

DOI:10.3969/j.issn.1673-4785.201412001
网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.tp.20150716.0934.004.html>

基于特征选择聚类方法的稀疏 TSK 模糊系统

张佳骊, 蒋亦樟, 王士同
(江南大学 数字媒体学院, 江苏 无锡 214122)

摘 要: 为避免模糊系统建模和估计领域的“维数灾难”, 将 TSK(Takagi-Sugeno-Kang)模糊系统建模转换为一个分块稀疏表示问题, 提出 FCA 稀疏 TSK 模糊系统(FCA-sparse TSK)。首先运用模糊聚类算法(FCA)对样本特征进行化简, 并产生模糊系统字典; 再利用存在于 TSK 模糊系统中的分块结构信息, 选取重要的模糊规则并对所选模糊规则的后件参数进行估计。该系统同时对模糊规则及模糊规则数进行化简, 在合成数据集和真实数据集上都表现出较好的性能。

关键词: T-S 模糊系统; 模糊系统字典; 模糊聚类; 特征选择; 分块结构; 稀疏表示; 规则约减; 参数估计

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2015)04-0583-09

中文引用格式: 张佳骊, 蒋亦樟, 王士同. 基于特征聚类方法的稀疏 TSK 模糊系统[J]. 智能系统学报, 2015, 10(4): 583-591.
英文引用格式: ZHANG Jiasu, JIANG Yizhang, WANG Shitong. Sparse TSK fuzzy system based on feature clustering method[J]. CAAI Transactions on Intelligent Systems, 2015, 10(4): 583-591.

Sparse TSK fuzzy system based on feature selection clustering method

ZHANG Jiasu, JIANG Yizhang, WANG Shitong
(School of Digital Media, Jiangnan University, Wuxi 214122, China)

Abstract: In order to solve the curse of dimensionality existing in fuzzy system identification and approximation, this paper proposes the FCA-sparseTSK fuzzy system by casting the Takagi-Sugeno-Kang(TSK) fuzzy system identification into a block sparse representation problem. First, FCA-sparseTSK fuzzy system uses the fuzzy clustering algorithm(FCA) to simplify sample features and generate fuzzy system dictionary. Then selects main important fuzzy rules and estimate the fuzzy rule's consequent parameter vector by taking into account the block-structured information that exists in the TSK fuzzy model. The FCA-sparseTSK fuzzy system simplifies the fuzzy rules and the number of fuzzy rules at the same time and shows good performance in artificial datasets and real-world datasets.

Keywords: TSK fuzzy system; fuzzy system dictionary; fuzzy clustering; feature selection; block structure; sparse representation; rules reduction; parameter estimation

在各种不同的模糊推理系统之中, Takagi-Sugeno-Kang(TSK)模糊系统提供了一个合理的框架, 将一个非线性系统分解成若干局部线性模型。因为传统的数学模型无法描述复杂非线性系统的行为, 所以 T-S 模糊系统的潜在应用范围很广。近年来, 在进行复杂非线性系统的控制器设计过程中, 将 TSK 模糊控制方法和各种控制策略(随机控制、滑模变结构控制等)相结合, 成为目前控制线性系统控制领域的一个研究热点。但是, 基于数据驱动的 TSK 模糊系统建模并不简单, 并且可能会产生一个非线性规划问题。在模糊系统模型中, 模糊规则包括前件隶属度函数以及相应的后件参数, 寻找到最优的模糊规则是模糊系统建模过程中较为重要的工作。

一些数值优化算法, 例如通过梯度下降优化技术^[2-3]的模糊神经网络技术、均衡模型复杂度和精确度的遗传算法^[4]、Levenberg-Marquardt 算法^[5]等,

收稿日期: 2014-12-01. 网络出版日期: 2015-07-16.
基金项目: 国家自然科学基金资助项目(61272210); 江苏省自然科学基金资助项目(BK20111417, BK20130155).
通信作者: 张佳骊. E-mail: jiasu0306@163.com.

已经被广泛应用于模糊系统建模领域。基于数据驱动的 TSK 模糊系统建模过程中有 2 个步骤非常关键:一是确定模糊规则前件,在这一步中将输入空间划分成一定数量的模糊域;二是估计模糊规则后件参数向量,在这一步中在每一个用于建立模糊系统的模糊域中对系统的行为进行描述。我们常常利用聚类技术划分输入空间并且确定模糊规则前件的隶属度函数,例如 K 均值算法^[6]、模糊 c 均值算法及其扩展^[7]、减法聚类算法^[8]、Gath-Geva 聚类算法^[9]、Gustafson-Kessel 聚类算法^[10]、向量量化(VQ)算法^[11]等。

在模糊规则前件确定之后,对后件参数的估计可以看作是在输入输出数据的积空间的一个线性回归问题。传统的回归算法将所有的后件参数视为回归系数并且将它们单独处理。但一方面,这些方法忽视了存在于 TSK 模糊模型中的分块结构信息;另一方面,冗余的模糊规则不可避免地导致模糊模型的复杂度变高以及过拟合的问题。

为了解决当前 TSK 模糊系统建模问题中的模糊规则复杂及规则冗余的问题。Luo M 等提出了 H-sparseFIS 算法^[12],这是一种层次结构稀疏表示方法,能有效地降低模糊规则的冗余度。本文对 TSK 模糊系统的建模过程做了进一步的优化:在前件学习时采用一种具备特征选择功能的聚类方法 FCA 算法来获取更为精简有效的特征组合,从而降低模糊规则的复杂度;在估计模糊规则后件参数过程中,利用存在于 TSK 模糊模型中的分块结构信息,将分块结构稀疏表示引入 TSK 模糊系统建模的框架中,利用块匹配追踪算法^[12]选择出较为重要的模糊规则,同时减少了非零后件参数的数目,进一步减少模糊规则的条数,有效地降低了规则的冗余度。

