第 47 卷 第 2 期 自 动 化 学 报 Vol. 47, No. 2
2021 年 2 月 ACTA AUTOMATICA SINICA February, 2021

结构辨识和参数优化协同学习的概率 TSK 模糊系统

顾晓清 [1] 倪彤光 [1] 张 聪 [1] 戴臣超 [1] 王洪元 [1]

摘 要 传统 Takagi-Sugeno-Kang (TSK) 模糊系统的结构辨识和参数优化往往分阶段进行, 同时模糊规则数需要预先设定, 因此 TSK 模糊系统的逼近性能和解释性往往不理想. 针对此问题, 提出了一种结构辨识和参数优化协同学习的概率 TSK 模糊系统 (Probabilistic TSK fuzzy system, PTSK). 首先, PTSK 使用概率模型表示模糊回归系统, 将结构辨识和参数优化作为一个整体来考虑. 其次, PTSK 不借助于专家经验, 使用粒子滤波方法对规则数和前后件参数协同学习, 得到系统全部参数的最优解. 实验结果表明, PTSK 具有良好的逼近性能, 同时能获得较少的模糊规则数.

关键词 Takagi-Sugeno-Kang 模糊系统, 概率模型, 回归, 粒子滤波

引用格式 顾晓清, 倪彤光, 张聪, 戴臣超, 王洪元. 结构辨识和参数优化协同学习的概率 TSK 模糊系统. 自动化学报, 2021, **47**(2): 349−362

DOI 10.16383/j.aas.c180298

Probabilistic TSK Fuzzy System in the Simultaneous Learning of Structure Identification and Parameter Optimization

GU Xiao-Qing[1] NI Tong-Guang[1] ZHANG Cong[1] DAI Chen-Chao[1] WANG Hong-Yuan[1]

Abstract Most of existing Takagi-Sugeno-Kang (TSK) fuzzy systems identify their structure and estimate antecedent/consequent parameter of rules in a stepwise manner. In addition, the number of fuzzy rules generally has to be decided manually. Therefore, the performance and interpretability of TSK fuzzy systems are often not satisfactory. In order to overcome these shortcomings, in this paper, a novel probabilistic TSK fuzzy system called probabilistic TSK fuzzy system (PTSK) is constructed from the perspective of probabilistic model. Firstly, PTSK uses a likelihood probability to characterize fuzzy regression tasks, while considering structure identification and parameter optimization as a whole. Secondly, without any prior expert experience, PTSK adopts a particle filter method to simultaneously learn the number of rules and antecedent/consequent parameter of fuzzy rules. According to the experimental results on several datasets, PTSK obtains good approximation performance and a relatively small number of fuzzy rules.

Key words Takagi-Sugeno-Kang fuzzy system, probabilistic model, regression, particle filter

Citation Gu Xiao-Qing, Ni Tong-Guang, Zhang Cong, Dai Chen-Chao, Wang Hong-Yuan. Probabilistic TSK fuzzy system in the simultaneous learning of structure identification and parameter optimization. *Acta Automatica Sinica*, 2021, **47**(2): 349−362

模糊推理系统 (Fuzzy inference system, FIS) 以模糊集合和模糊推理为基础, 能够将自然语言直接转译成计算机语言, 使得机器具有表达模糊语意的能力, 目前被广泛应用在时间序列分析、工业控制和故障诊断等方面[1−2]. 相比大多数智能模型, FIS 的优势在于: 1) FIS 具有很强的面向不确定系统的构建能力, 能模拟人类专家知识和推理的不确定性; 2) 不像 SVM 和神经网络等被视为一个黑箱, FIS 具有良好的基于规则的解释性; 3) FIS 具有强大的学习能力, 能利用模糊逻辑较强的结构性知识表达, 也可以像神经网络等模型一样利用数据集信息对模型参数进行优化学习. 由 Takagi, Sugeno 和 Kang 提出的 Takagi-Sugeno-Kang (TSK) 模糊系统, 又称 TSK 模糊模型, 因其结构简单和逼近能力强, 是一种常用的 FIS 工具[2]. TSK 模糊系统使用具有解释性的 "IF-THEN" 规则来定义系统的规则库, 规则库的构建工作由 2 部分组成: 规则结构的辨识和规则参数的优化. 规则结构的辨识指为系统的输入空间找到合适的模糊划分; 规则参数的优化则指确定模糊规则前件和后件的参数. 其中, 选择合适的模糊规则数是结构辨识的核心工作[3]. 模糊规则数过多

收稿日期 2018-05-10 录用日期 2018-12-12
Manuscript received May 10, 2018; accepted December 12, 2018
国家自然科学基金 (61806026, 61572085), 江苏省自然科学基金 (BK20180956), 江苏省教育科学 "十三五" 规划 2018 年度课题 (B-a/2018/01/41) 资助
Supported by National Natural Science Foundation of China (61806026, 61572085), National Natural Science Foundation of Jiangsu Province (BK20180956), Topics of the 13th Five-Year Plan for Educational Science in Jiangsu Province in 2018 (B-a/2018/01/41)
本文责任编委 许斌
Recommended by Associate Editor XU Bin
1. 常州大学信息科学与工程学院 常州 213164
1. School of Information Science and Engineering, Changzhou University, Changzhou 213164

会导致模糊系统复杂化, 易产生过拟合现象; 模糊规则数过少则导致系统逼近性能不佳.

目前, 确定模糊规则数最简单的方法是基于网格的输入空间划分法. 特征数是 d 的数据集, 如使用固定 m 网格的输入空间划分法, 共提取到 m^d 条模糊规则数. 显然这一方法不适用于高维数据[4]. 确定模糊规则数的另一类常用方法是聚类算法[5−6], 聚类法 TSK 模糊系统的一大优点是能获得较小规模的规则数, 但模糊规则数往往需要预先设定, 如文献 [5−7] 使用交叉验证的方法获得模糊规则数的最优值. 虽然一些聚类有效性指标如 Xie-Beni 指标和 Mountainpotential 指标等能用于聚类数的选择, 但这些有效性指标用于确定模糊规则数时往往效果不佳[8]. 此外, 聚类法 TSK 模糊系统在优化模糊规则的前件和后件参数时往往分阶段计算, 这种学习策略的优点是时间复杂度相对较低, 但其存在一个严重的缺陷: 无法捕捉输入空间和输出空间之间的内在联系, 因此得到的 TSK 模糊系统的逼近性能往往达不到最优. 为解决这一问题, 近年来一些学者开始研究前件和后件参数的联合学习方法, 如文献 [9] 使用迭代线性支持向量回归机来联合学习前件和后件参数, 文献 [10] 建立了前件和后件参数联合学习的贝叶斯推理模型, 并使用 Metropolis-Hastings (MH) 采样方法求解参数的最优解. 然而这两个算法仍需事先设定模糊规则数.

