

## 基于客观满意聚类的 pH 中和过程建模方法

王 娜<sup>1a,1b,2a</sup>, 胡超芳<sup>2b</sup>, 师五喜<sup>1a,1b</sup>

(1. 天津工业大学 a. 电气工程与自动化学院; b. 电工电能新技术天津市重点实验室, 天津 300387;  
2. 天津大学 a. 微光机电系统技术教育部重点实验室; b. 电气与自动化工程学院, 天津 300072)

**摘 要:** 针对一类生产中存在严重非线性的复杂工业过程——pH 中和过程, 基于客观聚类思想, 并结合 Gustafson-Kessel 聚类, 提出一种新的 T-S 模糊建模方法。根据用户对建模性能的满意度要求, 通过迭代模糊聚类, 进行模型前提结构和参数的辨识。仿真结果表明, 与传统的模糊聚类等方法相比, 该方法不依赖于系统的先验知识和预先定义的模糊隶属度函数, 具有较为精简的结构和更好的逼近性能, 对数据中的噪声具有一定的鲁棒性。

**关键词:** 复杂工业过程; T-S 模糊建模; 客观聚类; 模糊聚类; 满意聚类; pH 中和过程

**中文引用格式:** 王 娜, 胡超芳, 师五喜. 基于客观满意聚类的 pH 中和过程建模方法[J]. 计算机工程, 2018, 44(2): 298-303, 309.

**英文引用格式:** WANG Na, HU Chaofang, SHI Wuxi. Modelling Method Based on Objective Satisfactory Clustering for pH Neutralization Process[J]. Computer Engineering, 2018, 44(2): 298-303, 309.

## Modelling Method Based on Objective Satisfactory Clustering for pH Neutralization Process

WANG Na<sup>1a,1b,2a</sup>, HU Chaofang<sup>2b</sup>, SHI Wuxi<sup>1a,1b</sup>

(1a. School of Electrical Engineering and Automation; 1b. Tianjin Key Laboratory of Advanced Electrical Engineering and Energy Technology, Tianjin Polytechnic University, Tianjin 300387, China;  
2a. Key Laboratory of Micro Optical Electronic Mechanical System Technology, Ministry of Education;  
2b. School of Electrical Engineering and Automation, Tianjin University, Tianjin 300072, China)

**[Abstract]** For some manufacturing exists complex industrial processes with severe nonlinearity, named pH neutralization process, a new process T-S fuzzy modelling method via objective clustering idea, combined with Gustafson-Kessel (GK) clustering, is proposed. According to the satisfactory degree of modeling performance indexes, the premise structures and their parameters are identified through the iterated fuzzy clustering. Simulation results show that compared with the traditional fuzzy clustering, the proposed method does not rely on the prior-knowledge or defined fuzzy membership function, has a streamlined structure and better approximation performance, and is robust for the existing noise in data.

**[Key words]** complex industrial process; T-S fuzzy modelling; objective clustering; fuzzy clustering; satisfactory clustering; pH neutralization process

**DOI:** 10.3969/j.issn.1000-3428.2018.02.051

### 0 概述

pH 中和过程是具有严重非线性和大滞后的复杂工业过程<sup>[1-2]</sup>, 获得性能满意、能够反映其本质特征的模型是对其进行有效控制的重要前提。通常的解决方案有全局建模策略, 包括神经网络、支持向量机等, 但在处理 pH 中和这类数据分布离散、幅值变化较大的较强非线性对象时, 会因为结构的单一性而难以充分逼近其非线性特性, 导致模型的拟合精度和泛化能力较差<sup>[3]</sup>。

为此, 基于分解-合成思想的 T-S 模糊建模方法<sup>[3-5]</sup>, 通过在子空间建立起若干局部线性函数关系的组合来描述全局系统, 使规则库得以有效约简, 计算复杂性大大降低。

T-S 模型的构建通常采用模糊辨识, 即模糊子空间的划分来获取规则数、前提参数和后件参数。而规则数对应着模糊子空间的划分, 一般采用模糊聚类的方法实现<sup>[6-7]</sup>。但在实际应用中, 模糊聚类的初始参数, 即聚类个数和聚类中心通常是事先未知的, 一般通过经验或者人为试凑的方式给定, 有 2 类

**基金项目:** 天津大学微光机电系统技术教育部重点实验室开放课题基金“基于视觉的小型无人机跟踪系统研究”(MOMST2016-4)。

**作者简介:** 王 娜(1977—), 女, 讲师、博士, 主研方向为复杂工业过程建模、控制与优化; 胡超芳, 副教授、博士; 师五喜, 教授、博士。

**收稿日期:** 2016-11-16 **修回日期:** 2017-02-13 **E-mail:** wangna@tjpu.edu.cn

途径:比较法<sup>[3,8]</sup>或者融合法<sup>[3,9-10]</sup>。比较法通常采用若干聚类有效性指标的比较确定聚类数,易于受到人为决策的影响;而融合法中不重要聚类的判别和删减过程不仅增加了计算的复杂性,并且容易导致聚类空间出现空洞,使模型的泛化能力变差<sup>[9]</sup>。此外,由于聚类过程中易受到数据中噪声的影响,可能导致生成的聚类个数过少或者出现冗余,也会直接影响到模糊规则数的确定<sup>[11]</sup>。

