{ij}]$  和  $z_{ij} = [\varphi_j \ y_{ij}]$ . 以类为单位, 将拓展得到的2组数据分别通过最小二乘法

辨识, 得到线性多项式, 如下:

$$\begin{aligned} y_l^i &= a_{l0}^i + a_{l1}^i \varphi_j(1) + \dots + a_{lp}^i \varphi_j(p), \\ y_r^i &= a_{r0}^i + a_{r1}^i \varphi_j(1) + \dots + a_{rp}^i \varphi_j(p), \\ i &= 1, 2, \dots, c. \end{aligned}$$

将上式合并, 得到

$$[y_l^i, y_r^i] = [a_{l0}^i, a_{r0}^i] + [a_{l1}^i, a_{r1}^i] \varphi_j(1) + \dots + [a_{lp}^i, a_{rp}^i] \varphi_j(p), \quad i = 1, 2, \dots, c.$$

可得到后件参数区间集  $\tilde{a}_j^i (j = 1, \dots, p)$  的两端的端点值:  $\tilde{a}_j^i = [a_{ij}^i, a_{rj}^i]$ ,  $(j = 1, \dots, p)$ , 这样通过确定输出值区间  $[y_{ij}, y_{rj}]$  来进一步确定线性多项式的系数区间  $\tilde{a}_j^i$ . 最终得到的规则形式为

$$\begin{aligned} R^i: & \text{ if } [\varphi_j \ y_j] \text{ is } \tilde{Z}_i \\ & \text{ then } \tilde{y}^i = [y_l^i, y_r^i] = \\ & [a_{l0}^i, a_{r0}^i] + [a_{l1}^i, a_{r1}^i] \varphi_j(1) + \dots + \\ & [a_{lp}^i, a_{rp}^i] \varphi_j(p), \\ & i = 1, 2, \dots, c. \end{aligned}$$

最后模型的输出采用文献[11]提出的形式:

$$y = \left( \frac{\sum_{i=1}^c \mu_l^i y_l^i}{\sum_{i=1}^c \mu_l^i} + \frac{\sum_{i=1}^c \mu_r^i y_r^i}{\sum_{i=1}^c \mu_r^i} \right) / 2.$$

### 1.3 Ⅱ型T-S模糊建模的步骤

区间Ⅱ型T-S模糊建模可以分为以下步骤:

1) 采集系统输入输出的数据样本, 通过G-K算法将样本归类;

2) 由1.1中的方法确定模型中每条规则前件的隶属度偏差  $|\Delta\mu^i|$ ,  $i = 1, 2, \dots, c$ , 得到模型规则前件中的Ⅱ型模糊的隶属度区间集  $[\mu_l^i, \mu_r^i]$ ;

3) 由1.2中的方法确定样本输出的最大偏差  $|\Delta y|$ , 然后用随机数发生器在  $[0, |\Delta y|]$  之间取2个数值  $\Delta y1_j$  和  $\Delta y2_j$ , 令  $y_{ij} = y_j + \Delta y1_j$ ,  $y_{ij} = y_j - \Delta y2_j$ ,  $j = 1, 2, \dots, N$ . 将采样得到的数据样本  $z_j = [\varphi_j \ y_j]$  拓展得到另外2组数据:  $z_{ij} = [\varphi_j \ y_{ij}]$  和  $z_{ij} = [\varphi_j \ y_{ij}]$ . 然后以类为单位, 通过最小二乘法来分别得到Ⅱ型T-S模糊模型规则的后件参数  $a_{ij}^i, a_{rj}^i$ ,  $i = 1, \dots, c, j = 1, \dots, p$ ;

4) 最终得到Ⅱ型T-S模糊模型:

$$\begin{aligned} R^i: & \text{ if } [\varphi_j \ y_j] \text{ is } \tilde{Z}_i \\ & \text{ then } \\ \tilde{y}^i &= [y_l^i, y_r^i] = \tilde{a}_0^i + \tilde{a}_1^i \varphi_j(1) + \dots + \tilde{a}_p^i \varphi_j(p), \\ & i = 1, 2, \dots, c. \end{aligned}$$