众所周知, 模糊理论和概率模型是常用于描述复杂问题不确定的两类方法. 但两者的侧重点不同: 模糊理论能较好地描述自然语言的不确定性, 即语义的不确定性; 概率模型能较好地描述由系统固有偶然性或变异性带来的随机不确定性, 即系统性能或预测结果的不确定性[11]. Zadeh 在文献 [11] 中首次提出了 "概率和模糊互补多于竞争" 这一思想, 认为两者通过协同学习可以提高系统的性能. 受这一思想启发, 本文提出了一种结构辨识和参数优化协同学习的概率 TSK 模糊系统 (Probabilistic TSK fuzzy system, PTSK). PTSK 的核心思想是将数据的输入/输出空间、系统结构和规则参数作为一个整体来考虑, 并基于概率理论使用概率模型来构建模糊回归系统. 不借助于专家经验, 基于最大后验概率估计 (Maximum-a-posteriori, MAP), PTSK 使用粒子滤波方法[12] 同时得到模糊规则数、规则前件/后件参数的最优解. PTSK 模糊系统的优点有: 1) 以一种协同学习的形式构建了基于概率模型的 TSK 模糊系统. 该系统兼具统计学和模糊逻辑的优点, 能有效处理非线性回归问题. 2) 不同于传统聚类法 TSK 模糊系统使用 "黑盒" 策略 (如网格搜索法) 优化模糊规则数的方法, PTSK 无需任何专家经验, 使用粒子滤波方法能自动学习模糊规则的所有参数. 3) PTSK 充分挖掘数据集的整体特征, 同时考虑输入空间和输出空间对模糊规则参数的影响. 实验结果表明 PTSK 兼具强解释性和良好逼近性能的特点.

1 相关工作

1.1 TSK 模糊系统基本概念

TSK 模糊系统规则库中的第 k 个模糊规则可用以下形式表示:

$$R^k: \text{IF } x_1 \text{ is } A_{k1},\ x_2 \text{ is } A_{k2}, \cdots, x_d(k) \text{ is } A_{kd},$$
$$\text{THEN } f_k(\boldsymbol{x}) = v_{k0} + \sum_{j=1}^{d} v_{kj}x_j, k = 1, \cdots, K \quad (1)$$

其中 $x_1, x_2, \cdots, x_d$ 是输入向量 $\boldsymbol{x}$ 的 d 维分量, $\boldsymbol{A}_k$ ($\boldsymbol{A}_k = [A_{k1}, A_{k2}, \cdots, A_{kd}]^{\rm T}$) 是输入空间的模糊子集, K 是模糊规则数. 令 $\boldsymbol{v}_k = [\boldsymbol{v}_{k0}, \boldsymbol{v}_{k1}, \cdots, \boldsymbol{v}_{kd}]^{\rm T}$, $\tilde{\boldsymbol{x}} = [1, \boldsymbol{x}^{\rm T}]^{\rm T}$, 模糊规则 THEN 部分的 $f_k(\boldsymbol{x})$ 可以写成:

$$f_k(\boldsymbol{x}) = \boldsymbol{v}_k^{\rm T}\tilde{\boldsymbol{x}} \quad (2)$$

令 $\mu_k(\boldsymbol{x})$ 是第 k 条模糊规则的隶属度函数, 其值可由各维对应的隶属度值通过合取操作获得,

$$\mu_k(\boldsymbol{x}) = \prod_{i=1}^{d} \mu_{\boldsymbol{A}_{ki}}(x_i) \quad (3)$$

$\tilde{\mu}_k(\boldsymbol{x})$ 为第 k 条模糊规则对应的归一化后的隶属度函数, 其值可由下式计算得到:

$$\tilde{\mu}_k(\boldsymbol{x}) = \frac{\mu_k(\boldsymbol{x})}{\sum\limits_{k'=1}^{K} \mu'_k(\boldsymbol{x})} \quad (4)$$

若引入高斯函数作为隶属度函数, 式 (3) 中 $\mu_{\boldsymbol{A}_{ki}}(x_i)$ 可表示为:

$$\mu_{\boldsymbol{A}_{ki}}(x_i) = \exp\left(\frac{-\|x_i - c_{ki}\|^2}{\delta_{ki}}\right) \quad (5)$$

其中隶属度函数的中心 c_{ki} 和方差 δ_{ki} 被称为模糊规则的前件参数. TSK 模糊系统的实值输出 $\hat{y}$ 为:

$$\hat{y} = \sum_{k=1}^{K} \tilde{\mu}_k(\boldsymbol{x}) f_k(\boldsymbol{x}) \quad (6)$$

1.2 聚类法 TSK 模糊系统的参数学习

表 1 比较了聚类法 TSK 模糊系统中常用的规则前件/后件参数学习方法. 系统中每一个聚类划分转化为一条模糊规则. 此时式 (5) 中的隶属度函数

表 1 聚类法 TSK 模糊系统中常用的模糊规则前件/后件参数学习方法

Table 1 The common learning methods for the antecedent/consequent parameters in the clustering based TSK fuzzy system

前件参数学习		
FCM 模糊聚类[5]	优点	获得的空间划分具有模糊性, 算法实现简单
	缺点	聚类数需要预先设定
Gustafson-Kessel[13] 和 Gath-Geva 聚类[14]	优点	均使用诱导矩阵识别数据集的结构
	缺点	矩阵计算量较大
One-pass 聚类[15]	优点	数据集只需要遍历一遍即可完成空间划分, 适用于增量或在线学习模式
	缺点	对凸型数据分布识别较差; 且与聚类数有关的参数需要预先设定
后件参数学习		
最小二乘法[3, 8]	优点	最常用; 显式地得到后件参数的解析解, 计算简单
	缺点	对噪声数据敏感
支持向量回归机[15]	优点	保证参数的全局最优解, 逼近性能较强
	缺点	二次规划问题求解的计算量较大
进化计算[4]	优点	模拟自然随机优化算法, 不依赖对象的数学模型
	缺点	遗传编码的选择较难解决, 时间复杂度较高
反向传播算法[16]	优点	在小数据集上很快达到局部最优解
	缺点	收敛较慢, 不适用于大规模数据集

中心 $\boldsymbol{c}_k$ 为聚类中心, 方差 δ_k 可由下式计算得到:

$$\delta_{ki}^2=\sum_{\boldsymbol{x}\in\boldsymbol{c}_k}\frac{(x_{ji}-c_{ki})^2}{N_k},\quad i=1,\cdots,d,\ k=1,\cdots,K \tag{7}$$

其中 N_k 是第 k 个聚类中样本的个数. 若采用最小二乘法求解后件, TSK 模糊系统的目标函数可写成:

$$\min_{\boldsymbol{V}}\ \sum_{i=1}^N(y_i-\hat{y}_i)^2=\sum_{i=1}^N(y_i-\boldsymbol{V}^{\rm T}\phi(\boldsymbol{x}))^2=(\boldsymbol{Y}-\phi(\boldsymbol{X})\boldsymbol{V})^{\rm T}(\boldsymbol{Y}-\phi(\boldsymbol{X})\boldsymbol{V}) \tag{8}$$

其中 $\phi(\boldsymbol{x})=[(\mu_1(\boldsymbol{x})\boldsymbol{x}_e)^{\rm T},\ \cdots,(\mu_K(\boldsymbol{x})\boldsymbol{x}_e)^{\rm T}]^{\rm T}$, $\boldsymbol{x}_e=[1,\ \boldsymbol{x}^{\rm T}]^{\rm T}$. 后件参数矩阵 $\boldsymbol{V}$ 则可通过下式求解得到:

$$\boldsymbol{V}=(\phi(\boldsymbol{X})^{\rm T}\phi(\boldsymbol{X}))^{-1}\phi(\boldsymbol{X})^{\rm T}\boldsymbol{Y} \tag{9}$$

2 概率 TSK 模糊系统 (Probabilistic TSK Fuzzy System, PTSK)

2.1 PTSK 联合概率模型

给定输入数据集 $\boldsymbol{X}=\{\boldsymbol{x}_i,i=1,2,\cdots,N,\boldsymbol{x}_i\in\mathbf{R}^d\}$ 和对应的输出集 $\boldsymbol{Y}=\{y_i,i=1,2,\cdots,N,y_i\in\mathbf{R}\}$, PTSK 关于输入/输出数据、规则数和前件/后件参数的联合概率表示为 $p(\boldsymbol{X},K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$, 其中 4 个待优化参数分别是模糊规则数 K, 聚类中心矩阵 $\boldsymbol{C}$, 模糊划分矩阵 $\boldsymbol{U}$ 和后件参数矩阵 $\boldsymbol{V}$. 根据贝叶斯概率, $p(\boldsymbol{X},K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$ 可以表示为:

$$p(\boldsymbol{X},K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})=p(\boldsymbol{X}|K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})\cdot p(\boldsymbol{U}|K,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})p(\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}|K)p(K) \tag{10}$$

式 (10) 由 4 个因子构成. 下面对这 4 个因子展开叙述.