1 TSK 模糊系统概况

在这一部分,首先对 TSK 模糊系统的一些基本概念进行回顾,然后对模糊系统字典和模糊系统子字典的概念进行介绍。这些子字典构成了一个有意义的分块结构 TSK 模糊系统字典。

TSK 模糊系统是由前件和后件形式为“IF-THEN”的模糊规则组成的,模糊规则的后件被定义为一个仿射函数。对于一个 n 维输入变量 $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in})^T \in \mathbf{R}^n$, 第 j 条规则可以表示成下面这种形式:

R^j : If x_{i1} is A_{j1} and x_{i2} is A_{j2} ... and x_{in} is A_{jn} , then

$$y_i = w_{j0} + w_{j1}x_{i1} + w_{j2}x_{i2} + \dots + w_{jn}x_{in}$$

式中 $j = 1, 2, \dots, r$ 。在文中使用钟形隶属度函数对

模糊语言命题 $A_{jk}(x_{ik})$ 进行描述:

$$\mu_{A_{jk}}(x_{ik}) = \exp \left[- \left(\frac{x_{ik} - c_{jk}}{\sigma_{jk}} \right)^2 \right] \quad (1)$$

式中: $k = 1, 2, \dots, n$, c_{jk} 和 σ_{jk} 分别表示相应的钟形隶属函数的均值和方差。TSK 模糊系统的输出函数可以表示为

$$\hat{y}_i = \sum_{j=1}^r \varphi_j(\mathbf{x}_i) l_j(\mathbf{x}_i) \quad (2)$$

式中 $\varphi_j(\mathbf{x}_i)$ 表示输入变量 \mathbf{x}_i 对于第 j 条规则的触发强度:

$$\begin{cases} \varphi_j(\mathbf{x}_i) = \frac{\mu_j(\mathbf{x}_i)}{\sum_{i=1}^r \mu_i(\mathbf{x}_i)} \\ \mu_j(\mathbf{x}_i) = \prod_{k=1}^n \mu_{A_{jk}}(x_{ik}) \end{cases} \quad (3)$$

而 $l_j(\mathbf{x}_i) = w_{j0} + w_{j1}x_{i1} + w_{j2}x_{i2} + \dots + w_{jn}x_{in} = [1, \mathbf{x}_i^T]^T \mathbf{w}_j$ 是输入变量 \mathbf{x}_i 对于第 j 条规则的输出结果,其中 $\mathbf{w}_j = [w_{j0}, w_{j1}, w_{j2}, \dots, w_{jn}]^T$ 表示第 j 条规则后件的参数。

接下来介绍一下 TSK 模糊系统建模在数据驱动方面的一些概念。先来看一下输入输出数据集 $\mathbf{D} = \{(\mathbf{x}_i^T, y_i)^T; \mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T, i = 1, 2, \dots, N\}$ 式中 \mathbf{x}_i 和 y_i 分别表示第 i 个 n 维输入和输出变量。为了简洁,用 M'_D 表示有 r 条规则的 TSK 模糊系统,且这个系统是从数据集 \mathbf{D} 进行学习的。将 TSK 模糊系统 M'_D 的输出表示成 $\hat{\mathbf{y}} = [\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N]^T$:

$$\hat{\mathbf{y}} = \sum_{j=1}^r \Phi_j \mathbf{w}_j = \Phi \mathbf{w} \quad (4)$$

式中: $\Phi \in \mathbf{R}^{N \times (nr+1)}$ 是由 r 个块结构 $\Phi_j \in \mathbf{R}^{N \times (n+1)}$ 构成,也就是说, $\Phi = [\Phi_1, \Phi_2, \dots, \Phi_r]$ 。

$$\Phi_j = \text{diag}(\varphi_j(\mathbf{x}_1), \varphi_j(\mathbf{x}_2), \dots, \varphi_j(\mathbf{x}_N)) \mathbf{X}_e \quad (5)$$

式中: $j = 1, 2, \dots, r$, $\mathbf{X}_e = [1, \mathbf{X}^T]$, 其中 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ 。相应的, $\mathbf{w} = [\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_r^T]^T \in \mathbf{R}^{r(n+1)}$ 。根据定义, $\Phi \in \mathbf{R}^{N \times (nr+1)}$ 叫做 TSK 模糊系统 M'_D 的字典,成分 $\Phi_j \in \mathbf{R}^{N \times (n+1)}$ 是对应于第 j 条规则的字字典。很明显,这个 TSK 模糊系统字典是一个分块结构。在这个意义上讲,模糊模型输出 $\hat{\mathbf{y}}$ 可以用模糊规则子字典的一个线性组合表示。

2 FCA 稀疏 TSK 模糊系统

在模糊规则前件提取中,使用 FCA 聚类算法对样本输入空间进行划分并确定模糊规则前件的隶属度函数。在这个过程中, FCA 算法能够对噪声特征和不重要的特征进行约减。之后,对模糊系统字典的分块结构信息进行阐述,并将 TSK 模糊系统建模过程转化为一个分块结构稀疏恢复问题。在模糊规

则选取的过程之中,我们使用块正交匹配追踪算法,这样,就能减少模糊系统的规则数目。

2.1 模糊规则前件提取

在很多特定的情况下,专家可以根据先验知识提供一些模糊规则。但在大多数情况下,对于划分输入空间和确定模糊规则前件的隶属度函数,通常使用统计学方法或聚类技术。本文使用 FCA 算法从数据集 D 提取模糊规则前件。

通过 FCA 算法^[13],每个类均与一条模糊规则相联系。通过类在每个变量上的投影产生模糊规则前件的钟形隶属度函数 A_{jk} ,也就是说,在每一维上的类中心和方差被视为钟形隶属度函数 A_{jk} 的均值 c_{jk} 和方差 σ_{jk} ,这里, $j = 1, 2, \dots, r, k = 1, 2, \dots, n$ 。

