文章编号：1000－8152(2003)03－0783－05

基于满意聚类的多模型建模方法

李　柠，李少远，席裕庚
（上海交通大学 自动化研究所，上海 200030）

摘要：从系统输入输出数据出发，首先在GK模糊聚类算法的基础上，提出一种模糊满意聚类算法，该算法能快速对系统进行用户满意的模糊划分；继而将其引入多模型建模过程中，满意的系统划分数目即对应多模型个数，然后针对不同的聚类建立起相应的子系统模型，全局系统可视为各子模型的加权组合；最后通过几个典型实例验证了模糊满意聚类及基于此的多模型建模方法的有效性、准确性和快速性.

关键词：多模型；Gustafson-Kessel模糊聚类(GK)；满意聚类

中图分类号：TP273　　**文献标识码**：A

Multi-model modeling method based on satisfactory clustering

LI Ning, LI Shao-yuan, XI Yu-geng
(Institute of Automation, Shanghai Jiao Tong University, Shanghai 200030, China)

Abstract: First of all, from input-output data set, a satisfactory clustering algorithm based on GK fuzzy clustering was presented. Using this algorithm, a system could be quickly divided into multiple optimal fuzzy parts. Then the algorithm was used in multi-model modeling process. Satisfactory cluster number corresponded to the optimal number of sub-systems. For multiple clusters, multiple models could then be built, and the global system was described as their certain combination. Finally, examples are given to prove the effectiveness of the method.

Key words: multi-model; Gustafson-Kessel fuzzy clustering (GK); satisfactory clustering

1　引言(Introduction)

实际生产过程往往具有非线性、工况范围广、控制性能要求高等特点，采用单一模型对其进行描述已无法满足要求，即使能够获得满意的单一全局模型，其辨识过程也是相当困难的.基于分解合成法则(divide-and-conquer)的多模型建模方法从原理上解决了上述问题，并在实际中获得成功的应用[1~3].其基本思路是将系统划分为若干子系统，而整个系统可视作各子系统的某种组合.如何将复杂系统划分为多个子系统，换言之，原全局系统用多少个线性模型来表述并确定各自所在的有效区间就成为首先要解决的问题.本文提出的模糊满意聚类算法将用于解决上述问题.

聚类算法是将数据集根据相似性准则划分为若干子集，因此可利用聚类进行系统的多模型划分.然而，聚类算法往往要求聚类个数 c 事先给定，而 c 依赖于系统所呈现的非线性程度，因此若对系统没有充分的了解，准确的 c 初始值很难直接确定，这也一定程度上困扰了聚类算法的应用.目前较常见的 c 确定方法有两种：比较法(validity measure)[4]和融合法(cluster merging)[4,5].前者利用某种度量指标来评价聚类的质量，即将样本集进行若干次聚类($c \in [2, N]$)，N 表示样本个数，其中对应于最小度量指标的聚类个数即被视作最佳的聚类数目.后者首先从较大的聚类个数 $c_{\max}$ 开始聚类 $c(c \in [2, c_{\max}])$，$c_{\max}$ 足够大以覆盖整个系统的非线性特征，然后陆续将相近聚类中心进行合并，以此减少聚类数目.尽管后者计算量会逐步减小，其累积计算量仍很可观，这对于大样本集合尤其明显.本文从系统输入输出数据出发，针对比较法和融合法中计算量大，初始聚类个数选取盲目的缺点，依据用户满意为最终目标的原则，提出一种简单有效的快速满意 c 确定方法，

收稿日期：2001－05－14；收修改稿日期：2002－03－28.
基金项目：国家自然科学基金(69934020,60074004)资助项目.

并将其应用于系统多模型建模过程中.

2 基于满意聚类的多模型建模(Multi-model modeling method based on satisfactory clustering)

聚类方法有许多,其原理是按相似性将数据集划分为几个类别以表征系统的不同性征[6],同时各类别间应满足彼此间最小的重叠,以避免聚类的重复,即各聚类中心彼此之间应该在包含足够多相似样本的基础上最不相似,换句话讲,同一聚类中的样本应尽可能的靠近,而不同聚类中心之间的距离应尽可能的远.聚类算法也就是寻找若干包含一组与其相似的样本的最不相似样本中心,才能最大程度的代表系统的不同特征,同时各聚类中心应包含足够数目的样本以保证以尽可能少的聚类中心表达系统.