1) $p(\boldsymbol{X}|K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$: 聚类法 TSK 模糊系统中每一个聚类对应一条模糊规则. 此时, 条件似然 $p(\boldsymbol{X}|K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$ 仅与聚类数 K, 模糊隶属度矩阵 $\boldsymbol{U}$ 和聚类中心矩阵 $\boldsymbol{C}$ 有关, 等价于 $p(\boldsymbol{X}|K,\boldsymbol{U},\boldsymbol{C})$. PTSK 假设 $\boldsymbol{x}_n$ 的先验是 K 个正态分布的乘积, 正态分布的中心为聚类中心 $\boldsymbol{c}_k$, 协方差是其模糊隶属度分量 u_{nk} 构成的单位阵 $\frac{K\boldsymbol{I}}{u_{nk}^m}$, 即

$$p(\boldsymbol{x}_n|K,\boldsymbol{U},\boldsymbol{C})=\prod_{k=1}^K\mathcal{N}\left(\boldsymbol{x}_n|\mu=\boldsymbol{C}_k,\ \boldsymbol{\Lambda}=\frac{K\boldsymbol{I}}{u_{nk}^m}\right)=\frac{1}{2\pi^{\frac{d}{2}}}\prod_{k=1}^K\left(\frac{u_{nk}^m}{K}\right)^{\frac{d}{2}}\exp\left(-\frac{1}{2K}u_{nk}^m\|\boldsymbol{x}_n-\boldsymbol{C}_k\|^2\right) \tag{11}$$

其中 m 是模糊指数. 输入数据 $\boldsymbol{X}$ 中的每个样本都满足独立同分布, 条件似然 $p(\boldsymbol{X}|K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$ 可表示为全部样本的正态分布先验的乘积, 即

$$p(\boldsymbol{X}|K,\boldsymbol{U},\boldsymbol{Y},\boldsymbol{C})=\prod_{n=1}^N p(\boldsymbol{x}_n|K,\boldsymbol{U},C)\cdot$$

$$\propto \left(\prod_{n=1}^{N}\prod_{k=1}^{K}\left(\frac{u_{nk}^{m}}{K}\right)^{\frac{d}{2}}\right)\cdot$$
$$\exp\left(-\frac{1}{2K}\sum_{n=1}^{N}\sum_{k=1}^{K}u_{nk}^{m}\|\boldsymbol{x}_n-\boldsymbol{c}_k\|^2\right) \tag{12}$$

2) $p(\boldsymbol{U}|K,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$: 模糊划分矩阵 $\boldsymbol{U}$ 仅与聚类个数 K 和聚类中心矩阵 $\boldsymbol{C}$ 有关, 且每个样本对应的模糊隶属度相互独立, 因此 $\boldsymbol{U}$ 的条件似然可以写成 $p(\boldsymbol{U}|K,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{Y})=\prod_{n=1}^{N}p(\boldsymbol{U}_n|K,\boldsymbol{C})$. 狄利克雷 (Dirichlet) 分布是一种多变量连续概率分布, 其每个分量均大于 0 且每一维度之和为 1. 文献 [17] 使用狄利克雷分布来构造模糊聚类的模糊隶属度. 拉普拉斯 (Laplace) 分布较正态分布在中心点处有较高的峰度, 文献 [18] 使用拉普拉斯分布来提高聚类模型的稀疏性. 因此, PTSK 在模糊划分矩阵 $\boldsymbol{U}$ 的条件似然中同时考虑狄利克雷和拉普拉斯分布, $p(\boldsymbol{U}|K,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$ 可以写成:

$$p(\boldsymbol{u}_n|K,\boldsymbol{C})=\left(\prod_{k=1}^{K}\left(\frac{u_{nk}^{m}}{K}\right)^{-\frac{d}{2}}\right)\mathrm{Dirichlet}(\boldsymbol{u}_n|\alpha)^{\frac{1}{K}}\cdot$$
$$\prod_{k=1}^{K}\mathrm{Laplace}\left(u_{nk}|1,\frac{1}{\beta}\right)^{\frac{1}{K}} \tag{13}$$

式 (13) 第 1 项用于消去式 (12) 中的 $\prod_{k=1}^{K}(u_{nk}^{m}/K)^{d/2}$ 项. 狄利克雷分布 $\mathrm{Dirichlet}(\boldsymbol{u}_n|\alpha)$ 的形式为:

$$\mathrm{Dirichlet}(\boldsymbol{u}_n|\alpha)=\frac{1}{\boldsymbol{B}(\alpha)}\prod_{k=1}^{K}u_{nk}^{\alpha_k-1} \tag{14}$$

其中 $\alpha=[\alpha_1,\alpha_2,\cdots,\alpha_K]^{\mathrm{T}}$ 是狄利克雷参数, $\boldsymbol{B}(\alpha)$ 是一个正常数. 第 3 项 K 维拉普拉斯分布的形式为:

$$\prod_{k=1}^{K}\mathrm{Laplace}\left(u_{nk}|1,\frac{1}{\beta}\right)^{\frac{1}{K}}=$$
$$\prod_{k=1}^{K}\left(\frac{1}{\beta}\exp(-\beta\,|u_{nk}-1|)\right)^{\frac{1}{K}}=$$
$$\frac{\beta}{2}\exp(-\beta)\exp\left(\frac{\beta}{K}\right) \tag{15}$$

其中 β 是拉普拉斯分布的尺度参数. 因为 u_{nk} 在 [0, 1] 之间且满足 $\sum_{k=1}^{K}u_{nk}=1$, 化简后其值在上式中被消去. 对式 (13) ～ (15) 进行整理, $p(\boldsymbol{U}|K,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})$ 可以写成以下形式

$$p(\boldsymbol{U}|K,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})=\prod_{n=1}^{N}p(\boldsymbol{u}_n|K,\boldsymbol{C})\propto$$
$$\prod_{n=1}^{N}\left[\left(\prod_{k=1}^{K}\left(\frac{u_{nk}^{m}}{K}\right)^{-\frac{d}{2}}\right)\mathrm{Dirichlet}(\boldsymbol{u}_n|\alpha)^{\frac{1}{K}}\right]\cdot$$
$$\left(K^{-\frac{Kd}{2}}\right)\exp\left(\frac{\beta N}{K}\right) \tag{16}$$

3) $p(\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}|K)$: 在给定模糊规则数的情况下, PTSK 在条件似然 $p(\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}|K)$ 中考虑系统在 K 条模糊规则上的平均估计误差, 即

$$p(\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V}|K)=\exp\left(-\frac{1}{2K}\sum_{n=1}^{N}(y_n-\hat{y}_n)^2\right) \tag{17}$$

其中 $\hat{y}_n$ 是 PTSK 模糊系统中第 n 个样本的实际输出. 指数函数 $\exp(-x^2)$ 呈现单调递减性, 当 $\sum_{n=1}^{N}(y_n-\hat{y}_n)^2$ 趋近于最小值时, 使用最大后验概率估计能得到系统参数的最优值.