为此,本文提出一种基于客观满意聚类的 pH 中和过程建模方法。首先改进原始客观聚类分析算法<sup>[12-14]</sup>,其具有对数据中噪声的较强鲁棒性,从而克服噪声对于聚类结果,即聚类个数和聚类中心的影响,可以直接确定一个比较适宜的初始规则数和初始聚类中心,并将其与 Gustafson-Kessel(GK)模糊聚类<sup>[15-16]</sup>相结合,利用 GK 聚类进一步优化原有的聚类中心及对应的模糊划分。此外,针对建模者对于模型的满意度,在初始模糊划分的基础上,再利用模型的满意度指标进行模糊聚类的迭代,以确定满意的模糊规则数和前提参数,使模型结构更加灵活有效,最后再利用稳态卡尔曼滤波算法<sup>[5]</sup>估计模型的后件参数,以克服传统最小二乘算法在求解病态矩阵时易导致非数值解的问题,提高计算的稳定性。

### 1 pH 中和过程描述

从系统建模的角度而言,pH 中和过程是一个多入多出过程,输入量为酸液流量  $F_a$  及所含成分浓度  $C_a$ ,碱液流量  $F_b$  及所含成分浓度  $C_b$ ,输出量为酸液浓度  $\omega_a$  和碱液的浓度  $\omega_b$ 。CSTR 的动态模型为:

$$V \frac{d\omega_a}{dt} = F_a C_a - (F_a + F_b) \omega_a \quad (1)$$

$$V \frac{d\omega_b}{dt} = F_b C_b - (F_a + F_b) \omega_b \quad (2)$$

其中,  $V$  为反应器的容积。

不失一般性,考虑弱酸强碱过程,其中和滴定方程为:

$$w_b + 10^{-pH} - 10^{pH-14} - \frac{w_a}{1 + 10^{pK_a - pH}} = 0 \quad (3)$$

其中,  $w_a = [Cl^-]$ ,  $w_b = [Na^+]$ ,  $K_a = 1.76 \times 10^{-5}$  为弱酸的电离常数,且  $pK_a = -\lg K_a$ 。中和滴定曲线如图 1 所示,可见 pH 中和值在中点附近灵敏度很高,而在远离中和点处灵敏度很低。

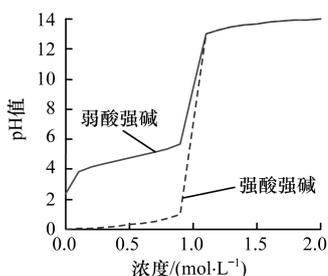


图 1 弱酸强碱及强酸强碱中和滴定曲线

### 2 T-S 模糊模型

在非线性系统建模中,T-S 模糊模型是一种常用的有效模型结构,规则形式如下<sup>[5]</sup>:

$$\begin{aligned} &\text{IF } u_1 \text{ is } A_1^i \text{ and } u_2 \text{ is } A_2^i \text{ and } \dots \text{ and } u_p \text{ is } A_p^i \\ &\text{THEN } y_i = a_{i0} + a_{i1}u_1 + \dots + a_{ip}u_p \end{aligned} \quad (4)$$

其中,  $n$  为规则总数,给定输入变量  $u = [u_1, u_2, \dots, u_p]$ ,则  $A_j^i$  表示为输入数据  $u_j$  对于第  $i$  条规则所属的模糊隶属度,  $y_i$  为第  $i$  条规则的输出,  $[a_{i0}, a_{i1}, \dots, a_{ip}]$  为规则  $i$  的后件参数矩阵。

采用加权平均法,T-S 模型的推理可简化如下:

$$y = \sum_{i=1}^n \lambda_i(u) y_i \quad (5)$$

其中,  $\lambda_i(x) = \frac{\beta_i(x)}{\sum_{j=1}^n \beta_j(x)}$  为第  $i$  条规则在全部规则集中的权重,  $\beta_i(x) = A_1^i \wedge A_2^i \wedge \dots \wedge A_p^i$  为规则  $i$  对  $u$  的激活程度。

### 3 T-S 模型前提参数辨识

#### 3.1 客观满意聚类原理

客观满意聚类首先引入原始的客观聚类算法<sup>[12-14]</sup>对其进行改进,并将其引入 Gustafson-Kessel(GK)聚类<sup>[15-16]</sup>来提供比较合适的初始聚类个数和聚类中心,以此作为原始输入-输出数据空间的初始模糊划分,并依据用户提供的建模满意度准则,进行后续新的聚类个数生成和聚类中心的更新,以确定最终的模糊划分,其原理如图 2 所示。

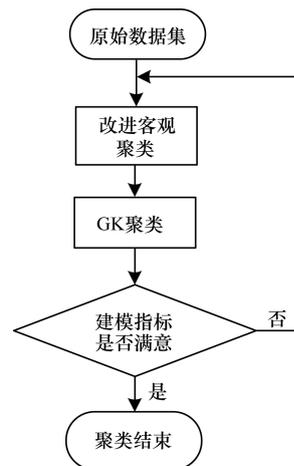


图 2 客观满意聚类原理

#### 3.1.1 GK 聚类

与传统模糊聚类算法相比,GK 聚类利用自适应距离度量取代固定的欧式距离,进行类内距离的计算,从而有效克服了模糊 C 均值算法易受初始值影响而陷入局部极值的缺点,因而更加适合模糊辨识。