建模的流程图如下:

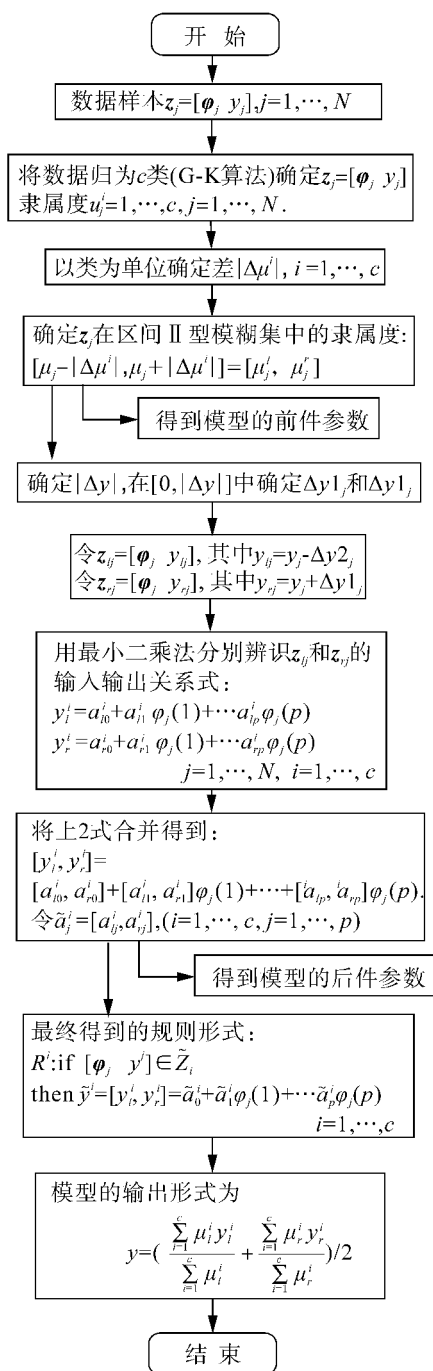


图1 流程图

Fig. 1 Flow chart

相比于文献[7]中的建模方法,该建模方法中参数是一次确定的,不需要反复调整,并且该方法简便易行,计算步骤和计算量较少。以下通过仿真来验证模型的精度。

## 2 pH 中和过程仿真

pH 中和过程是复杂工艺过程,广泛存在于化工和污水处理等工艺之中。它具有严重非线性,同时不可避免地存在各种干扰。

pH 中和过程是通过反应物(酸液/碱液)对流

入物(碱液/酸液)进行中和,并检测流出物来获取反应容器中的 pH 值。实际的 pH 过程由于各种物料成分或者容器的不同会有复杂多变的性质,但可以用数学动态模型来描述 pH 过程在连续搅拌式反应器(CSTR)中的主要特性<sup>[12]</sup>。

假设 CSTR 中各处等温而且反应物料均完全混合,由文献[12]中给出的 pH 动态数学模型,设流入容器中的物料有 2 种,流入物(酸液)和反应物(碱液)。其中,酸液的流量为  $F_a$ ,所含成分的浓度为  $C_a$ ,碱液流量为  $F_b$ ,所含成分浓度为  $C_b$ ,那么 CSTR 中的动态模型可以描述为

$$V \frac{dw_a}{dt} = F_a \times C_a - (F_a + F_b) \times w_a, \quad (4)$$

$$V \frac{dw_b}{dt} = F_b \times C_b - (F_a + F_b) \times w_b. \quad (5)$$

式中:  $V$  为容器的体积,  $w_a$  为容器中酸液的浓度,  $w_b$  为容器中碱液的浓度。式(4)和(5)是反映了容器内的酸碱浓度随着酸碱液流量的变化而变化的过程。由  $w_a$  和  $w_b$  可以得到容器流出物的 pH 值,考虑 HAC(弱酸)和 NaOH(强碱)的反应过程,其中和滴定曲线方程为

$$w_b + 10^{-\text{pH}} - 10^{\text{pH}-14} - \frac{w_a}{1 + 10^{pK_a - \text{pH}}} = 0. \quad (6)$$