4) $p(K)$: 模糊规则数是正整数, 其服从离散分布. 其分布可采用两种方法: 一种是假设模糊规则数服从离散均匀分布, 即 $p(K)=\mathrm{discrete}(K)=1/l$, 其中 l 是区间内离散值的个数. 此时模糊规则数的选取等价于网格搜索法. 另一种是使用泊松过程或泊松分布[12]. 因为离散均匀分布的区间上界不易设定, 本文设定模糊规则数的先验分布服从泊松分布:

$$p(K)=\mathrm{Poisson}(K,\lambda)=\frac{\lambda^K\exp(-\lambda)}{K!} \tag{18}$$

其中 λ 是形状参数, 参照文献 [19], $\lambda=\lg N$.

将式 (12), (16), (17) 和 (18) 相乘, 得到 PTSK 关于数据、规则数和前件/后件参数的联合概率模型:

$$p(\boldsymbol{X},K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})=p(\boldsymbol{X}|K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})\cdot$$
$$p(\boldsymbol{U}|K,\boldsymbol{C},\boldsymbol{Y},\boldsymbol{V})p(\boldsymbol{C},\boldsymbol{V},\boldsymbol{Y}|K)p(K)\propto$$
$$\exp\left(-\frac{1}{2K}\sum_{n=1}^{N}\sum_{k=1}^{K}u_{nk}^{m}||\boldsymbol{x}_n-\boldsymbol{c}_k||^2\right)\cdot$$
$$\prod_{n=1}^{N}\prod_{k=1}^{K}(u_{nk}^{\alpha_k-1})^{\frac{1}{K}}\exp\left(\frac{\beta N}{K}\right)\cdot$$
$$\exp(-\frac{1}{2K}\sum_{n=1}^{N}(y_n-\hat{y}_n)^2)\frac{\lambda^K}{K!}\exp(-\lambda) \tag{19}$$

对上式取自然对数, 可以得到 PTSK 的目标函数:

$$J = -\frac{1}{2K}\sum_{n=1}^{N}\sum_{k=1}^{K}u_{nk}^{m}||\boldsymbol{x}_n-\boldsymbol{c}_k||^2+\frac{1}{K}\left(\sum_{n=1}^{N}\sum_{k=1}^{K}(\alpha_k-1)\lg u_{nk}\right)+\frac{\beta N}{K}-\frac{1}{2K}\sum_{n=1}^{N}(y_n-\hat{y}_n)^2+K\lg\lambda-\sum_{k=1}^{K}\lg k \quad (20)$$

从式 (19) 和 (20) 可以看成, PTSK 将数据的输入/输出空间、规则数的识别和前件/后件参数的优化视作一个整体, 得到的规则数和前件/后件参数一定是相互依赖、密切相关的. 当式 (20) 联合概率模型达到 MAP 值时, PTSK 中参数 $\{K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{V}\}$ 同时得到最优解.

2.2　PTSK 结构辨识和参数优化的协同学习

粒子滤波方法是一种序贯蒙特卡罗方法, 使用带权值的随机粒子按照序贯重要性采样方法递归估计状态变量的后验概率分布. 粒子滤波方法还具有噪声容忍性强和模型初始状态不敏感的优点. PTSK 采用粒子滤波方法求得式 (20) 的最大后验估计. PTSK 结构辨识和参数优化的协同学习示意图如图 1 所示.

从图 1 可以看出, PTSK 结构辨识和参数优化的协同学习由一系列迭代过程构成. 在算法初始化阶段, PTSK 创建一组带权重的离散粒子 $WDP[r]=\{\{K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{V}\},ll\}^{\rm T}$ $(r=1,2,\cdots,P)$, 其中 ll 是式 (20) 对应的目标函数的值, P 是粒子数. PTSK 设置各粒子模糊规则数 K 的初值为 1, 模糊隶属度矩阵 $\boldsymbol{U}$ 的元素初值为 $u_{i1}=1$ $(i=1,2,\cdots,N)$, 聚类中心矩阵 $\boldsymbol{C}$ 的元素初值 c_{i1} 为输入数据 $\boldsymbol{X}$ 的均值. PTSK 创建候选最优粒子集 $CAND$, 设置 $CAND[1]=WDP[1]$. 下面详细介绍第 n 次迭代过程, 其主要由 4 个步骤构成:

1) 采样. 使用式 (18) 对模糊规则数 K^* 进行采样, 其泊松分布的均值为当前规则数 K. 如果新采样的模糊规则数 K^* 小于当前规则数 K, 从聚类中心矩阵 $\boldsymbol{C}$ 中随机选择 K^* 个中心作为当前聚类中心矩阵; 如果新采样的模糊规则数 K^* 大于当前规则数 K, 则保留当前聚类中心矩阵 $\boldsymbol{C}$, 并使用 d 维拉普拉斯分布采样 K^*-K 个新聚类中心

$$\boldsymbol{c}_k\sim\prod_{i=1}^{d}{\rm Laplace}(e_i,\gamma)=\prod_{i=1}^{d}\frac{1}{2\gamma}\exp\left(-\frac{|e_i|}{\gamma}\right) \quad (21)$$

其中拉普拉斯分布的位置参数 $\boldsymbol{e}$ 是输入数据 $\boldsymbol{X}$ 的均值, γ 是尺度参数. 经过大量实验, γ 取值为 5.

2) 参数优化. 这一步骤的工作是根据采样得到的模糊规则数 K^* 对参数 $\{\boldsymbol{U},\boldsymbol{C},\boldsymbol{V}\}$ 进行优化.

a) 优化模糊隶属度矩阵 $\boldsymbol{U}$. 随着模糊规则数 K 的变化, 模糊隶属度矩阵 $\boldsymbol{U}$ 也相应变化. 由于对模糊隶属度的取值无先验知识, PTSK 假设模糊隶属度服从平坦型狄利克雷 (Flat Dirichlet) 分布[20], 此时狄利克雷分布参数 α 中各分量为 1. 狄利克雷分布能保证所采样的模糊隶属度元素满足 $u_{nk}\geq 0$ 且 $\sum_{k=1}^{K}u_{nk}=1$, 将其作为约束条件改造式 (20) 可得:

$$J' = -\frac{1}{2K}\sum_{n=1}^{N}\sum_{k=1}^{K}u_{nk}^{m}||\boldsymbol{x}_n-\boldsymbol{c}_k||^2+\frac{\beta N}{K}-\frac{1}{2K}\sum_{n=1}^{N}\sum_{k=1}^{K}\tilde{w}_{k,n}(y_n-\boldsymbol{v}_k^{\rm T}\tilde{\boldsymbol{x}}_n)^2+K\lg\lambda-\sum_{k=1}^{K}\lg k+\sum_{n=1}^{N}\eta_n(\sum_{k=1}^{K}u_{nk}-1) \quad (22)$$

其中 η_n 是拉格朗日乘子. 上式得到极值的必要条件为 $\partial J'/\partial u_{nc}=0$, 可得 u_{nc} 的解析解:

$$u_{nk}=\frac{\|\boldsymbol{x}_n-\boldsymbol{c}_k\|^{\frac{2}{1-m}}}{\sum\limits_{k=1}^{K}\|\boldsymbol{x}_n-\boldsymbol{c}_k\|^{\frac{2}{1-m}}} \quad (23)$$