给定数据集  $Z$ ,GK 通过求取目标函数  $J$  的极小值来获取给定聚类个数  $c$  下的模糊隶属度矩阵  $U$  和对应的聚类中心  $V$ :

$$\min J(Z, V, U) = \sum_{i=1}^c \sum_{j=1}^N \mu_{i,j}^m d^2(z_j, v_i) \quad (6)$$

其中,  $d(z_j, v_i)$  表示样本  $z_j$  与聚类中心  $v_i$  之间的距离, 受实际的聚类形状影响, 并由聚类协方差矩阵  $F_i$  决定。

$$F_i = \frac{\sum_{j=1}^N \mu_{i,j}^m (z_j - v_i)(z_j - v_i)^T}{\sum_{j=1}^N \mu_{i,j}^m} \quad (7)$$

$$d^2(z_j, v_i) = (z_j - v_i)^T M_i (z_j - v_i) \quad (8)$$

$$M_i = \det(F_i)^{\frac{1}{c+1}} F_i^{-1} \quad (9)$$

但 GK 聚类的初始参数, 即聚类个数  $c$  和聚类中心  $V$  难以确定, 常见有比较<sup>[3,8]</sup>和融合法<sup>[3,9-10]</sup>。但在如何确定聚类的初始个数上, 比较法依赖于聚类有效性指标的相互比较, 而融合法采用最大聚类个数开始, 依次逐步合并的方式确定最终聚类个数。两者均比较依赖于经验, 或者某些假定的数据条件, 比如必须满足高斯正态分布等。而对于缺乏先验知识, 或者不满足外在假设条件的数据集则缺少有效的提炼聚类个数的方法。并且, 还容易受到数据中噪声、数据分布等因素的影响, 从而给模型前件参数的确定带来较多的不确定性, 进而导致不准确的辨识结果, 影响整个模型的精度。为此, 提出改进客观聚类算法, 并与 GK 聚类相结合, 形成客观满意聚类方法, 以改善这一状况。

### 3.1.2 客观聚类原理

客观聚类的基本原理如图 3 所示。主要采用相似性数据分级和自组织聚类实现对相似数据子集的共同聚类和比较, 其中相似性数据分级利用两两距离最近的数据对排序, 以将整个数据集划分为训练集  $A$ 、 $B$  或者测试集  $C$ 、 $D$ , 即  $Z = \{A, B\}$  or  $\{C, D\}$ , 并分别在这 4 个子集上进行自组织聚类, 将各自的聚类结果两两比较以决定最终的聚类结果。

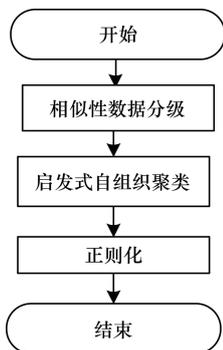


图 3 客观聚类的基本原理

为克服原有算法的最近邻聚类效应, 导致难于形成较小聚类, 以及原有一致性计算中出现的多值解问题, 本文采用改进客观聚类算法<sup>[14]</sup>, 即引入相对不相似性测度 (Relative Dissimilarity Measure, RDM)<sup>[17]</sup> 来确定待融合的聚类, 以降低最近邻引起

聚类过程的滚雪球问题, 提高算法的准确性, 并提出改进一致性准则来克服多值解的收敛问题, 提高求解的稳定性。

与传统硬聚类相比, 改进客观聚类的特点主要为:

#### 1) 相似性数据分级

通过相似性分级, 原始的数据集  $Z$  被重新划分为 2 组数据子集  $\{A, B\}$  或者  $\{C, D\}$ 。相比之下, 子集  $A$  与  $B$  的相似度最高, 而  $C$  与  $D$  亦然。并且,  $\{A, B\}$  的相似度高于  $\{C, D\}$ , 因此将  $\{A, B\}$  作为内部训练集进行聚类, 来获得原始数据集  $Z$  的一种可能的聚类结果, 而  $\{C, D\}$  作为测试集来校验  $\{A, B\}$  结果的准确性, 为最终的聚类结果的确定提供外部依据, 因此, 在内外 2 种准则的共同检验下, 最终聚类结果的有效性得以保证。

以训练集  $A$ 、 $B$  为例, 其构成的偶极子如图 4 所示,  $C$ 、 $D$  子集的构成亦然。给定待划分的数据集  $Z = \{z_i\}, i = 1, 2, \dots, N$ , 形如  $\begin{pmatrix} z_i \\ z_j \end{pmatrix}$  的数据对称为偶极子, 其中,  $i \neq j, i, j = 1, 2, \dots, N, k \in \{1, 2, \dots, \frac{N \times (N-1)}{2}\}$ 。  $d_{ij} = \|z_i - z_j\|$  表示偶极子  $O_{ij}^k$  的值,  $\|\cdot\|$  表示欧式距离。则按照定义, 数据集  $Z$  共产生  $C_n^2$  对偶极子,  $n = 1, 2, \dots, \frac{N \times (N-1)}{2}$ 。