式中:  $pK_a = -\lg K_a$ ,  $K_a$  为弱酸 HAC 的电离常数,

$$K_a = \frac{[\text{AC}^-][\text{H}^+]}{\text{HAC}} = 1.76 \times 10^{-5}.$$

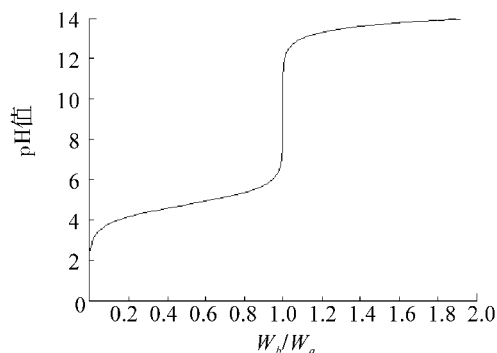


图2 弱酸强碱中和滴定曲线

Fig. 2 Neutralization titration of weak acid and strong base

考虑式(4)~(6)描述的弱酸强碱中和反应的过程,具体参数见文献[12]。假设给定酸液流量  $F_a = 81$ ,由碱液流量  $F_b$  的变化来影响输出的 pH 值  $y_{\text{pH}}$ 。设 T-S 模型  $k$  时刻的输出关系式包含  $(k-1)$  和  $(k-2)$  时刻的共 4 个输入量,表达形式为

$$\hat{y}_{\text{pH}} = \psi(F_b(k-1), y_{\text{pH}}(k-1), F_b(k-2), y_{\text{pH}}(k-2)).$$

式中:  $\hat{y}_{\text{pH}}$  和  $y_{\text{pH}}$  分别为模型输出和实际输出。

在中和反应过程中加入随机不确定性后,测量获得 300 组数据样本,形式为

$$\{F_b(k-1), y_{\text{pH}}(k-1), F_b(k-2), y_{\text{pH}}(k-2), y_{\text{pH}}(k)\},$$

采用本文中的方法,可以由样本拓展得到 300 组

$$\{F_b(k-1), y_{\text{pH}}(k-1), F_b(k-2), y_{\text{pH}}(k-2), y_{\text{pH}}(k)\}$$

数据和 300 组

$$\{F_b(k-1), y_{\text{pH}}(k-1), F_b(k-2), y_{\text{pH}}(k-2), y_{\text{pH}}(k)\}$$

数据来辨识 II 型 T-S 模糊模型。

从以上得到的数据出发,采用提出的方法,设定模型的规则数为 6,规则如下:

设

$$\varphi(1) = F_b(k-1), \varphi(2) = y_{\text{pH}}(k-1),$$

$$\varphi(3) = F_b(k-1), \varphi(4) = y_{\text{pH}}(k-2).$$

$R^1$ : if  $[\varphi \ y]$  is  $\tilde{Z}_1$

then

$$\begin{aligned} \tilde{y}_{\text{pH}}^1 = & [-5.2302, -3.9410] + \\ & [0.0102, 0.0093]\varphi(1) + \\ & [1.0249, 1.0633]\varphi(2) + \\ & [-0.0011, -0.0013]\varphi(3) + \\ & [-0.0038, 0.0008]\varphi(4). \end{aligned}$$

$R^2$ : if  $[\varphi \ y]$  is  $\tilde{Z}_2$

then

$$\begin{aligned} \tilde{y}_{\text{pH}}^2 = & [-7.7676, -5.4739] + \\ & [0.0107, 0.0120]\varphi(1) + \\ & [0.6702, 0.0632]\varphi(2) + \\ & [0.0025, 0.0068]\varphi(3) + \\ & [0.3494, 0.5754]\varphi(4). \end{aligned}$$

$R^3$ : if  $[\varphi \ y]$  is  $\tilde{Z}_3$

then

$$\begin{aligned} \tilde{y}_{\text{pH}}^3 = & [-3.0864, -4.6647] + \\ & [0.0179, 0.2120]\varphi(1) + \\ & [0.0785, 0.0307]\varphi(2) + \\ & [0.0047, 0.0067]\varphi(3) + \\ & [0.0319, 0.0432]\varphi(4). \end{aligned}$$