当前隶属度函数的宽度矩阵 δ 和后件参数矩阵 $\boldsymbol{V}$ 可以分别通过式 (7) 和式 (9) 计算得到.

b) 优化聚类中心矩阵 $\boldsymbol{C}$. 固定模糊规则数 K 和模糊隶属度矩阵 $\boldsymbol{U}$, 此时式 (20) 得到极值的必要条件为 $\partial J/\partial c_{ki}=0$, 可得 c_{ki} 的解析解:

$$c_{ki}=\frac{\dfrac{\sum\limits_{n=1}^{N}u_{nk}^{m}x_{ni}}{K}+2\sum\limits_{n=1}^{N}(y_n-\hat{y}_n)\times\dfrac{f_k(\boldsymbol{x}_n)\times\sum\limits_{k=1}^{K}\mu_k(\boldsymbol{x}_n)-\sum\limits_{k=1}^{K}f_k(\boldsymbol{x}_n)\times\mu_k(\boldsymbol{x}_n)}{(\sum\limits_{k=1}^{K}\mu_k(\boldsymbol{x}_n))^2}\times\mu_k(\boldsymbol{x}_n)\times\dfrac{x_{ni}}{\delta_i^k}}{\dfrac{\sum\limits_{n=1}^{N}u_{nk}^{m}}{K}+2\sum\limits_{n=1}^{N}(y_n-\hat{y}_n)\times\dfrac{f_k(\boldsymbol{x}_n)\times\sum\limits_{k=1}^{K}\mu_k(\boldsymbol{x}_n)-\sum\limits_{k=1}^{K}f_k(\boldsymbol{x}_n)\times\mu_k(\boldsymbol{x}_n)}{(\sum\limits_{k=1}^{K}\mu_k(\boldsymbol{x}_n))^2}\times\mu_k(\boldsymbol{x}_n)\times\dfrac{1}{\delta_i^k}} \quad (24)$$

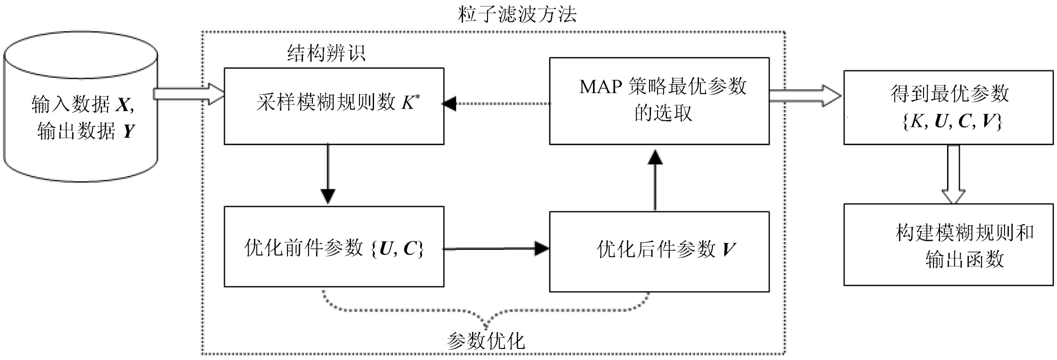


图 1 PTSK 结构辨识和参数优化的协同学习示意图

Fig. 1 The diagram of simultaneous learning of structure identification and parameter optimization in PTSK

c) 优化后件参数矩阵 $\boldsymbol{V}$. 优化模糊划分矩阵 $\boldsymbol{U}$ 和聚类中心矩阵 $\boldsymbol{C}$ 后, 此时式 (20) 得到极值的必要条件为 $\partial J/\partial \boldsymbol{V}=0$, 可得 $\boldsymbol{V}$ 的解析解的形式与式 (9) 相同.

d) 计算粒子对应的目标函数值. 在优化了每个粒子对应的参数 $\{K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{V}\}$ 后, 每个粒子对应的目标函数值 ll 可通过计算式 (20) 得到.

3) 粒子更新. 检查粒子集 WDP 中的每个粒子的 ll 值能否提高当前模糊规则数对应的目标函数值, 如果是, 则将该粒子替换当前模糊规则数的候选最优粒子 $CAND[K]$, 并加入到粒子集 $CAND$ 中, 如果不是, 则保留 $CAND[K]$, 即

$$CAND[K]=\begin{cases}WDP[r], & 若\ \ WDP[r].ll>CAND[K].ll\\ CAND[K], & 其他\end{cases}\tag{25}$$

然后使用 WDP 和 $CAND$ 构建粒子集 PS,

$$PS=\{WDP,CAND\}\tag{26}$$

4) 权重计算和重采样. 为了减少粒子退化的影响, PTSK 根据粒子权重执行重采样操作, 更新粒子集 WDP. 每一个粒子的权重值 w_i 的计算式为

$$w_i=\frac{\exp(PS[i].ll)}{\sum\limits_{i=1}^{|PS|}\exp(PS[i].ll)}\tag{27}$$

其中 $|PS|$ 表示粒子集 PS 中粒子的个数. 每个粒子的重采样概率与权重值 w_i 成正比, 从粒子集 PS 中重采样 P 个粒子并使用它们更新粒子集 WDP. 这样 WDP 中权重小的粒子被剔除掉, 权重大的粒子被保留, 甚至被多遍复制. 最终, 在经过若干次迭代后, 粒子集 WDP 中最大 ll 值对应的粒子的参数 $\{\boldsymbol{U},\boldsymbol{C},\boldsymbol{V}\}$ 为 PTSK 模糊系统的最优参数.

2.3 算法描述和分析

首先给出 PTSK 模糊系统的构建算法描述, 如算法 1 所示.

在重采样步骤中, 相比传统粒子滤波方法[21] 仅使用固定规模的粒子集 WDP, PTSK 使用不固定规模的粒子集 PS, PS 粒子集由 2 部分构成: 粒子集 WDP 和 $CAND$, 其中 $CAND$ 由采样得到的不同模糊规则数的候选最优粒子构成, 粒子数不固定. 粒子集 $CAND$ 的作用是进一步减小粒子退化的影响, 加快算法的收敛.

接下来, 我们分析 PTSK 模糊系统的收敛性. 在实际应用中, 粒子数的规模有限, 系统需要在收敛速度和系统性能之间进行平衡. 因此, PTSK 在算法的终止条件上除了设置最大迭代次数外, 还计算当前目标函数值与上一次迭代目标函数值之间的差值, 若其值小于阈值 ε, 则统计其次数. 当累计次数超过设定值 $miter$ 时, 算法终止. 因为随着迭代次数的增加, 模糊规则数趋于固定, 规则的前件/后件参数也仅在最优值附近微调, 此时使用式 (17) 计算全部规则上的平均估计误差几乎不变, 系统的性能趋于稳定. 因此, 依据文献 [22] 粒子滤波方法求解系统静态参数可得局部最优解的结论, 算法 1 也可保证所得模糊规则数和规则前件/后件参数的最优解是局部最优解. 最后, 我们分析 PTSK 模糊系统的时间复杂度. 由算法 1 步骤可知, PTSK 模糊系统的时间复杂度主要集中在对参数 $\{\boldsymbol{U},\boldsymbol{C},\boldsymbol{V}\}$ 的优化部分. 使用式 (23) 优化模糊隶属矩阵 $\boldsymbol{U}$ 的时间复杂度是 $\mathrm{O}(NK)$. 使用式 (24) 优化聚类中心 $\boldsymbol{c}_k$ 的时间复杂度是 $\mathrm{O}(NK^2(d+1))$. 基于平均估计误差的模糊规则后件参数的时间复杂度是 $\mathrm{O}(N^3)$. 因此, 算法 1 执行单次迭代的时间复杂度为 $\mathrm{O}(P(NK+K^2N(d+1)+N^3))$, 其中 N, K, d 和 P 分别表示训练样本个数, 模糊规则数, 样本维数和粒子数.