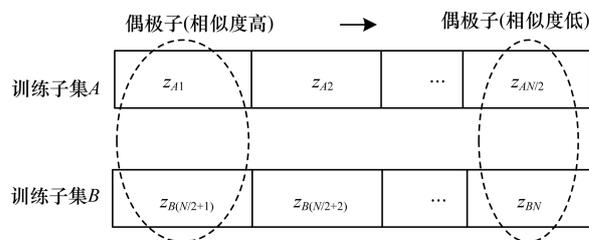


图 4 训练子集  $A$ 、 $B$  的偶极子构成

#### 2) 启发式自组织聚类

不同于传统的层次聚类, 启发式自组织聚类利用一种新定义的距离测度-相对不相似性测度 (RDM) 来衡量<sup>[17]</sup> 每次聚类中待融合的 2 个聚类  $i$  和  $j$  之间的相似性, 并且还考虑了未被合并的聚类  $k$  对于本次聚类合并的影响, 对于数据的分布更具有较强鲁棒性。

$$D_{ij} = \frac{d_{ij}}{\min(\bar{d}_{ij}, \bar{d}_{ji})} \quad (10)$$

$$\bar{d}_{ij} = \frac{\sum_{k \in C, k \neq j} d_{ik}}{c-2} \quad (11)$$

其中,  $c$  代表聚类个数,  $d_{ij}$  代表聚类  $i$ 、 $j$  的距离,  $\bar{d}_{ij}/\bar{d}_{ji}$  为聚类  $i$  ( $j$ ) 到除  $j$  ( $i$ ) 外其他聚类的平均距离,  $D_{ij}$  表示所有聚类中  $i$  和  $j$  间的相对相似性。

改进一致性准则定义如下:

$$ECC = \frac{1}{mc} \sum_{i=1}^m \sum_{j=1}^c (o_{ij}^A - o_{ij}^B)^2 \quad (12)$$

其中,  $m$  表示特征变量维数,  $c$  表示当前聚类个数,  $o_{ij}^A$  和  $o_{ij}^B$  分别表示训练子集  $Z_A$  和  $Z_B$  的相对应的第  $i$  个聚类中心向量的  $j$  分量。

与原始的一致性准则相比,改进的一致性计算采用同一位置下不同子集的聚类中心之间的差异来代替数据点之间的差值,从而降低了数据中噪声点对一致性计算的影响,确保了聚类结果的准确性。

### 3.2 基于客观满意聚类的 T-S 模型前提参数辨识

采用式(1)所示的 T-S 模型,给定训练数据集  $Z = \{z_1, z_2, \dots, z_N\}$ , 其中  $z_k = (u_1^k, u_2^k, \dots, u_r^k, y_1^k, y_2^k, \dots, y_q^k) = (x_k^1, x_k^2, \dots, x_k^{r+q})$ ,  $k = 1, 2, \dots, N$ ,  $r$  为输入变量维数,  $q$  为输出变量维数,  $N$  为数据集的长度。采用客观满意聚类算法对  $Z$  进行输入输出空间上的模糊聚类,获得聚类个数  $c$ , 初始聚类中心为  $V_0$ , 得到模糊划分矩阵  $U_F = [\mu_{ki}]$ , 其中  $\mu_{ki} \in [0, 1]$  表示第  $k$  个数据属于第  $i$  类的隶属度,  $i = 1, 2, \dots, c, k = 1, 2, \dots, N$ 。

则对于任意给定的输入数据  $z_l \in Z, l \in \{1, 2, \dots, N\}$ , 利用投影法可以确定前提参数为  $A_j^i(z_l^j) = \mu_{li}$ ,  $j = 1, 2, \dots, r + q, i = 1, 2, \dots, c$ 。这种直接确定前提参数的方式,可以降低隶属度函数的参数估计时易导致的拟合误差,从而提高辨识精度,并减少融合法中迭代辨识前提参数和结论参数的计算量,也避免了各个聚类有效性函数下聚类结果的计算和比较,计算效率大大提高。

## 4 基于客观满意聚类的 T-S 模糊辨识

如 2.3 节所述,利用所提的客观满意聚类算法的聚类结果可以确定 T-S 模型的初始结构,规则数即为聚类个数,模糊划分矩阵中的隶属度元素即为所对应规则的前提参数,但这一结果并非模糊模型的最优结构。其主要原因在于聚类算法在使用数据建模时,主要利用聚类的相似性原理,以寻找最优的聚类为目的,而建模对象之间的输入-输出关系的某些非线性特性就可能被忽视,造成欠拟合,所以最终的聚类结果可能并不能满足建模精度要求。

如图 5 所示,在非线性曲线中,曲线  $l_1$  和  $l_2$  段的动态特性仅用聚类  $C_1$  来表示,  $l_4$  和  $l_5$  段仅用聚类  $C_3$  所覆盖,显然建模的拟合精度不足。VS 表示非常小, S 表示小, M 表示中, L 表示大, VL 表示非常大。为此,在客观满意聚类基础上,引入一种误差反馈机制,利用聚类后模型与实际数据输出之间的最大误差来反馈所建模型的拟合效果,并以此为聚类是否继续迭代的依据,直至模型与数据输出之间的误差降低至某一允

许的精度范围内终止。在此前提参数的辨识基础上,利用稳态卡尔曼滤波算法<sup>[5]</sup>估计规则的后件参数,完成整个 T-S 模糊模型的结构和参数辨识。

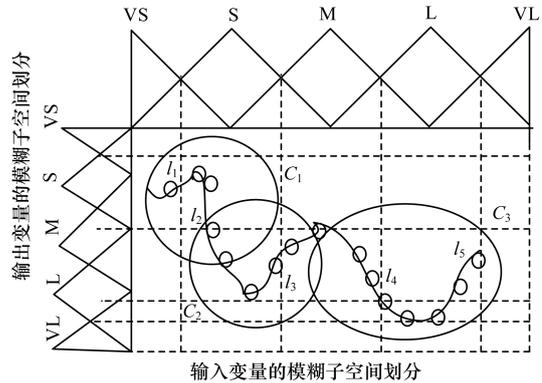


图 5 模糊聚类划分的欠拟合

算法步骤归纳如下:

**步骤 1** 采用 EOCA 算法确定一个初始聚类结果  $\{c_0, V_0\}$ ,  $c_0$  为聚类个数,  $V_0$  为初始聚类中心集。

**步骤 2** 以上述聚类结果初始化 GK 聚类算法, 得到聚类中心集  $V$ , 以得到的模糊划分矩阵  $U_F = [\mu_{ij}]_{c \times N}$  确定前提参数  $A_j^i(x_j^l)$ , 并采用稳态卡尔曼滤波方法估计后件参数  $\hat{a}$ 。

**步骤 3** 采用均方误差指标  $MSE$ ,  $MSE = \frac{1}{N} \sum_{i=1}^N [\hat{y} - y]^2$ , 其中  $y$  为实际输出,  $\hat{y}$  为模型估计输出, 计算所建模型与实际输出数据之间的误差指标, 判断:若  $MSE$  满足用户的建模需求, 则算法结束; 否则转到步骤 4。

**步骤 4** 计算每个数据的输出  $y_j$  与其对应的模型输出  $\hat{y}_j$  的误差  $err_j$ , 取其最大者  $err_{\max} = \max \{err_j\}, j = 1, 2, \dots, N$  为数据集的长度, 将数据  $z_{\max}$  加入到原始聚类中心集  $\{V\}$  中, 得到新的聚类中心集  $V_1 = \{V, z_j\}$ , 并将聚类个数增加 1, 即  $c = c + 1$ , 并转到步骤 1 ~ 步骤 3, 继续计算新模型的均方误差指标, 直至所得  $MSE$  满足用户的建模要求精度为止, 终止迭代过程。

## 5 仿真研究

考虑式(1)~式(3)描述的弱酸强碱中和过程, 给定酸液流量  $F_a$ , 则系统输入输出分别为碱液流量  $F_b$  和流出物的 pH 值。样本集  $(F_b, y_{pH})$  可通过机理模型式(1)~式(3)得到。在  $F_b$  中加入  $[-51.5, +51.5]$  范围内的变化量, 由此产生 300 组样本。假设 T-S 模型包含 2 个输入量, 其表达式为:

$$\hat{y}_{pH}(t) = F(F_b(t-1), y_{pH}(t-1)) \quad (13)$$

其中,  $y_{pH}$  和  $\hat{y}_{pH}$  分别表示实际系统的输出和模型的估计值。

从上述训练集  $(F_b(t-1), y_{pH}(t-1), y_{pH}(t))$  出发,采用所提的客观模糊聚类方法辨识该过程。首先,利用客观聚类方法对训练数据依次进行偶极子划分、启发式自组织聚类和聚类结果的正则化处理,由训练子集  $Z_A$  和  $Z_B$  得到改进一致性指标  $ita_{AB}=0.0041, c=3, Z_C、Z_D$  子集上得到  $ita_{CD}=0.0081, c=2$ 。取  $ita_{AB}$  和  $ita_{CD}$  两者之中的最小值,则得到  $c=2$  和对应的初始聚类中心  $V_i = \begin{bmatrix} 0.56 & 0.70 & 0.87 \\ 0.51 & 0.20 & 0.14 \end{bmatrix}$ 。训练子集  $Z_C、Z_D$  上的一致性准则与聚类个数关系曲线如图6所示。

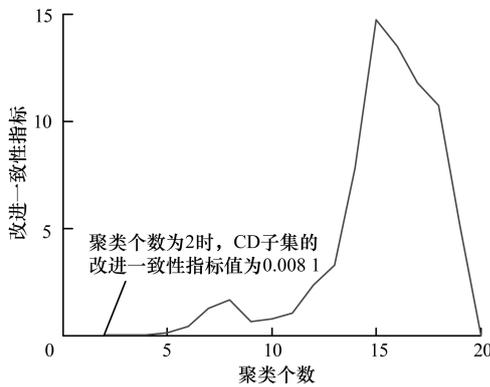


图6 一致性准则与聚类个数关系曲线

给定初始规则数  $n=c=2$ , 初始聚类中心  $V_i$ , 采用 GK 和稳态卡尔曼滤波法分别辨识前提参数和结论参数之后,得到第  $i$  条规则  $R_i$  为:

$$R^i: \text{if } F_b(t-1) \text{ is } A_1^i \text{ and } y_{pH}(t-1) \text{ is } A_2^i, \\ \text{then } y_{pH}(t) = a_{i0} + a_{i1}F_b(t-1) + a_{i2}y_{pH}(t-1) \quad (14)$$

计算此时模型的误差性能指标  $MSE_{n=2} = 1.0981$ , 此时的误差曲面如图7所示,其中输入分别为  $t_1$  时刻的碱液流量  $u_1 = F_b(t-1)$  和  $t_1$  时刻的流出物 pH 值  $u_2 = y_{pH}(t-1)$ , 输出为  $t$  时刻实际 pH 值与模型输出之间的误差  $error = y_{pH}(t) - \hat{y}_{pH}(t)$ 。由误差曲面图可见,实际系统与模型的拟合误差分布比较广泛,并且存在着最大和最小误差所形成的误差凸峰,其中最大误差为 3.45, 最小误差为 -2.37。