由此,选取GK(Gustafson-Kessel)模糊聚类[7]作为基本聚类算法,并在此基础上提出基于GK的模糊满意聚类算法.简单地说,给定初始聚类个数 $c=2$,对系统进行初次聚类后,若聚类效果尚未令人满意,则从样本集中找出一个与各聚类中心点 $v_1 \sim v_c$ 最不相似的样本作为新的样本中心 v_{c+1},并将 $v_1 \sim v_{c+1}$ 作为初始聚类中心,在此基础上粗略计算新的非随机的隶属度矩阵,继而用GK算法重新对系统进行 $c+1$ 类划分,根据性能指标的要求重复上述步骤,直到得出令人满意的结果.获得满意的聚类个数后(聚类个数 c 对应于系统的子模型个数),可以认为系统有了满意划分,继而采用最小二乘法等辨识出各子模型的参数,最终以各子模型的加权求和形式获得系统的整体描述.

考虑一MISO系统,其样本集由系统的输入输出数据组成,假设其样本表示为 (φ_j, y_j), $j=1,\cdots,N$, φ_j 表示影响系统输出的递推向量,一般选择系统当前及以往的输入输出作为其向量分量,y_j 是系统输出.定义 $z_j=[\varphi_j, y_j]^{\mathrm{T}}$,则样本集可表示为 $Z=[z_1,\cdots,z_N]$,其中 $z_j \in \mathbb{R}^{d+1}$.假定样本集 Z 被分成 c 个聚类 $\{Z_1, Z_2, \cdots, Z_c\}$,则系统可由 c 个子模型 $\{M_1, M_2, \cdots, M_c\}$ 表征,图1给出了基于GK模糊满意聚类的多模型方法的结构示意图.

基于GK模糊满意聚类的多模型方法的具体步骤为

Step 1 令初始聚类个数 $c=2$;

Step 2 由初始隶属度矩阵 U_0,利用GK算法将样本集合 Z 进行分类,得出隶属度矩阵 $U=[\mu_{i,j}]_{c\times N}$,然后根据每组样本所属各子集的隶属度选取最大值进行分类,将 Z 分为 c 个子集 $\{Z_1, Z_2, \cdots, Z_c\}$;

Step 3 对聚类后生成的每个子集采用稳态Kalman滤波器迭代算法[8]辨识出各子模型的参数,借助Step 2中生成的隶属度矩阵可以方便的得出参数集 $P=[p_0^1,\cdots,p_0^c,p_1^1,\cdots,p_1^c,\cdots,p_d^1,\cdots,p_d^c]^{\mathrm{T}}$,则对应各聚类中心的子模型可以描述为

$$\begin{aligned} M_1&: y^1 = p_0^1 + p_1^1\varphi(1) + \cdots + p_d^1\varphi(d),\\ M_2&: y^2 = p_0^2 + p_1^2\varphi(1) + \cdots + p_d^2\varphi(d),\\ &\vdots\\ M_c&: y^c = p_0^c + p_1^c\varphi(1) + \cdots + p_d^c\varphi(d). \end{aligned} \tag{2.1}$$

Step 4 计算出来的隶属度矩阵 $U=[\mu_{i,j}]_{c\times N}$ 可直接作为输入 z_j 隶属于第 j 条规则的程度,则对应输入 z_j 的系统输出为

$$\hat{y} = \sum_{i=1}^{c}\mu_{ij}\,y_i \Big/ \sum_{i=1}^{c}\mu_{ij}. \tag{2.2}$$

若要预测新的样本输入 $\tilde{\varphi}$ 对应的输出 $\tilde{y}$,则回到GK算法通过下式计算 $\tilde{\varphi}\in\mathbb{R}^d$ 对应第 i 条规则的隶属度 $\tilde{\mu}_i$[9],

$$\tilde{\mu}_i(\tilde{\varphi}) = \frac{1}{\sum_{j=1}^{c}\left(D_{A_i^x}(\tilde{\varphi}, v_i^x)/D_{A_i^x}(\tilde{\varphi}, v_j^x)\right)^{2/(m-1)}}. \tag{2.3}$$