一旦模糊规则前件确定下来,模糊字典 Φ 以及所包含的模糊规则子字典 $\Phi_j(j = 1, 2, \dots, r)$ 也可以通过式(5)获得。一般来说,通过对模糊规则回归模型的后件参数估计的方法来对 TSK 模糊系统进行建立。基于 FCA 算法的 TSK 模糊系统的建模流程图如图 1 所示。

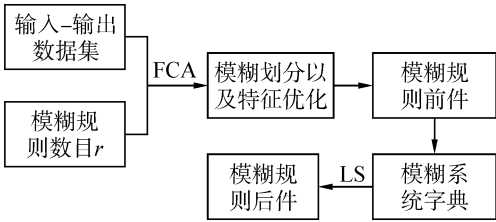


图 1 基于 FCA 聚类算法的 TSK 模糊系统模型流程图
Fig.1 Flowchart of TSK fuzzy system modeling on the basis of FCA clustering

FCA 算法是一种加权模糊聚类算法,该算法在实现对数据进行聚类的时候,可以按样本特征对聚类的贡献进行排序。因此,本文先使用 FCA 算法实现模糊划分并挑选出对聚类贡献较大的若干特征生成模糊规则前件,从而化简了模糊规则的复杂度。

依据上文,样本集的输入部分可以表示为 $X = [x_1, x_2, \dots, x_N], \forall x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n, C$ 为聚类数目,集合 $V = [v_1, v_2, \dots, v_C], \forall v_i = [v_{i1}, v_{i2}, \dots, v_{in}]^T \in R^n$ 表示聚类中心, μ_{ij} 表示样本 x_i 隶属于第 j 类的隶属度, $W = [w_1, w_2, \dots, w_n]$ 是样本特征的权重,此处 $w_k > 0$ 且 α 为其参数, $\beta = [\beta_1, \beta_2, \dots, \beta_n]$ 是样本的权重,此处 $\beta_i > 0$ 且 γ 为其参数,建立目标函数如下:

$$J(U, V, W, \beta) = \sum_{i=1}^N \sum_{j=1}^C \frac{\mu_{ij}^m}{\beta_i^\gamma} \left(\sum_{k=1}^n w_k^\alpha d(x_{ik}, v_{jk}) \right) \tag{6}$$

式中: $d(x_{ik}, v_{jk}) = (x_{ik} - v_{jk})^2, \sum_{k=1}^n w_k = 1, \sum_{i=1}^N \beta_i =$

$$1, \sum_{j=1}^C \mu_{ij} = 1, m > 1。$$

在 FCA 算法中, U, V, W 和 β 在 γ 和 α 满足一定条件的时候能够使得 $J(U, V, W, \beta)$ 达到局部最优,下面分别给出 U, V, W 和 β 的迭代更新表达式:

$$w_k = \frac{1}{\sum_{p=1}^n \frac{D_k^{\frac{1}{\alpha-1}}}{D_p^{\frac{1}{\alpha-1}}}} \tag{7}$$

式中 $D_k = \sum_{i=1}^N \sum_{j=1}^C \frac{\mu_{ij}^m}{\beta_i^\gamma} d(x_{ik}, v_{jk})$ 。

$$\beta_i = \frac{F_i^{\frac{1}{\gamma+1}}}{\sum_{s=1}^N F_s^{\frac{1}{\gamma+1}}} \tag{8}$$

式中 $F_i = \sum_{j=1}^C \mu_{ij}^m \left(\sum_{k=1}^n w_k^\alpha d(x_{ik}, v_{jk}) \right)$ 。

$$\mu_{ij} = \frac{1}{\sum_{q=1}^C \frac{f_{ij}^{\frac{1}{m-1}}}{f_{iq}^{\frac{1}{m-1}}}} \tag{9}$$

式中 $f_{ij} = \sum_{k=1}^n w_k^\alpha d(x_{ik}, v_{jk})$ 。

$$v_{jk} = \frac{\sum_{i=1}^N \frac{\mu_{ij}^m}{\beta_i^\gamma} x_{ik}}{\sum_{i=1}^N \frac{\mu_{ij}^m}{\beta_i^\gamma}} \tag{10}$$

根据上述更新规则,在有限次迭代后, $J(U, V, W, \beta)$ 将会收敛到局部极小值或鞍点。具体 FCA 算法(算法 1)如下。

1) 输入:输入输出数据集的输入部分 $X = [x_1, x_2, \dots, x_N]$, 聚类数 r , 初始聚类中心 V^0 以及误差控制量 $\varepsilon > 0$ (其中 V^0 可不设定)。

2) 初始化:根据数据集信息随机初始化 U^0 , 随机产生满足条件的 W^0, β^0 , 若未输入初始聚类中心,则通过公式(10)计算初始聚类中心。

3) 步骤:
FOR $n = 1, 2, \dots$
 使用式(7)迭代产生 W^{n+1}
 使用式(8)迭代产生 β^{n+1}
 使用式(9)迭代产生 U^{n+1}
 使用式(10)迭代产生 V^{n+1}
IF
| $J(U^{n+1}, V^{n+1}, W^{n+1}, \beta^{n+1}) - J(U^n, V^n, W^n, \beta^n)$ | < ε
 break
ELSE $n = n + 1$
4) 输出: $U^{n+1}, V^{n+1}, W^{n+1}$ 和 β^{n+1} 。

2.2 模糊规则选取的分块结构稀疏表示

通常,使用 LS 算法对模糊规则后件参数向量 w_i 估计的方式有 2 种,一种是通过求解 LS 问题的 $w=[w_1, w_2, \dots, w_r]$ 全局学习,另外一种是通过求解 r 个独立的加权 LS 问题 $w_i(i=1, 2, \dots, r)$ 的局部学习。然而,全局学习在估计过程中将后件参数独立对待,忽略了存在于模糊系统字典 Φ 中的分块结构。局部学习是对每一条模糊规则分别估计后件参数,其实质是一种结构化学习;然而所有的结果参数 $w_i(i=1, 2, \dots, r)$ 都未经选择就进行估计了,这就导致了所有的模糊规则都参与了模糊系统的建模。事实上,一些模糊规则是不必要的,删去它们并没有很显著的影响一个模型的效果。更为严重的是,过多的模糊规则难免会导致过拟合问题,并使得泛化效果变差。