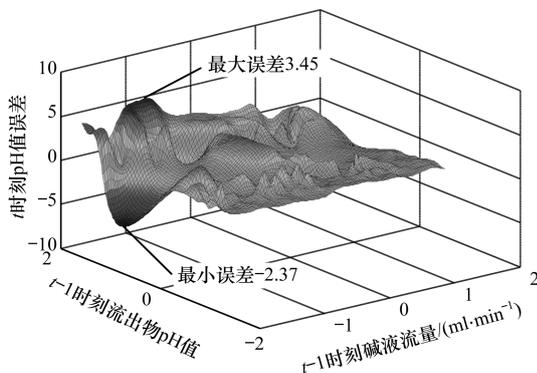


图7 无噪声下实际系统输出与模型估计的误差曲面

由图7可见,该误差显示当前所建模型与实际系统的拟合效果不能令人满意,因此计算每个数据点上的输出  $y_j$  与模型的估计输出  $\hat{y}_j$  之间的误差  $error_j, j \in \{1, 2, \dots, 150\}$ , 找出正负误差中绝对值最大者  $error_{max} = 3.45$ , 将其对应的数据设为新增的聚类中心矢量  $v_j = [-1.61 \ 1.87 \ 1.81]$ , 则当前聚类中心更新为

$$\begin{pmatrix} (V_i)_{2 \times 3} \\ v_j \end{pmatrix} = \begin{pmatrix} 0.64 & 1.54 & 1.51 \\ 0.76 & -0.63 & -0.67 \\ -1.09 & -0.63 & -0.67 \end{pmatrix}_{3 \times 3}$$

$n=3$  时的  $MSE_{n=3} = 1.2221$ , 增加了一条规则,但与规则数为 2 时误差反而增大了,说明此时模型的拟合能力下降,反而不能达到满意,而且  $n=2$  时显然对应的规则数也不是最好的覆盖,而可能是一个局部极值。为此,利用绝对误差最大的原则,继续增加新的聚类中心为  $v_k = [1.21 \ 1.67 \ -0.65]$  和规则数,令新产生的聚类中

$$\text{心为 } \begin{pmatrix} V_i \\ v_j \\ v_k \end{pmatrix}_{4 \times 3}$$

为 5 时,聚类中心为  $v_l = [-1.02 \ 1.58 \ 1.33]$ , 得到规则数

$$\begin{pmatrix} 0.64 & 1.54 & 1.51 \\ 0.76 & -0.63 & -0.67 \\ -1.09 & -0.63 & -0.67 \\ 0.55 & -0.36 & 1.11 \\ -1.02 & 1.58 & 1.33 \end{pmatrix}_{5 \times 3}$$

应  $MSE_{c=5} = 0.082$ , 误差的精度等级明显较聚类个数为 4 时下降一个量级,说明该聚类中心对模型精度具有重要作用,故不再继续增加聚类中心,整个模糊辨识迭代过程结束,转入基于稳态卡尔曼滤波的后件参数辨识过程,得到后件参数:

$$[a_{10}, a_{11}, a_{12}, \dots, a_{50}, a_{51}, a_{52}] = [1.57, -0.90, -1.20, 0.96, 1.67, 0.01, 0.04, -0.09, -0.49, -0.33, 0.04, -0.23, -0.63, -1.52, -0.4]$$

最终产生的 5 个聚类中心在整个数据集中的分布如图8所示。

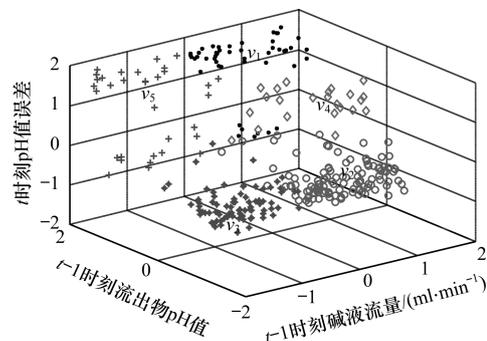


图8 无噪声下  $c=5$  时的聚类子集与聚类中心分布

由图 8 可见,整个数据集被均匀划分为 5 个聚类,每个聚类中心即为模糊规则的中心,数据与聚类中心之间的模糊隶属度矩阵决定了其前提参数。以第 1 个数据为例,其在模糊矩阵中表示为:  $U_{c5}(1,:) = [0.03 \ 0.79 \ 0.06 \ 0.09 \ 0.02]$ ,以其最大的隶属度值 0.79 应该归为第 2 类,其余数据以此类推,形成上述 5 个聚类子集  $\{c_1, c_2, c_3, c_4, c_5\}$ ,对应聚类中心  $\{v_1, v_2, v_3, v_4, v_5\}$ 。系统输出与模型输出的比较曲线如图 9 所示。