算法 1. PTSK 模糊系统的构建

//初始化

1) 创建粒子集 WDP, 设置 $K=1$, 分别使用式 (21) 和 (23) 初始化 c 和 u;

2) 创建粒子集 $CAND$, 设置 $CAND[1]=WDP[1]$;

3) 设置迭代次数 $t=1$, $r=1$;

Repeat $t=t+1$;

Repeat $r=r+1$;

//采样模糊规则数 K

4) 用式 (25) 更新 $K*$;

//优化参数 $\{\boldsymbol{U},\boldsymbol{C},\boldsymbol{V}\}$

5) 更新 $WDP[r].K=K^{*}$;

6) 使用式 (23) 计算 $WDP[r].\boldsymbol{U}$;

7) 使用式 (24) 计算 $WDP[r].\boldsymbol{C}$;

8) 使用式 (9) 计算 $WDP[r].\boldsymbol{V}$;

9) 使用式 (20) 计算 $WDP[r].ll$;

//粒子更新

10) 使用式 (25) 得到候选最优粒子集 $CAND$;

11) 构建粒子集 $PS=\{WDP, CAND\}$;

Until $r>P$

//权重计算和重采样

12) 使用式 (27) 计算 PS 中粒子的权重值 w_i;

13) 以 w_i 为重采样概率, 在 PS 中重采样 P 个粒子, 并更新粒子集 WDP;

Until $t\geq t_{\max}$ 或者 $count(\left|\max\{PS[r_1].ll\}_t-\max\{PS[r_2].ll\}_{t-1}\right|<\varepsilon)>miter, (r_1,r_2=1,2,\cdots,|PS|)$;

14) 选择 $CAND$ 中 ll 值最大的粒子, 得到 $\{K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{V}\}$ 最优解;

//构建模糊规则

15) 使用最优解 $\{K,\boldsymbol{U},\boldsymbol{C},\boldsymbol{V}\}$ 构建模糊规则, 并由式 (6) 得到输出函数.

3　实验研究

为了验证本文方法的有效性, 本节将通过 28 个回归数据集对 PTSK 模糊系统进行分析与验证. 实验安排如下: 第 3.1 节对数据集和实验的设置进行了介绍; 第 3.2 节分析了 PTSK 和 7 种对比算法在 28 个数据集上的实验结果; 第 3.3 节对实验结果进行了统计分析; 最后给出 PTSK 收敛性和参数敏感性分析.

3.1　实验设置

表 2 列出了实验中使用的 28 个回归数据集的基本信息. mexihat、abalone、housing、mg 和 bodyfat 来自 LIBSVM 数据集[23], gc-s、gc-x 和 gc-p 数据集来自文献 [6], 其余数据集来自 KEEL 数据集[24]. 实验比较了两类共 7 种回归算法, 一类是 TSK 模糊系统: L2-TSK-FS[5], TSK-IRL-R[25], MOGUL-TSK-R[26] 和 B-ZTSK-FS[10]. 另一类是经典的回归算法: WM-R[27], ENSEMBLE-R[28] 和 PSVR[29]. L2-TSK-FS, B-ZTSK-FS, PSVR 和 PTSK 使用 MATLAB2016b 实现; 其余方法使用 KEELtoolbox 软件实现[24]. 实验中各算法参数的设置如表 3 所示, 7 种对比算法的参数设置均使用相应文献的默认设置. 参数的选取使用 5 重交叉验证的方法. 实验采用 3 个评价指标: 1) 均方误差 MSE (Mean squared error) 和方差; 2) 平均训练时间; 3) TSK 模糊系统的平均规则数. 同时, 为了使得性能对比更具有统计意义, 本文采用无参统计学方法中的 Friedman 检验[30] 和 Host 事后检验[31] 进行统计测试. 本文全部实验在 Intel i7-3770 CPU 3.4 GHz, 16 GB RAM, Windows 7 的环境下执行.

表 2　数据集基本信息

Table 2　Basic information of datasets

数据集	规模	维数	数据集	规模	维数
abalone	4 177	8	gc-x	56 000	6
anacalt	4 052	7	gc-p	56 000	6
autompg6	392	5	housing	506	13
autompg8	392	7	mexihat	2 500	2
bodyfat	252	14	mg	1 385	6
compactiv	8 192	21	mortgage	1 049	15
concrete	1 030	8	plastic	1 650	2
dee	365	6	pole	14998	26
delta-ail	7 129	5	puma32h	8 192	32
delta-elv	9 517	6	quake	2 178	3
diabert	43	2	stock	950	9
elevators	16 599	18	treasury	1 049	15
friedman	1 200	5	wankara	1 609	9
gc-s	56 000	6	wizmir	1 461	9

3.2　实验结果分析

实验比较了 PTSK 和另外 7 种回归算法在 28 个数据集上的实验结果, 表 4 和表 5 分别显示了各算法的 MSE (标准差) 和平均训练时间, 表 6 比较了 6 种 TSK 模糊系统得到的平均模糊规则数.

1) 从表 4 可以看出, PTSK 在所有数据集上取得了令人满意的 MSE 值, PTSK 在 28 个数据集上胜出 17 次, ENSEMBLE-R、PSVR、TSK-IRL-R 和 WM-R 各胜出 2、5、3 和 1 次. L2-TSK-FS 和 B-ZTSK-FS 采用网格搜索法求得模糊规则最优解, 但设定合适的规则数搜索区域较困难, 因而 L2-TSK-FS 和 B-ZTSK-FS 的平均 MSE 值不理想. 由此可知挖掘数据输入空间和输出空间内在联系对 TSK 模糊系统的性能有着重要的影响, 亦说明 PTSK 采用的基于概率模型的结构辨识和优化参数的协同学习机制有利于找到合适的模糊规则数.

2) 从表 5 可以看出, 8 种回归算法中 MOGUL-R 和 TSK-IRL-RPTSK 训练时间最长, PSVR 和 L2-TSK-FS 训练时间最短, PTSK 在样本规模较大

表 3 算法参数设置

Table 3 Parameter setting

算法	参数设置
L2-TSK-FS	模糊规则数 $K \in \{2^2, 3^2, \cdots, 11^2\}$, 尺度参数 $h \in \{0.2^2, 0.4^2, \cdots, 2^2\}$, 模糊指数 $m = 2$, 正则化参数 $C \in \{2^{-4}, 2^0, \cdots, 2^7\}$.
TSK-IRL-R	适应性函数最小匹配度 = 1.5, 种群数 = 61, 交叉概率 = 0.1, 种群比例 =0.2.
MOGUL-R	正类样本的匹配度参数 = 0.05, 负类样本的允许比例参数 = 1.5, 适应性函数最小匹配度 = 0.1, 种群数 = 15.
B-ZTSK-FS	模糊规则数 $K \in \{2^2, 3^2, \cdots, 11^2\}$, 尺度参数 $h \in \{0.2^2, 0.4^2, \cdots, 2^2\}$, 模糊指数 $m = 2$, 狄利克雷参数 = 1 000.
WM	标签数 = 5.
ENSEMBLE	隐层数 = 2, 隐结点数 = 15, 学习系数 = 0.15, 动量系数 = 0.1, 集成方法类型: BEM, 网络数 = 10.
PSVR	正则化参数 $C \in \{10^{-3}, 10^{-2}, \cdots, 10^3\}$, 高斯核核参 $\sigma \in \{10^{-3}, 10^{-2}, \cdots, 10^3\}$.
PTSK	模糊指数 $m = 2$, 最大迭代次数 10^3, 阈值 $\varepsilon = 10^{-3}$, 收敛阈值 $miter = 50$, 稀疏参数 $\beta \in \{1, 2, \cdots, 8\}$, 粒子数 $P = 10$.