为了克服上述问题,利用分块结构的优势,并利用一个 LS 算法对分块结构稀疏表示的 TSK 模糊系统进行建模,作为对传统稀疏表示的继承,分块结构稀疏表示首次在组 LASSO^[14] 中进行介绍和研究。它介绍了一种回归模型,在这种回归模型中,很多贡献小的分块的回归系数能够在保证模型精度较高的条件下削减接近于零。

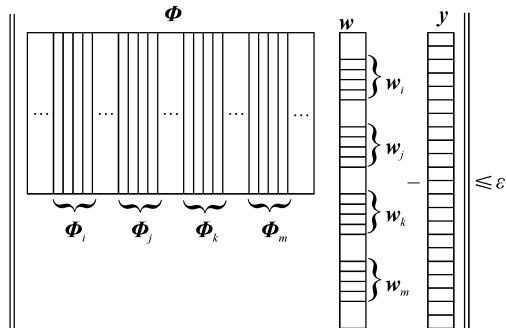


图 2 分块结构稀疏表示

Fig.2 Block-structured sparse representation

在 TSK 模糊系统的建模过程中,模糊规则的数目越多,模型的精度一般更高;但是,太多的模糊规则也会不可避免地将模糊系统变得复杂,甚至会导致过拟合问题。所以,模糊规则的选取变成了一个很关键的问题,要选取一个合理的参数化问题既要能控制规则数目也能够控制训练误差。利用 TSK 模糊模型的块结构的优势,找到了一种 TSK 模糊规则选择的块结构回归方法。在系统建模过程中,这种方法将许多贡献小的块的后件参数 $w_i(i=1, 2, \dots, r)$ 缩减至零。它们可以表示成如下优化问题

$$\begin{aligned} & \text{minimize } \|w\|_{2,0} \\ & \text{subject to } \frac{1}{2} \|y - \sum_{i=1}^r \Phi_i w_i\|_2^2 \leq \varepsilon \quad (11) \end{aligned}$$

式中: $\|w\|_{2,0} = \|(\|w_1\|_2, \|w_2\|_2, \dots, \|w_r\|_2)\|_0$ 实际上反映的是模糊规则的数目,并且 ε 表示模型精度的上限。优化问题(11)的目的是使得结果参数向量 w 尽可能的块稀疏并均衡模型的精度和复杂度。通过图 2 来说明这种块结构稀疏回归。

2.3 块正交匹配追踪算法应用于 TSK 模糊系统建模

在过去的研究中,优化问题(11)已被证明是 NP 难问题;然而,通常有 2 种方法求近似解:1) 松弛方法,例如 $\|\cdot\|_{2,1}$ 范数凸优化近似求解的一般方法;2) 分块(分组)贪婪选择算法,例如块正交匹配追踪算法(block OMP 算法)、块匹配追踪算法(OMP 算法)等。在本文中,使用块正交匹配追踪算法去解决优化问题(11),因为该算法在计算代价方面有优势。

作为 OMP 算法的一个继承,在稀疏回归时,分块 OMP 算法进行的是块变量的选择而不是对单个变量的选择。它以一种非常直观的方式运作。在每一步迭代中,根据减少的残差选择最优块。一旦块被选定,在已经选定的块上执行 LS 最小化获得对相应系数的估计。每次迭代结束,算法检查停止条件是否已经达到。这个停止条件通常是残差的最小范围或者是迭代的最大次数。在 TSK 模糊系统建模的框架中,因为块变量 w_i 是与第 i 条模糊规则的后件参数向量相联系的,故对块变量的选择,实际上就是对模糊规则的选择。

使用分块 OMP 算法(算法 2)对模糊规则进行选取的过程如下。

1) 输入:依据数据集 D 生成的 TSK 模糊系统字典 $\Phi=[\Phi_1, \Phi_2, \dots, \Phi_r]$ 。

2) 初始化:初始化标签集 $I^{(0)} = \emptyset$, 后件参数向量 $w^{(0)} = 0$, 残差 $r^{(0)} = y$ 。

3) 步骤:

FOR $K = 1, 2, \dots$

令 $j^{(k)} = \arg \max_j (r^{(k-1)})^T \Phi_j (\Phi_j^T \Phi_j)^{-1} \Phi_j^T r^{(k-1)}$

IF 终止条件为真, break

设定 $I^{(k)} = I^{(k-1)} \cup j^{(k)}$

通过 $w_{I^{(k)}}^{(k)} = \arg \min_{w_{I^{(k)}}} \|\Phi_{I^{(k)}} w_{I^{(k)}} - y\|_2$ 计算当前的后件参数向量 $w^{(k)}$ 并且设置 $w_{(I^{(k)})^c}^{(k)} = 0$

更新当前残差 $r^{(k)} = y - \Phi w^{(k)}$

4) 输出: $I^{(k)}$ 中选定的模糊规则子字典,以及模糊规则后件参数向量 w^k 。

在第 k 次迭代, $I^{(k)} \subset \{1, 2, \dots, r\}$ 表示已选模糊子字典的标签集; $w^{(k)} \in R^{r(n+1)}$ 表示模糊规则后件参数向量; $r^{(k)} = y - \Phi w^{(k)}$ 表示相应的残差。初始化后件参数向量为 $w^{(0)} = 0$, 残差为 $r^{(0)} = y$ 。 $\Phi_{I^{(k)}}$ 表示