其中,v_i^x 表示第 i 个聚类中心除去输出分量后剩余的向量部分,$v_i^x\in\mathbb{R}^d$,$D_{A_i^x}(\tilde{\varphi}, v_i^x)$ 表示新输入向量与第 i 个聚类之间的按照GK聚类算法中定义的距离函数,$m>1$ 是表征聚类模糊程度的可调参数,m 越大各聚类之间的重叠越多,通常取 $m=2$,则预测输出 $\tilde{y}$ 可按式(2.2)得出;

Step 5 计算用户给定的系统性能指标的当前值 $S_c=\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i-y_i)^2}$;若 $S_c\leqslant S_{TH}$,S_{TH} 为用户认为满意的性能指标阈值,则认为多模型建模结束;否则,认为系统聚类不成功,转Step 6;

Step 6 在样本集中,根据隶属度矩阵 U 找出一个与各子集均不相似样本 z_n, $n=1,\cdots,N$,不相似性可按下式给出

$$n = \arg\min_{n}\sum_{\substack{1\leqslant i,j\leqslant c\\ i\neq j}}(\mu_{ni}-\mu_{nj}). \tag{2.4}$$

为避免噪声,一般应找出几个类似的样本求其平均值,作为新的聚类中心 v_{c+1};

Step 7 以 $v_1,\cdots,v_c,v_{c+1}$ 为新的聚类初始中心,按照常用的隶属度函数粗略计算相应的新的初

始隶属度矩阵 U_0，而非比较法中重新初始化的随机矩阵；

Step 8　令 $c = c + 1, U = U_0$，转 Step 2.

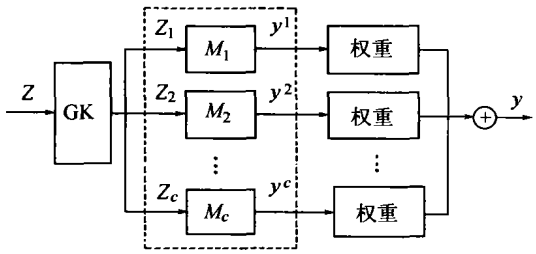

图 1　基于模糊满意聚类的多模型方法的结构示意图

Fig. 1　Diagram of multi-model identification method based on fuzzy satisfactory clustering

满意聚类算法避免了聚类融和方法中根据系统非线性特征确定 c_{max}，而直接采用 $c = 2$ 为初始化条件；而且除初次聚类外，以后聚类初始化参数，如隶属度矩阵等，可根据上次聚类结果预先确定，不必再从随机量开始重新聚类，因此计算的收敛速度将明显加快，对于大样本量的数据集，快速性更为明显.

3　仿真实例(Simulation examples)

3.1　非线性函数(Nonlinear function)

考虑一双输入单输出的非线性静态系统[10]

$$z = (1 + x^{-2} + y^{-1.5})^2,\ 1 \leqslant x, y \leqslant 5. \quad (3.1)$$

针对文献[10]中提供的50组输入输出数据，采用本文提出的多模型方法对系统进行建模，其结果如表1和图2所示.

表 1　非线性系统建模误差

Table 1　Modeling error comparison of nonlinear system

模型	规则数	RMSE
Sugeno[11]	6	0.564
本文方法	4/5	0.337/0.319

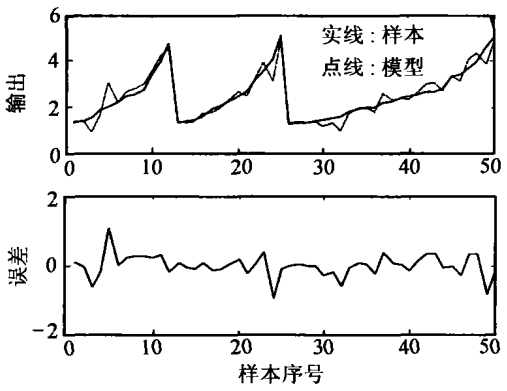

图 2　非线性系统建模结果

Fig. 2　Nonlinear system modeling result

3.2　Box-Jenkins 煤气炉数据(Box-Jenkins gas furnace)

Box-Jenkins 煤气炉是系统辨识的一个典型例子，其数据集包含296组输入输出观测数据[13]，其中输入 $u(t)$ 为进入煤气炉的煤气流量，输出 $y(t)$ 为释放出的煤气中的 CO_2 浓度.本例采用 $y(t-1)$ 和 $u(t-4)$ 作为模型的输入.其聚类及辨识结果如下：聚类中心 $v_i(i = 1,2)$ 和结论参数集 P，