表 4 8 种算法在 28 个数据集上的 MSE (标准差) 比较

Table 4 MSE (Standard deviation) comparison of 8 algorithms on 28 datasets

数据集	ENSEMBLE-R	PSVR	WM-R	MOGUL-R	TSK-IRL-R	L2-TSK-FS	B-ZTSK-FS	PTSK
mexihat	0.0205	0.0222	0.0219	0.0207	0.0231	0.0226	0.0239	**0.0204**
	1.004×10^{-3}	1.023×10^{-3}	1.011×10^{-3}	1.025×10^{-3}	1.017×10^{-3}	1.003×10^{-3}	1.020×10^{-3}	1.001×10^{-3}
abalone	4.1738	5.0107	6.6225	4.7323	5.8258	5.1987	4.1186	**4.0810**
	0.383	0.376	0.301	0.273	0.326	0.298	0.213	0.196
anacalt	0.0531	0.0478	0.0595	0.0521	0.0829	0.0542	0.0496	**0.0435**
	1.452×10^{-3}	1.023×10^{-3}	4.520×10^{-3}	1.102×10^{-3}	4.239×10^{-3}	2.358×10^{-3}	2.102×10^{-3}	1.404×10^{-3}
autompg6	11.8797	11.8827	13.3320	14.1473	**8.2574**	13.6705	12.3298	11.0676
	4.831	2.681	1.821	8.553	1.325	1.964	1.759	1.553
autompg8	8.6182	8.2351	8.7713	9.6884	**6.6799**	9.9872	8.8537	8.2122
	1.131	1.756	1.098	1.205	0.975	1.210	1.026	1.003
bodyfat	5.1200×10^{-4}	$\mathbf{2.0201}\times10^{-4}$	5.8424×10^{-4}	9.9749×10^{-4}	4.4523×10^{-4}	5.2704×10^{-4}	5.0125×10^{-4}	3.4402×10^{-4}
	1.711×10^{-5}	2.310×10^{-5}	6.418×10^{-5}	6.528×10^{-5}	3.245×10^{-5}	4.718×10^{-5}	3.610×10^{-5}	2.036×10^{-5}
compactiv	**34.8932**	37.6408	35.7480	37.1190	37.4625	39.6401	38.3607	36.8515
	4.544	5.008	4.646	5.121	4.673	4.786	4.145	4.006
concrete	52.6524	**50.5003**	56.2245	55.1906	58.2542	58.8612	55.8678	50.5285
	3.535	4.388	6.541	3.325	3.875	3.764	3.757	3.400
dee	0.2524	0.2627	**0.2296**	0.3034	0.6920	0.2956	0.2765	**0.2294**
	0.024	0.032	0.086	0.184	0.085	0.030	0.027	0.021
delta-elv	2.7764×10^{-6}	2.2719×10^{-6}	3.3007×10^{-6}	3.5924×10^{-6}	$\mathbf{1.9863\times10^{-6}}$	4.3942×10^{-6}	3.5075×10^{-6}	2.1432×10^{-6}
	6.892×10^{-7}	5.231×10^{-7}	3.390×10^{-7}	6.743×10^{-7}	4.721×10^{-7}	5.432×10^{-7}	5.121×10^{-7}	3.976×10^{-7}
delta-ail	3.5813×10^{-8}	3.5842×10^{-8}	5.7242×10^{-8}	3.6223×10^{-8}	2.9804×10^{-8}	3.7946×10^{-8}	5.1313×10^{-8}	$\mathbf{2.7236\times10^{-8}}$
	2.017×10^{-9}	4.654×10^{-9}	8.523×10^{-9}	7.487×10^{-9}	6.754×10^{-9}	5.987×10^{-9}	6.000×10^{-9}	3.003×10^{-9}
diabert	0.4932	0.5785	0.7266	0.6225	0.9208	1.1231	0.7334	**0.4883**
	0.259	0.236	0.699	0.354	0.435	0.500	0.465	0.398
elevator	2.5947×10^{-4}	2.5398×10^{-4}	$\mathbf{2.3237\times10^{-4}}$	5.6023×10^{-4}	5.7413×10^{-4}	7.6498×10^{-4}	6.5429×10^{-4}	5.5531×10^{-4}
	2.646×10^{-5}	1.658×10^{-5}	1.765×10^{-5}	2.991×10^{-5}	3.102×10^{-5}	3.832×10^{-5}	3.801×10^{-5}	3.124×10^{-5}
friedman	2.7852	2.2769	3.1595	2.1445	3.0082	3.0603	2.4206	**2.1408**
	0.352	0.615	0.371	0.201	0.308	0.312	0.311	0.353
gc-s	0.5734	0.6024	0.4216	0.5015	0.2601	0.3267	0.2602	**0.2304**
	0.064	0.078	0.060	0.078	0.012	0.038	0.014	0.010
gc-x	4.6492×10^{-3}	4.9121×10^{-3}	4.8530×10^{-3}	4.8000×10^{-3}	3.5912×10^{-3}	3.8955×10^{-3}	3.4279×10^{-3}	$\mathbf{3.2687\times10^{-4}}$
	3.042×10^{-5}	3.550×10^{-5}	3.706×10^{-5}	3.743×10^{-5}	3.001×10^{-5}	3.330×10^{-5}	3.328×10^{-5}	2.004×10^{-5}
gc-p	0.0826	0.0980	0.0900	0.0900	0.0807	0.0856	0.0754	**0.0717**
	2.998×10^{-3}	3.005×10^{-3}	3.071×10^{-3}	3.026×10^{-3}	3.053×10^{-3}	3.251×10^{-3}	3.117×10^{-3}	3.010×10^{-3}

表 4 8 种算法在 28 个数据集上的 MSE (标准差) 比较 (续表)

Table 4 MSE (Standard deviation) comparison of 8 algorithms on 28 datasets (Continued table)