对 Φ 进行限制的变量集是 $\{\Phi_i; i \in I^{(k)}\}$ 。

在第 k 次迭代中选取的最好的规则要使残差减少得越多越好,事实上,在第 k 次迭代中第 j 条模糊规则子字典产生的误差为

$$\varepsilon(j) = \|\Phi_j w_j - r^{(k-1)}\|_2^2 \quad (12)$$

为了将其最小化,则 $w_j^* = (\Phi_j^T \Phi_j)^{-1} \Phi_j^T r^{(k-1)}$, 此处, $j=1,2,\cdots,r$, 将 w_j^* 带入 $\varepsilon(j)$ 得

$$\begin{aligned} \varepsilon(j) &= \min_{w_j \in \mathbb{R}^{n+1}} \|\Phi_j w_j - r^{(k-1)}\|_2^2 = \\ &\|\Phi_j w_j^* - r^{(k-1)}\|_2^2 = \\ &\|\Phi_j (\Phi_j^T \Phi_j)^{-1} \Phi_j^T r^{(k-1)} - r^{(k-1)}\|_2^2 = \\ &\|r^{(k-1)}\|_2^2 - (r^{(k-1)})^T \Phi_j (\Phi_j^T \Phi_j)^{-1} \Phi_j^T r^{(k-1)} \end{aligned} \quad (13)$$

为了保证误差尽可能多地减少,在第 k 次迭代过程中,选择出的模糊规则子字典 Φ_{j^*} 要满足

$$j^* = \arg \max_j (r^{(k-1)})^T \Phi_j (\Phi_j^T \Phi_j)^{-1} \Phi_j^T r^{(k-1)} \quad (14)$$

在 $I^{(k)}$ 中已经选定的模糊规则子字典将不会再被选取,假设在 k 次迭代之后选取的模糊规则为 $R^{(k)}$, 那么通过式 (15) 估计后件参数向量 $w^{(k)}$:

$$w_{I^{(k)}}^{(k)} = \arg \min_{w_{I^{(k)}}} \|\Phi_{I^{(k)}} w_{I^{(k)}} - y\|_2 \quad w_{(I^{(k)})^c}^{(k)} = 0 \quad (15)$$

在算法 2 中 $(I^{(k)})^c$ 表示 $I^{(k)}$ 的补集。 $w_{I^{(k)}}^{(k)}$ 的最优值是通过将下面的二次导数置零获得:

$$\Phi_{I^{(k)}}^T (\Phi_{I^{(k)}} w_{I^{(k)}} - y) = -\Phi_{I^{(k)}}^T r^{(k)} = 0 \quad (16)$$

所以,所有在标签集 $I^{(k)}$ 中选择过的模糊规则子字典将不会再被式 (14) 选取。

2.4 已选模糊规则后件的稀疏正则化

本文提出了一个分块结构的稀疏回归模型来挑选一个 TSK 模糊模型最重要的模糊规则。在这一部分,稀疏表示进一步被应用于减少 TSK 模糊规则最小非零后件参数。通过块 OMP 算法选择出最重要的 k 条规则之后,可以用 Φ^k 表示模糊系统字典。最普通的用于寻找后件参数的方法是

$$\min \frac{1}{2} \|y - \Phi^k w\|_2^2 + \lambda \|w\|_0 \quad (17)$$

这里的参数 λ 是用于平衡模型的误差以及后件中非零参数的数目。不过已经证明了带有 $\|\cdot\|_0$ 范数的稀疏正则项的补偿函数是 NP 难的。一般情况下,会用平滑补偿函数的松弛算法对上述问题进行近似求解。在多种估计方法中,一种较为常用的方法就是将 $\|\cdot\|_0$ 范数改成 $\|\cdot\|_1$ 范数。这样,优化问题 (17) 被改写成

$$\min \frac{1}{2} \|y - \Phi^k w\|_2^2 + \lambda \|w\|_1 \quad (18)$$

上述优化问题在信号和图像处理领域有广泛和深入的研究,跟经典的 l_1 范数支持向量回归一样。有几种较为成熟的算法去解决上述问题,比如迭代重加权最小二乘算法、最小角回归算法以及迭代收缩算法,本文采用 l_1 范式最小二乘方法^[15] 对上述问题进行求解。

2.5 FCA-sparseTSK 应用于 TSK 模糊系统建模

在这一部分中,将算法 1 与块算法 2 相结合,提出了一种有效的 TSK 模糊系统建模的方法,称之为 FCA-sparseTSK(算法 3)。该模型在划分输入空间过程中采用了 FCA 聚类算法,对权重较小的特征进行去除,从而化简了模糊规则;在选择模糊规则过程中采用了 Block OMP 算法,挑选出较为重要的若干规则,去除了冗余规则,并利用稀疏表示的方法对后件参数进行估计。所以,通过 FCA-sparseTSK 建立的模糊系统的模糊规则数及非零后件参数的数目均得到了约减。

FCA-稀疏 TSK(FCA-sparseTSK)算法如下:

- 1) 输入:输入输出数据集 D 。
- 2) 特征优化:通过 FCA 聚类算法进行模糊划分并挑选出权重较大的特征生成 TSK 模糊系统字典。
- 3) 块结构稀疏:运行 Block OMP 算法选择出 k 条重要的模糊规则子字典并得到相应的类中心。
- 4) 优化学习:将上一步得到的类中心最为初始类中心,再次运行 FCA 聚类算法,并产生一个新的模糊系统字典。
- 5) 稀疏表示:根据稀疏表示对相应的模糊规则后件参数进行估计。
- 6) 输出:优化的 TSK 模糊系统。

3 实验

用数值实验证明所提出 FCA-sparseTSK 模型的有效性并与相关算法比较。在本实验中,将初始的模糊规则条数设置为 30。设输入输出数据集的形式如下所示:

$$D = \{[x_i^T, y_i]^T: x_i = [x_{i1}, x_{i2}, \cdots, x_{in}]^T, i = 1, 2, \cdots, N\}$$