$R^4$ : if  $[\varphi \ y]$  is  $\tilde{Z}_4$

then

$$\begin{aligned} \tilde{y}_{\text{pH}}^4 = & [-11.4480, -19.2874] + \\ & [0.0202, 0.0359]\varphi(1) + \\ & [0.5230, 0.7959]\varphi(2) + \\ & [0.0095, 0.0067]\varphi(3) + \\ & [0.1724, 0.1462]\varphi(4). \end{aligned}$$

$R^5$ : if  $[\varphi \ y]$  is  $\tilde{Z}_5$

then

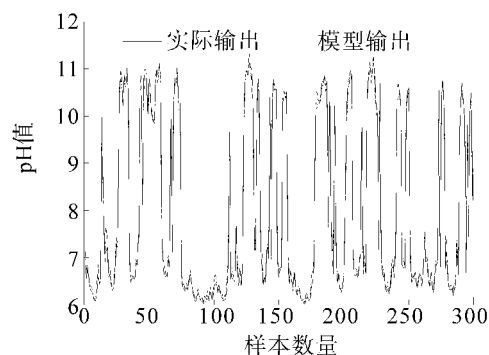
$$\begin{aligned} \tilde{y}_{\text{pH}}^5 = & [-26.1793, -18.8988] + \\ & [0.0349, 0.0296]\varphi(1) + \\ & [0.1900, 0.1663]\varphi(2) + \\ & [0.0195, 0.0126]\varphi(3) + \\ & [0.3926, 0.3876]\varphi(4). \end{aligned}$$

$R^6$ : if  $[\varphi \ y]$  is  $\tilde{Z}_6$

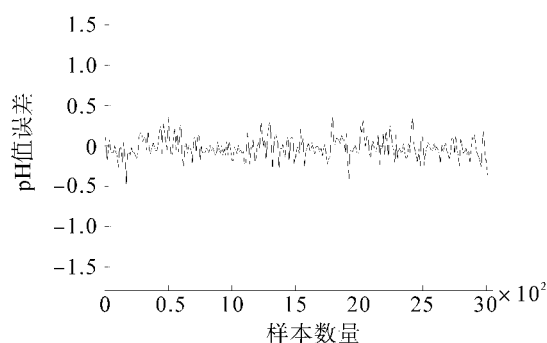
then

$$\begin{aligned} \tilde{y}_{\text{pH}}^6 = & [-9.8802, 8.0025] + \\ & [0.0200, 0.0008]\varphi(1) + \\ & [0.0640, 0.2299]\varphi(2) + \\ & [0.0093, -0.0055]\varphi(3) + \\ & [0.1441, -0.0432]\varphi(4). \end{aligned}$$

得到的模型输出结果如图 3 所示:



(a) 模型输出与实际输出的对比



(b) 模型输出的误差曲线

图 3 II 型 T-S 模糊模型辨识结果

Fig. 3 The results of type-II T-S fuzzy model identification  
模型的均方根误差(RMSE)为

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{j=1}^N (\hat{y}_j - y_j)^2}.$$

式中:  $\hat{y}_j$  为实际输出,  $y_j$  为模型输出,  $N$  为样本的个数. 对 pH 中和过程建立的模型, 本文中的 II 型 T-S 模糊模型的 RMSE 与 I 型 T-S 模糊模型的 RMSE 相比较如表 1:

表1 I型和 型T-S模糊模型误差比较

Table1 The comparison of type-I and type- T-S fuzzy models

算 法	规则数	RMSE
I型T-S模糊模型2	71	0.7483
改进后 型T-S模糊模型 <sup>[1]</sup>	6	0.2023
本文的 型T-S模糊模型	6	0.1342