$v_1 = [-0.1701, 53.7942, 53.7422]$,

$v_2 = [0.0268, 53.3764, 53.5230]$,

$P = [28.2875, 7.3565, -1.4375, -1.0771, 0.4635, 0.8719]$.

图3则给出了采用本文算法在聚类个数为2时获得的建模效果，并在表2中比较了不同建模方法下的均方根误差，结果令人满意.

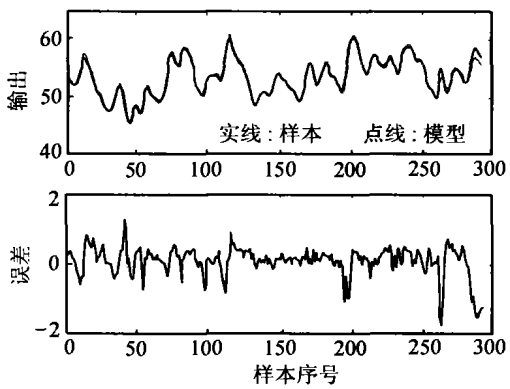

图 3　Box-Jenkins 煤气炉建模结果

Fig. 3　Box-Jenkins gas furnace modeling result

表 2　Box-Jenkins 煤气炉建模方法误差比较[12]

Table 2　Modeling error comparison of Box-Jenkins gas furnace[12]

模型	模型输入个数	规则数	RMSE
Tong'77	2	19	0.684
Pedrycz'84	2	81	0.565
Xu'87	2	25	0.572
Peng'88	2	49	0.548
Sugeno'91	6	2	0.261
Sugeno'93	3	6	0.435
Wang'96	2	5	0.397
本文方法	2	2	0.426

3.3　pH 中和过程(pH neutralization process)

pH 中和过程是具有严重非线性和滞后性的复杂工业过程，其建模与控制也是工业过程控制的难题之一.考虑一弱酸强碱中和过程[14]，采用上一时

刻碱流量及 pH 值作为模型输入，在碱流量中加入范围在[-51.5, +51.5]的随机扰动，从而产生 300 组输入输出数据，在此基础上采用本文算法进行系统辨识，结果如表 3 及图 4 所示.