根据相关实验的具体情况选择不同的评价指标,例如均方误差 $MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$ 和平均绝对误差 $MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$ 。这里 \hat{y}_i 表示 TSK 模糊系统对应变量 $x_i (i=1,2,\cdots,N)$ 的输出。

本文所有实验均对数据进行标准化处理,并采用五折交叉验证的方法对 FCA-sparseTSK 算法中所涉及到的参数进行寻优。算法中, m 和 α 的寻优范围设

置为 $(1,3]$, γ 的寻优范围为 $(0,2]$, 寻优步长均为 0.1。根据五折交叉验证,数据集被分成 5 个子集,对于每种算法,均进行 5 次建模。每次建模,取 5 个子集中的 1 个作为测试集,剩余的 4 个作为训练集。本文的算法与 genfis2、genfis3 以及 H-sparseFIS 算法^[12]相比较,验证 FCA-sparseTSK 算法的有效性。

3.1 合成数据集

为了体现本文方法的优势,利用下面生成函数产生数据集。对于任意输入 $\mathbf{x}_i=[x_{i1},x_{i2},x_{i3}]^T$,产生输出 $y_i=x_{i1}\sin(x_{i1})+x_{i2}\cos(x_{i2})+x_{i3}$ 。根据此方法,在取值范围 $x_{ik}\in[-10,10]$ ($k=1,2,3$) 内等间距取样,生成 2 000 个样本。为了测试本文提出算法对噪音和特征识别的能力,使用 normrnd(0,1,2 000,2) 生成 $2\,000\times 2$ 的噪声矩阵,将该矩阵与原矩阵合成为新的数据集,该数据集共计 2 000 个样本,5 个特征(后 2 个特征为噪声特征)。

经过 FCA 聚类算法,得到各样本特征的权重如图 3 所示,从图中显示的结果中发现,数据集中人工引入的两维噪声特征的权重都非常低,故可以将这 2 个特征去除,对模糊规则进行了优化,减少噪声特征对系统的干扰,提高系统的精度。另外,测试数据集的真实输出与模型输出的比较如图 4 所示。将 FCA-sparseTSK 和其他 3 种相关算法进行比较,结果如表 1 所示,表中列举出五折交叉验证中测试集的 MAE 并计算出平均 MAE 及其标准差。

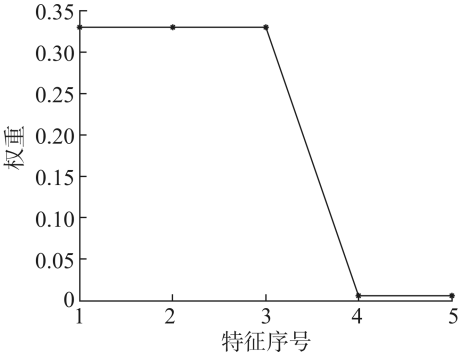
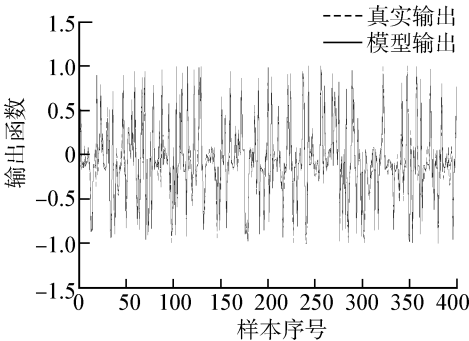
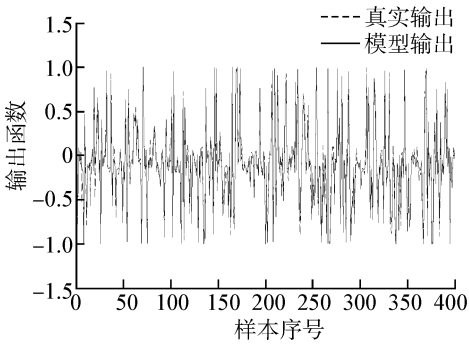


图 3 合成数据集样本特征权重

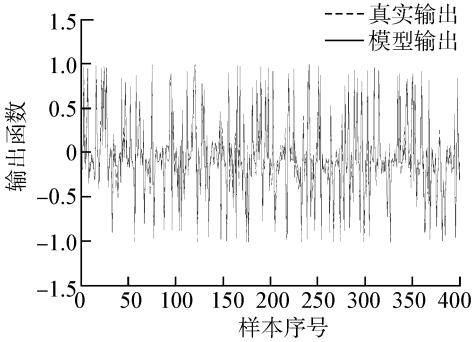
Fig.3 Feature weights of synthetic datasets



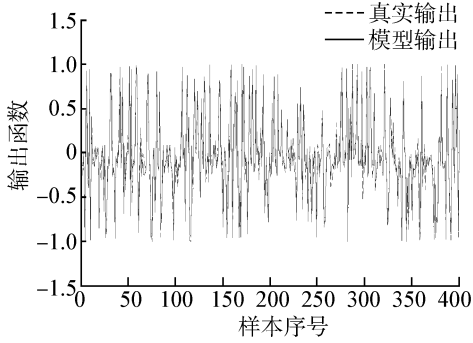
(a) 第 1 组交叉验证输出比较



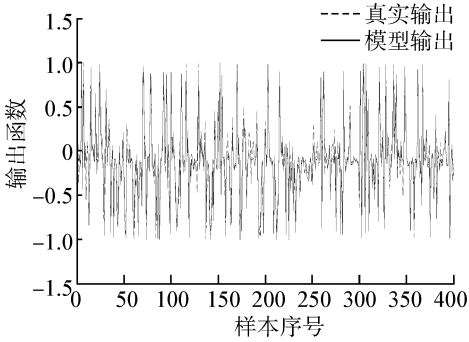
(b) 第 2 组交叉验证输出比较



(c) 第 3 组交叉验证输出比较



(d) 第 4 组交叉验证输出比较



(e) 第 5 组交叉验证输出比较

图 4 五折交叉验证测试数据集实际输出和模型输出比较

Fig.4 Comparison of actual output and model's output over testing data

从表 1 中能够发现,通过对样本特征的优化,去除掉两条噪声特征,不仅使得模糊规则得到化简,而且相对于其他算法,模型的精度更高;通过本文模型与未约减规则数的模型相比较,在保证模型精度基