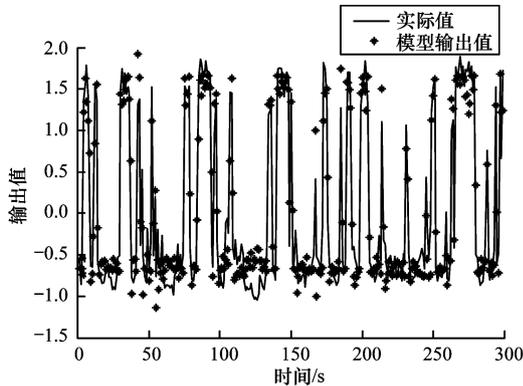


图 9 无噪声下系统实际输出与模型输出的误差比较曲线

上述结果为未添加噪声下的模型,为验证所建模型的鲁棒性,对上述数据的输入和输出同时添加信噪比为 5 dB 的白高斯噪声,采用同样方法训练,结果得到  $(ita_{CD} = 0.0033)_{c=2} \angle (ita_{AB} = 1.1832)_{c=6}$ ,故选取初始聚类个数为 2。经过 GK 模糊聚类和迭代,以及稳态卡尔曼滤波后,得到聚类个数为 4 时  $MSE$  为 0.64,而聚类个数为 5 时  $MSE$  为 0.65,误差精度变化不大,故确定最终规则数为 4,  $MSE$  为 0.64,如图 10 所示。

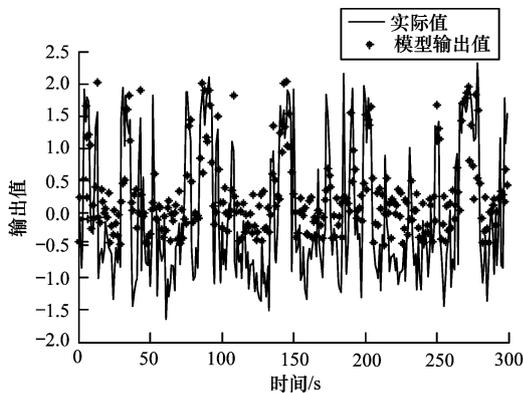


图 10 有噪声下系统实际输出与模型输出的误差比较曲线

无噪声环境下,采用上述训练数据集,所提方法与其他模糊聚类建模效果比较如表 1 所示。由表 1 可见,与 USOCPN 相比,本文方法规则数较少,并且训练精度处于同一量纲内,因此具有更加精简的模型结构,其结构和参数辨识的计算效率相应就得以大大降低。与 GK 模糊聚类相比,本文方法取规则

数为 5 时,其建模精度处于同一量纲范围内,但其初始聚类个数和初始聚类中心均可由客观聚类方法直接一次给定,避免了 GK 模糊聚类中初始聚类参数难以确定,采用试凑而带来的不确定性,更加直接有效。并且,在添加噪声之后,与 GK 模糊聚类相比,模型的拟合精度处于同一量级范围内,但却避免了 GK 模糊聚类难以确定模糊规则数的缺点,使得初始结构的确定更加准确。同时,能够以较少的规则数获得较为满意的建模精度。

表 1 客观满意聚类与其他模糊聚类的建模性能比较

方法	规则数	均方误差
USOCPN <sup>[2,18]</sup> (无噪声)	71	0.560
GK 模糊聚类 <sup>[18]</sup> (无噪声)	4	0.097
GK 模糊聚类 <sup>[18]</sup> (添加 5 dB 白高斯噪声)	4	0.211
	5	0.203
客观满意聚类 (无噪声)	4	1.720
	5	0.082
客观满意聚类 (添加 5 dB 白高斯噪声)	4	0.640
	5	0.650

## 6 结束语

本文利用客观模糊聚类辨识 pH 中和过程,提供了一种新型的 T-S 模糊建模方法。该方法能从系统的实际输入输出数据出发,自适应地快速确定系统的模糊规则数、前提参数和后件参数。对 pH 过程的仿真研究验证了所建模型对系统具有较强的拟合能力,并且对于噪声数据的拟合有较强的鲁棒性,结构简单,易于实现。

## 参考文献

- [1] BUCHHOLT F, KUMMEL M. Self-tuning Control of a pH Neutralization Process [J]. Automatica, 1979, 15(6):665-671.
- [2] 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.
- [3] 李 柠, 李少远, 席裕庚. 利用模糊满意聚类建立 pH 中和过程模型[J]. 控制与决策, 2002, 17(2):143-147.
- [4] 李香娜. 基于 T-S 模糊模型的酸碱中和过程的系统辨识[J]. 华北科技学院学报, 2010, 7(3):56-60.
- [5] TAKAGI T, SUGENO M. Fuzzy Identification of Systems and Applications to Modeling and Control[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1985, 15(1):116-132.
- [6] YI Yang, FAN Xiangxiang, ZHANG Tianping. Anti-disturbance Tracking Control for Systems with Nonlinear Disturbances Using T-S Fuzzy Modeling [J]. Neurocomputing, 2016, 171(3):1027-1037.