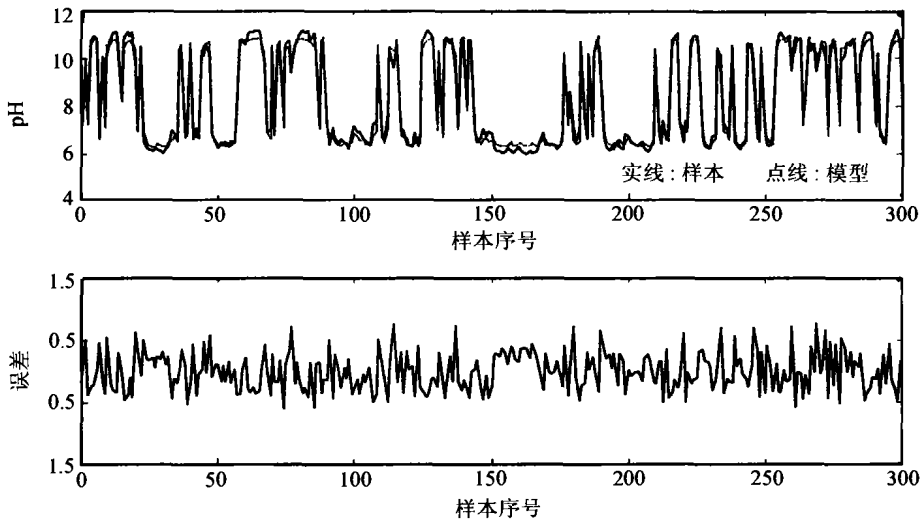

图 4 pH 中和过程建模结果

Fig. 4 pH neutralization process modeling result

表 3 pH 中和过程模型误差比较

Table 3 Modeling error comparison of pH process

模型	规则数	RMSE
Nie[14]	71	0.560
本文方法	4	0.312

以上是采用本文提出的建模方法对典型的 pH 中和过程辨识的结果. 而为了进一步说明采用满意聚类较原有聚类方法在计算快速性上的优势，文中就此对 pH 中和过程作了仿真比较，图 5 给出了当聚类中心 $c = 4$ 时，GK 算法分别采用满意聚类及比较法时的收敛曲线. 其中实线代表本文算法，并利用了 $c = 3$ 时的聚类结果，迭代步数为 12. 而虚线表示采用比较法的聚类收敛曲线，图中给出 3 个随机初始隶属度矩阵 U_0，其迭代步数分别为 205, 172, 244. 类似的结论在其他实验中同样能够得到.

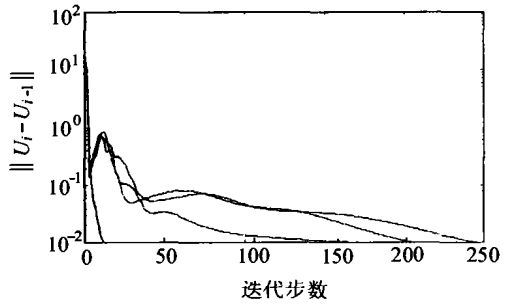

图 5 pH 中和过程聚类收敛曲线

Fig. 5 Convergence curve of modeling results of pH neutralization process

4 结束语(Conclusions)

在 GK 算法的基础上，提出一种满意聚类思想. 该算法从 2 个聚类出发，根据需要增加新的聚类，聚类过程中充分利用以往聚类的结果，避免了重新选取聚类隶属度矩阵及聚类中心的随机性，收敛速度明显加快. 文中将满意聚类引入建模中，在此基础上提出一种基于 GK 模糊满意聚类的多模型建模方法，并对几个典型实例进行了仿真，结果充分说明了满意聚类个数确定方法以及基于此的多模型建模方法的快速性、准确性和有效性.

参考文献(References):

[1] MURRY-SMITH R, JOHANSEN T A. *Multiple Model Approaches to Modeling and Control* [M]. London: Taylor and Francis, 1997.

[2] EIKENS B, KARIM M N. Process identification with multiple neural network models [J]. *Int J Control*, 1999, 72(7/8): 576 - 590.

[3] POTTMANN M, UNBEHAUEN H, SEBORG D E. Application of a general multi-model approach for identification of highly nonlinear processed—a case study [J]. *Int J Control*, 1993, 57(1): 97 - 120.

[4] KRISHNAPURAM R, CHIN-PIN F. Fitting an unknown number of lines and planes to image data through compatible cluster merging [J]. *Pattern Recognition*, 1992, 25(4): 385 - 400.

[5] KAYMAK U, BABUSKA R. Compatible cluster merging for fuzzy modeling [A]. *Proc of IEEE Int Conf on Fuzzy Systems* [C]. Yokohama: IEEE Press, 1995: 897 - 904.

[6] ZHONG W. *Studies on soft-sensing & advanced control strategies with applications in petrochemical processes* [D]. Shanghai: East China University of Science and Technology, 1999.

[7] GUSTAFSON D, KESSEL W C. Fuzzy clustering with a fuzzy co-

variance matrix [A]. *Proc of IEEE Conference on Decision and Control* [C]. San Diego, CA: IEEE Press, 1979: 761 - 766.

[8] TAKAGI T, SUGENO M. Fuzzy identification of systems and its applications to modeling and control [J]. *IEEE Trans on Systems, Man, and Cybernetics*, 1985, 15(1): 116 - 132.

[9] BABUSKA B. *Fuzzy Modeling for Control* [M]. Boston: Kluwer Academic Publishers, 1998.

[10] NAKANISHI H, TURKSEN I B, SUGENO M. A review and comparison of six reasoning method [J]. *Fuzzy Sets and Systems*, 1992, 57(2): 257 - 294.