本一致的前提下,本文算法平均约减规则 13.2 条,故对冗余规则的去除是有效的。

表 1 合成数据集各模型性能比较

Table 1 The performance of each model on synthetic datasets

算法	特征数	性能指标	1	2	3	4	5	均值
genfis2	5	MAE	0.048 4	0.045 6	0.067 1	0.061 8	0.044 8	0.053 5±0.010 2
		规则数	21	18	20	21	19	19.8
genfis3	5	MAE	0.277 3	0.266 4	0.285 2	0.302 3	0.278 9	0.282 0±0.013 2
		规则数	20	20	20	20	20	20
H-sparseFIS	5	MAE	0.044 7	0.048 5	0.075 8	0.088 4	0.082 4	0.068 0±0.020 0
		规则数	25	25	16	15	19	20
FCA-sparseTSK	3	MAE	0.011 2	0.010 7	0.010 8	0.011 7	0.011 6	0.011 2 ±0.000 5
		规则数	30	30	30	30	30	30
		MAE	0.014 3	0.012 7	0.015 9	0.018 6	0.016 2	0.015 5±0.002 2
		规则数	13	18	16	16	17	16.8

3.2 真实数据集

3.2.1 Airfoil Self-Noise 数据集

该数据集从 UCI machine learning repository 获取,记录了不同尺寸的 NACA 0012 机翼在不同风动速度和攻角下的性能,共计 1 503 个样本,5 条特征。输入分别为频率、攻角、弦长、自由流速度以及吸力面位移厚度,输出为声压等级。各算法在五折交叉验证中测试集的 MSE 及规则数目如表 2 所示。

表 2 结果显示,对于 airfoil self-noise 数据集,使用 FCA-sparseTSK 构建出的模型性能最好。本文所提出的模型相比于 genfis2、genfis3 和 H-sparseFIS 性能有所提升,并且 FCA-sparseTSK 去除掉 2 条不重要的特征,化简了模糊规则的复杂程度,使模糊系统变得简洁;通过与未约减规则数的实验相比较,本文算法平均约规则 19.2 条,且误差与未约减规则时基本一致,故对冗余规则的去除是有效的。

表 2 Airfoil self-noise 数据集各模型性能比较

Table 2 The performance of each model on airfoil self-noise datasets

算法	特征数	性能指标	1	2	3	4	5	均值
genfis2	5	MSE	0.050 3	0.054 7	0.052 3	0.056 5	0.045 8	0.051 9±0.004 2
		规则数	11	10	11	10	11	10.6
genfis3	5	MSE	0.057 9	0.067 5	0.064 0	0.074 6	0.064 3	0.065 7±0.006 1
		规则数	15	15	15	15	15	15
H-sparseFIS	5	MSE	0.042 3	0.048 8	0.045 5	0.059 3	0.048 2	0.048 8±0.006 4
		规则数	12	13	11	12	13	12.2
FCA-sparseTSK	3	MSE	0.045 3	0.047 4	0.039 4	0.051 6	0.042 4	0.045 2 ±0.004 7
		规则数	30	30	30	30	30	30
		MSE	0.040 4	0.049 7	0.035 4	0.053 2	0.039 5	0.043 6±0.007 5
		规则数	12	11	10	10	11	10.8

3.2.2 Machine CPU performance 数据集

该数据集从 knowledge extraction based on evolutionary learning(KEEL)获取,记录了相对 CPU 性能数据,共计 209 个样本,6 条特征。数据输入分别为机器周期时间(MYCT)、最小内存(MMIN)、最大内存(MMAX)、快速存储器(CACH)、最小通道(CHMIN)、最大通道(CHMAX),输出为 CPU 的相对性能(PRP)。各算法在五折交叉验证中测试集的 MAE 及规则数目如表 3 所示。

表 3 结果显示,对于 machine CPU performance 数据集,使用 FCA-sparseTSK 构建出的模型性能最好。本文所提出的模型相比于 genfis2、genfis3 和 H-sparseFIS 性能有所提升,并且 FCA-sparseTSK 去除掉 1 条不重要的特征,化简了模糊规则的复杂程度;通过与未约减规则数的实验相比较,本文算法平均约规则 25.4 条,且精度与未约减规则的模型相比略有提升,故对冗余规则的去除是有效的。

表 3 Machine CPU performance 数据集各模型性能比较

Table 3 The performance of each model on machine CPU performance datasets

算法	特征数	性能指标	1	2	3	4	5	均值
genfis2	6	MAE	0.056 9	0.046 9	0.070 7	0.056 9	0.074 8	0.061 2±0.011 4
		规则数	6	5	6	5	4	5.2
genfis3	6	MAE	0.074 6	0.058 3	0.076 8	0.069 1	0.091 0	0.074 0±0.011 9
		规则数	5	5	5	5	5	5
H-sparseFIS	6	MAE	0.056 9	0.035 7	0.048 9	0.052 7	0.060 0	0.050 8±0.009 4
		规则数	6	4	5	4	7	5.2
FCA-sparseTSK	5	MAE	0.053 2	0.036 7	0.062 8	0.048 6	0.055 9	0.051 4 ± 0.009 7
		规则数	30	30	30	30	30	30
		MAE	0.043 9	0.040 0	0.047 8	0.040 7	0.058 6	0.046 2±0.007 6
		规则数	4	4	5	4	6	4.6