## 参考文献

- [1] SAITO T, ISO Y, KIRYU N. A Study of Technical Aspect of Web Browser Fingerprinting [C]//Proceedings of the 31st Symposium on Cryptography and Information Security. Berlin, Germany: Springer, 2010.
- [2] 张梦媛. 浏览器的安全访问及指纹识别技术 [D]. 南京: 南京邮电大学, 2012.
- [3] 吴建涛. 浏览器指纹技术研究与实现 [D]. 西安: 西北大学, 2014.
- [4] 王研昊, 马媛媛, 杨明, 等. 基于隐性标识符的零权限 Android 智能终端识别 [J]. 东南大学学报(自然科学版), 2015, 45(6): 1046-1050.
- [5] 王研昊. Android 设备指纹识别技术的研究与实现 [D]. 南京: 东南大学, 2015.
- [6] NIKIFORAKIS N, KAPRAVELOS A, JOOSEN W, et al. On the Workings and Current Practices of Web-based Device Fingerprinting [J]. IEEE Security & Privacy, 2014, 12(3): 28-36.
- [7] ACAR G, JUAREZ M, NIKIFORAKIS N, et al. Fpdetective: Dusting the Web for Fingerprinters [C]//Proceedings of ACM SIGSAC Conference on Computer & Communications Security. Washington D. C., USA: IEEE Press, 2013: 1129-1140.
- [8] 张玉清, 武倩如, 刘奇旭, 等. 第三方追踪的安全研究 [J]. 通信学报, 2014, 35(9): 1-11.
- [9] ECKERSLEY P. How Unique is Your Web Browser [C]//Proceedings of the 10th Privacy Enhancing Technologies Symposium. Berlin, Germany: Springer, 2010: 1-18.
- [10] NIKIFORAKIS N, KAPRAVELOS A, JOOSEN W. Cookieless Monster: Exploring the Ecosystem of Web-based Device Fingerprinting [C]//Proceedings of the 34th IEEE Symposium of Security and Privacy. Washington D. C., USA: IEEE Press, 2013: 541-555.
- [11] MOWERY K, SHACHAM H. Pixel Perfect: Fingerprinting Canvas in HTML5 [C]//Proceedings of Web 2.0 Security and Privacy Conference. Washington D. C., USA: IEEE Press, 2012: 1-12.
- [12] TAKASU K, SAITO T, YAMADA T. A Survey of Hardware Features in Modern Browsers: 2015 Edition [C]//Proceedings of the 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. Washington D. C., USA: IEEE Press, 2015: 520-524.
- [13] MOWERY K, BOGENREIF D, YILEK S, et al. Fingerprinting Information in JavaScript Implementations [C]//Proceedings of IEEE International Conference on Acoustics, Speech & Signal Processing. Washington D. C., USA: IEEE Press, 2013: 9-12.
- [14] MULAZZANI M, RESCHL P, HUBER M, et al. Fast and Reliable Browser Identification with Javascript Engine Fingerprinting [C]//Proceedings of Web 2.0 Security and Privacy Conference. Washington D. C., USA: IEEE Press, 2013: 503-506.
- [15] UPATHILAKE R, LI Yingkun, MATRAWY A. A Classification of Web Browser Fingerprinting Techniques [C]//Proceedings of the 7th International Conference on New Technologies, Mobility and Security. Washington D. C., USA: IEEE Press, 2015: 1-5.
- [16] 王维, 王伟平, 王建新. 一种基于升级浏览器指纹的用户识别方法 [C]//第七届信息安全漏洞分析与风险评估大会论文集. 南京: [出版者不详], 2014: 189-197.
- (上接第 303 页)
- [7] 宋俐, 谢刚, 杨云云. 基于模糊聚类的社团划分算法 [J]. 计算机工程, 2016, 42(8): 126-133.
- [8] QIAO Junfei, LI Wei, ZENG Xiaojun. Identification of Fuzzy Neural Networks by Forward Recursive Input-output Clustering and Accurate Similarity Analysis [J]. Applied Soft Computing, 2016, 49: 524-543.
- [9] KRISHNAPURAM R, FREG C P. Fitting an Unknown Number of Lines and Planes to Image Data Through Compatible Cluster Merging [J]. Pattern Recognition, 1992, 25(4): 385-400.
- [10] KAYMAK U, BABUSKA R. Compatible Cluster Merging for Fuzzy Modeling [C]//Proceedings of IEEE International Conference on Fuzzy System. Washington D. C., USA: IEEE Press, 1995: 897-904.
- [11] 刘士荣, 余建军, 林卫星, 等. 启发式模糊聚类学习在函数逼近和非线性系统建模中的应用 [J]. 模式识别与人工智能, 2003, 16(2): 230-235.
- [12] 王娜. 基于客观聚类的模糊建模方法研究 [D]. 上海: 上海交通大学, 2009.
- [13] 贺昌政. 自组织数据挖掘与经济预测 [M]. 北京: 科学出版社, 2005.
- [14] 王娜, 杨煜普. 一种基于改进客观聚类分析的模糊辨识方法 [J]. 控制与决策, 2009, 24(1): 13-17, 22.
- [15] QIU Haobo, XU Yanjiao, GAO Liang, et al. Multi-stage Design Space Reduction and Metamodeling Optimization Method Based on Self-organizing Maps and Fuzzy Clustering [J]. Expert Systems with Applications, 2016, 46(3): 180-195.
- [16] GUSTAFSON D, KESSEL W C. Fuzzy Clustering with a Fuzzy Covariance Matrix [C]//Proceedings of IEEE Conference on Decision & Control. Washington D. C., USA: IEEE Press, 1979: 761-766.
- [17] MOLLINEDA R A, ENRIQUE V. A Relative Approach to Hierarchical Clustering [M]. Amsterdam, the Netherlands: IOS Press, 2000: 19-28.
- [18] 李柠. 多模型建模与控制的若干问题研究 [D]. 上海: 上海交通大学, 2002.

编辑 顾逸斐

编辑 顾逸斐