[11] SUGENO M, YASUKAWA T. A fuzzy-logic-based approach to qualitative modeling [J]. *IEEE Trans on Fuzzy Systems*, 1993, 1(1): 7 - 31.

[12] GOMEZ-SKARMETA A F, DELGADO M, VILA M A. About the use of fuzzy clustering techniques for fuzzy model identification [J]. *Fuzzy Sets and Systems*, 1999, 106(2): 180 - 188.

[13] BOX G E P, JENKINS G M. *Time Series Analysis Forecasting and Control* [M]. San Francisco, CA: Holden-Day, 1970.

[14] NIE J H, 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：

李　柠　(1974 —)，女，上海交通大学自动化所博士生，研究领域为复杂系统建模与控制，预测控制，模糊系统等；

李少远　(1965 —)，男，上海交通大学自动化所教授，研究领域为预测控制，模糊控制，自适应控制理论与应用，E-mail: syli@sjtu.edu.cn；

席裕庚　(1946 —)，男，上海交通大学自动化所教授，博士生导师，研究领域包括预测控制，大系统及智能机器人等。

《控制理论与应用》第五届编委会会议在湖北神农架召开

《控制理论与应用》第五届编委会会议于2003年8月7日至10日在湖北省宜昌市神农架举行．来自大陆及香港的24所高校和科研单位的40多位顾问、编委及特邀代表参加了会议．会议由主编陈翰馥院士主持，吴捷主编做了第四届编委会工作的总结报告．两位主编还分别代表两个主办单位向到会的顾问和编委颁发了聘书．

与会代表对第四届编委会5年来的工作进行了讨论和总结，肯定了上一届编委会的成绩，并对下一届编委会的主要工作进行了展望．

1．第四届编委会5年来的工作成绩显著：成功举办了纪念关肇直先生诞辰80周年学术研讨会；申请创办《控制理论与应用》英文刊(JOURNAL OF CONTROL THEORY AND APPLICATIONS) 获得教育部和科技部批准；本刊继续被美国《数学评论》、英国《科学文摘》、莫斯科《文摘杂志》和美国《工程索引》Ei page one等国际检索系统收录外，又相继被德国《数学文摘》(Zentralblatt MATH)、美国《剑桥科学文摘社网站：电子与通讯文摘》(CSA: ECA)和美国《剑桥科学文摘社网站：计算机与信息系统文摘》(CSA: CISA)收录；此外还开通了电子投稿，试行网上审稿与邮寄相结合，大大缩短了审稿周期．

2．第五届编委会的工作继续围绕以提高刊物质量和在国际学术界的知名度为中心进行运作．加强选题与组稿，编委要带头审稿、组稿、推荐好文章，并撰写高质量的稿件；稿件实行责任编委负责制和推荐制，对于具创造性、创新性的优秀文章优先发表．

3．从作者、审稿人到编委都应重视中英文摘要尤其英文摘要的作用，特别是要符合国际重要检索数据库的摘要撰写要求，进一步扩大本刊在国际同行中的影响．摘要的语言要规范，信息量要大．

4．《控制理论与应用》英文刊将于今年年底出版创刊号，她将在中文刊的起点上逐渐实现论文作者、编委及审稿专家、读者的国际化，争取早日进入SCI．欢迎专家、学者向本刊投稿，对于有创造性和创新性的、居国际水平的文章优先发表，并给予某些优惠．

5．在提高稿件送审准确性的同时，花大力气缩短审稿周期．

6．坚决反对、杜绝学术上的不良风气，对投稿中出现的剽窃、一稿多投等现象，本刊将严肃处理．

与会代表特别缅怀本刊已故的前任主编关肇直教授、许国志院士，顾问张钟俊院士、张学铭教授、疏松桂教授以及编委程勉教授、王恩平研究员、李训经教授；对多年来一直关心和支持本刊工作，但因年事已高、已离开编委队伍的老专家、老教授表示衷心感谢．大家一致称赞5年来华南理工大学和中国科学院系统科学研究所的各级领导及其编委对出版发行《控制理论与应用》所给予的关心和支持，对编辑部同志们的认真、努力工作表示衷心感谢，并预祝本刊在两个主办单位的领导和支持下，进一步加强编辑力量，为将《控制理论与应用》中英文刊办成国际知名的刊物而努力．