3.2.3 Stock prices 数据集

该数据集从 KEEL 获取,由 10 个航空航天公司从 1988 年 1 月到 1991 年 10 月的日常股票价格,共计 950 个样本,9 条特征。用前 9 个公司的股票价

格作为输入,对第 10 个公司的股票价格进行估计。各算法在五折交叉验证中测试集的 MAE 集规则数目如表 4 所示。

表 4 Stock prices 数据集各模型性能比较

Table 4 The performance of each model on stock prices datasets

算法	特征数	性能指标	1	2	3	4	5	均值
genfis2	9	MAE	0.041 0	0.039 5	0.043 9	0.041 1	0.043 1	0.041 7±0.001 8
		规则数	21	22	21	21	24	21.8
genfis3	9	MAE	0.132 3	0.130 9	0.129 8	0.130 0	0.133 7	0.131 3±0.001 6
		规则数	20	20	20	20	20	20
H-sparseFIS	9	MAE	0.038 6	0.044 7	0.049 4	0.046 9	0.042 5	0.044 4±0.004 1
		规则数	22	17	20	22	22	20.6
FCA-sparseTSK	6	MAE	0.047 4	0.044 4	0.049 6	0.046 7	0.045 5	0.046 7 ± 0.002 0
		规则数	30	30	30	30	30	30
		MAE	0.047 3	0.047 6	0.047 1	0.047 2	0.047 2	0.047 6±0.000 7
		规则数	17	17	18	17	17	17.6

表 4 结果显示,对于 stock 数据集,genfis2 构建出的模型性能最好。本文所提出的模型与 genfis2 和 H-sparseFIS 相比性能大体一致,但 FCA-sparseTSK 去除掉 3 条不重要的特征,大大简化了模糊规则的复杂程度,使模糊系统变得简洁;通过与不约减规则数的实验相比较,本文算法平均约减掉 12.4 条规则,且误差与未约减规则时基本持平,故对冗余规则的去除是有效的。

4 结束语

本文提取出存在于 TSK 模糊系统中的分块结构信息,将模糊系统建模问题转化为一个结构化的稀疏恢复问题。FCA-sparseTSK 算法在构建模糊系统的过程中,对样本特征进行优化,选取重要的模糊规则并对规则的后件参数进行估计。在合成数据集

以及真实数据集上的实验证明,FCA-sparseTSK 模型能够在保证系统性能的前提下,同时简化模糊规则并消减冗余模糊规则。

参考文献:

[1] DING B. Dynamic output feedback predictive control for nonlinear systems represented by a Takagi-Sugeno model [J]. IEEE Transactions on Fuzzy Systems, 2011, 19(5): 831-843.

[2] JANG J S R. ANFIS: adaptive-network-based fuzzy inference system[J]. IEEE Transactions on Systems, Man and Cybernetics, 1993, 23(3): 665-685.

[3] QIN Hao, YANG S X. Adaptive neuro-fuzzy inference systems based approach to nonlinear noise cancellation for images[J]. Fuzzy Sets and Systems, 2007, 158(10): 1036-1063.

- [4] WANG Di, ZENG Xiaojun, KEANE J A. An evolving-construction scheme for fuzzy systems[J]. IEEE Transactions on Fuzzy Systems, 2010, 18(4): 755-770.
- [5] LUGHOFFER E. Evolving fuzzy systems-methodologies, advanced concepts and applications[M]. Heidelberg: Springer, 2011: 81-85.
- [6] JAIN A K. Data clustering: 50 years beyond K-means[J]. Pattern recognition letters, 2010, 31(8): 651-666.
- [7] TSEKOURAS G E. On the use of the weighted fuzzy c -means in fuzzy modeling[J]. Advances in Engineering Software, 2005, 36(5): 287-300.
- [8] YAGER R R, FILEV D P. Generation of fuzzy rules by mountain clustering[J]. Journal of Intelligent and Fuzzy Systems, 1994, 2(3): 209-219.
- [9] GATH I, GEVA A B. Unsupervised optimal fuzzy clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989, 11(7): 773-780.
- [10] GUSTAFSON D E, KESSEL W C. Fuzzy clustering with a fuzzy covariance matrix[C]//IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes. San Diego, USA, 1978: 761-766.
- [11] LUGHOFFER E. Extensions of vector quantization for incremental clustering[J]. Pattern Recognition, 2008, 41(3): 995-1011.
- [12] LUO Minnan, SUN Fuchun, LIU Huaping. Hierarchical structured sparse representation for T-S fuzzy systems identification[J]. IEEE Transactions on Fuzzy Systems, 2013, 21(6): 1032-1043.
- [13] 皋军, 王士同. 具有特征排序功能的鲁棒性模糊聚类方法[J]. 自动化学报, 2009, 35(2): 145-153.
- GAO Jun, WANG Shitong. Fuzzy clustering algorithm with

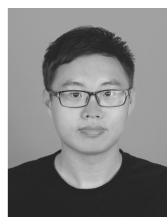
ranking features and identifying noise simultaneously[J].

Acta Automatica Sinica, 2009, 35(2): 145-153. [14]

YUAN Ming, LIN Yi. Model selection and estimation in regression with grouped variables[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2006, 68(1): 49-67.

[15] KIM S J, KOH K, LUSTIG M, et al. A interior-point method for large-scale ℓ_1 -regularized least squares[J]. IEEE Journal on Selected Topics in Signal Processing, 2007, 1(4): 606-617.

作者简介:



张佳骥,男,1990年生,硕士研究生,主要研究方向为人工智能与模式识别、模糊系统。



蒋亦樟,男,1988年生,博士研究生,主要研究方向为人工智能与模式识别、模糊系统。



王士同,男,1964年生,教授,博士生导师,主要研究方向为人工智能、模式识别和生物信息。

[责任编辑:刘畅]