可以看到 型模糊模型可以得到更高的精度,并且由于 型模糊模型的参数是区间模糊集,相比 I型包含了更多的信息。

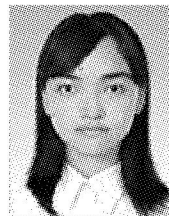
### 3 结 束 语

由于实际工业过程中不可避免地存在干扰噪声,采集到的数据与实际的数据会有或多或少的偏差。型模糊集在对过程的描述中增加了模糊性,增强了处理不确定因素的能力,能有效地减少不确定性带来的影响。提出了一种基于数据驱动的 型T-S模糊模型的建模方法,从现场采集到的数据样本出发,在I型T-S模糊模型的基础上,通过对数据样本的分析来判断模糊化的程度,然后对I型T-S模糊模型的前件和后件参数都分别进行模糊化,得到前件参数为区间 型模糊集,后件参数为区间I型模糊集的一种 型T-S模糊模型。最后通过对 pH中和过程的仿真来验证了建模算法的有效性,为存在不确定性因素影响的工业过程的研究提供了良好的前提条件。

### 参考文献:

- [1] ZADEH L A. The concept of a linguistic variable and its application to approximate reasoning-I[J]. Information Sciences, 1975, 8(9): 199-249.
- [2] KARNIK N N, MENDEL J M, LIANG Qilian. Type- I fuzzy logic systems[J]. IEEE Transactions on Fuzzy Systems, 1999, 7(6): 643-658.
- [3] 陈 薇, 孙增圻. 二型模糊系统研究与应用[J]. 模糊系统与数学, 2005, 19(1): 126-135.  
CHEN Wei, SUN Zengqi. Research on type-2 fuzzy logic system and its application[J]. Fuzzy Systems and Mathematics, 2005, 19(1): 126-135.
- [4] MENDEL J M, ROBERT J J, LIU Feilong. Interval type-2 fuzzy logic systems made simple[J]. IEEE Transactions on Fuzzy Systems, 2006, 14(6): 808-821.
- [5] LIANG Qilian, MENDEL J M. Interval type-2 fuzzy logic systems theory and design[J]. IEEE Transactions on Fuzzy Systems, 2000, 8(5): 535-550.
- [6] LANG Qilian, MENDEL J M. An introduction to type-2 TSK fuzzy logic systems[C]//Proceedings of IEEE International Conference on Fuzzy Systems. Seoul, Korea, 1999: 1534-1539.
- [7] MENDEZ G M, CASTILLO O. Interval type-2 TSK fuzzy logic systems using hybrid learning algorithm [C]//Proceedings of IEEE International Conference on Fuzzy Systems. Reno, USA, 2005: 230-235.
- [8] LI Ning, LI Shaoyuan, XI Yugeng. Modeling pH neutralization processes using fuzzy satisfactory clustering[C]//Proceedings of IEEE International Conference on Fuzzy Systems. Melbourne, Australia, 2001: 308-311.
- [9] GUSTAFSON D, KESSEL W C. Fuzzy clustering with a fuzzy covariance matrix[C]//Proc of IEEE CDC. San Diego, 1979: 761-766.
- [10] TAKAGI T, SUGENO M. Fuzzy identification of systems and its applications to modeling and control[J]. IEEE Transactions on Systems, 1985, 15(1): 116-132.
- [11] REN Qun, BARON L, BALAZINSKI M. Type-2 Takagi-Sugeno-Kang fuzzy logic modeling using subtractive clustering[C]//Fuzzy Information Processing Society (NAFIPS 2006). [S. l.], 2006: 120-125.
- [12] NIE Junhong, LOH A P, HANG C C. Modeling pH neutralization processes using fuzzy-neural approaches[J]. Fuzzy Sets and Systems, 1996, 78(1): 5-22.

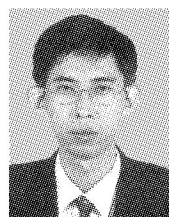
作者简介:



廖倩芳,女,1983年生,硕士研究生,主要研究方向为型模糊建模与控制



李 柠,女,1974年生,副研究员,硕士生导师,主要研究方向为复杂系统建模与控制、预测控制等。



李少远,男,1965年生,教授,博士生导师,主要研究方向为预测控制、自适应智能控制等。2006年获得上海市自然科学一等奖(第一完成人)。承担了包括国家自然科学基金、国家"863"计划在内的国家级科研项目10余项。在国内外学术杂志上发表学术论文180余篇,其中被SCI和EI检索100余篇。