数据集	ENSEMBLE-R	PSVR	WM-R	MOGUL-R	TSK-IRL-R	L2-TSK-FS	B-ZTSK-FS	PTSK
housing	**29.6062**	33.7403	34.8514	30.4763	34.9782	37.5164	34.0108	33.0672
	6.899	7.041	6.948	6.389	5.839	5.214	5.317	5.215
mg	0.0203	0.0214	0.0179	0.0163	0.0166	0.0188	0.0176	**0.0157**
	1.873×10^{-3}	1.431×10^{-3}	1.351×10^{-3}	1.572×10^{-3}	1.313×10^{-3}	1.082×10^{-3}	1.277×10^{-3}	1.139×10^{-3}
mortgage	0.0843	0.0448	0.0925	0.6160	0.0881	0.0589	0.0409	**0.0407**
	3.985×10^{-3}	2.751×10^{-3}	3.618×10^{-3}	1.916×10^{-2}	2.936×10^{-3}	2.643×10^{-3}	2.603×10^{-3}	2.517×10^{-3}
plastic	2.6657	2.3495	2.3646	2.3735	2.8642	2.9098	2.8477	**2.2153**
	0.401	0.098	0.446	0.110	0.100	0.218	0.231	0.200
pole	206.5032	203.5998	229.1911	207.0983	233.7895	225.5987	216.9002	**200.9751**
	4.167	5.531	6.258	4.003	5.638	4.980	4.678	4.236
puma	6.2417×10^{-3}	$\mathbf{2.3405\times10^{-3}}$	4.6104×10^{-3}	5.8963×10^{-3}	4.3104×10^{-3}	4.7403×10^{-3}	4.7612×10^{-3}	3.7208×10^{-3}
	7.835×10^{-6}	7.230×10^{-6}	7.200×10^{-6}	7.737×10^{-6}	6.875×10^{-6}	6.943×10^{-6}	7.032×10^{-6}	7.053×10^{-6}
quake	0.0591	**0.0350**	0.0538	0.0371	0.0461	0.0570	0.0532	0.0356
	4.752×10^{-3}	1.995×10^{-3}	6.863×10^{-3}	2.700×10^{-3}	3.286×10^{-3}	3.274×10^{-3}	2.863×10^{-3}	2.965×10^{-3}
stock	1.5008	**1.1681**	1.3863	1.6803	1.5430	1.6518	1.7626	1.1938
	0.321	0.477	0.459	0.287	0.300	0.348	0.372	0.265
treasury	0.4287	0.4562	0.4199	0.5553	0.5421	0.4198	0.4568	**0.4124**
	0.102	0.021	0.122	0.099	0.078	0.063	0.067	0.066
wankara	1.6955	2.0040	2.6569	2.8312	2.7434	1.8889	1.7313	**1.6096**
	0.409	0.143	0.588	0.164	0.296	0.302	0.228	0.199
wizmir	2.1475	1.6275	2.2321	2.2404	2.3343	1.9649	1.9772	**1.5245**
	0.467	0.146	0.502	0.161	0.278	0.222	0.234	0.211

表 5 8 种算法在 28 个数据集上的平均训练时间 (s) 的比较

Table 5 Comparison of the average training time (s) of 8 algorithms on 28 datasets

数据集	ENSEMBLE-R	PSVR	WM-R	MOGUL-R	TSK-IRL-R	L2-TSK-FS	B-ZTSK-FS	PTSK
mexihat	18.82	5.03	20.36	50.14	160.23	**4.25**	90.23	70.18
abalone	21.31	**3.00**	28.22	5 238.68	7 192.52	3.17	51.89	40.35
anacalt	36.94	**2.31**	27.58	1 006.54	60.48	2.45	22.31	18.96
autompg6	8.38	**1.37**	7.14	320.26	70.69	1.89	8.15	6.38
autompg8	8.55	**1.55**	8.01	309.65	392.45	1.81	20.64	9.10
bodyfat	7.10	0.65	7.71	420.68	856.98	**0.52**	6.69	5.98
compactiv	44.88	**28.38**	98.39	4 124.81	4 792.17	30.34	729.93	640.63
concrete	30.95	**5.01**	41.30	1 100.41	8 649.22	5.33	72.66	41.80
dee	10.43	**0.36**	9.31	30.82	89.47	0.93	9.39	5.61
delta-elv	29.47	**3.56**	35.72	965.03	7536.12	3.88	79.38	51.69
delta-ail	30.10	**5.78**	27.26	1 167.60	2 546.60	6.06	109.63	69.61
diabert	7.81	0.45	3.64	24.10	16.935	**0.43**	1.96	1.74
elevator	186.64	**30.07**	181.18	3 508.54	3 286.65	100.64	828.80	559.20
friedman	53.26	**2.81**	47.96	124.85	1 003.47	2.95	68.036	31.47
gc-s	245.27	79.34	268.96	3 976.33	4 024.20	**79.11**	964.45	520.76
gc-x	252.85	78.02	270.82	4 034.87	4 035.65	**78.00**	968.02	518.32
gc-p	248.38	80.13	268.40	3 956.67	4 020.55	**80.05**	966.27	523.44
housing	20.37	**3.87**	13.86	398.50	445.20	4.12	75.82	47.83
mg	32.54	**3.25**	23.54	786.92	765.32	3.56	60.43	23.67
mortgage	20.15	**3.88**	22.16	1 039.27	1238.48	4.02	68.26	57.20
plastic	22.74	**4.79**	23.58	305.20	268.09	5.07	120.3	78.64
pole	96.36	**40.86**	720.44	6 290.63	8 533.63	42.75	1075.56	713.70
puma	40.58	**30.05**	130.83	4 903.38	4 893.30	32.70	1073.12	796.57
quake	28.57	**2.54**	19.80	360.25	400.39	2.97	63.92	18.48
stock	19.43	**3.50**	25.32	1 490.34	2 175.56	3.89	78.46	22.46
treasury	19.45	**5.83**	18.21	1 202.40	2 543.82	6.02	117.20	78.30
wankara	28.76	**4.38**	27.33	2 009.35	2 464.50	4.85	164.64	79.43
wizmir	19.29	**4.02**	16.36	1 344.52	2 032.38	4.21	100.28	70.35

表 6 6 种 TSK 模糊系统在 28 个数据集上的平均模糊规则数比较

Table 6 Comparison of the average number of fuzzy rules of six TSK fuzzy systems on 28 datasets

数据集	WM-R	MOGUL-R	TSK-IRL-R	L2-TSK-FS	B-ZTSK-FS	PTSK
mexihat	8.8	10.4	28.2	6.8	6.6	**4.2**
abalone	217.6	114.4	6434.0	16.0	9.0	**6.0**
anacalt	124.6	313.6	185.4	16.0	25.0	**4.8**
autompg6	117.0	81.2	786.0	36.0	36.0	**18.2**
autompg8	182.0	380.0	2658.0	36.0	25.0	**6.6**
bodyfat	190.6	101.2	1715.2	9.0	6.0	**3.2**
compactiv	1 599.6	536.2	2097.0	25.0	16.0	**11.0**
concrete	309.2	360.4	1 497.2	49.0	36.0	**8.6**
dee	161.4	112.2	3051.4	36.0	36.0	**34.6**
delta-elv	708.8	220.6	6 510.0	36.0	25.0	**5.8**
delta-ail	241.8	104.6	1 476.6	25.0	36.0	**8.8**
diabert	16.4	32.8	22.8	25.0	16.0	**8.4**
elevator	4286.7	801.0	191	25.0	25.0	**23.8**
friedman	767.8	432.2	3 043.2	25.0	16.0	**6.6**
gc-s	62.8	40.2	226.2	16.0	9.0	**6.4**
gc-x	60.8	43.2	220.8	16.0	9.0	**6.2**
gc-p	61.4	40.0	218.8	16.0	9.0	**6.4**
housing	291.2	288.4	2 673.0	49.0	49.0	**40.0**
mg	240.0	175.0	3 887.0	9.0	9.0	**5.4**
mortgage	198.2	62.8	122.6	25.0	16.0	**14.4**
plastic	14.8	97.4	87.8	49.0	25.0	**21.4**
pole	3 228.8	100.2	1 775.0	36.0	36.0	**23.4**
puma	6 553.4	188.0	3 221.0	81.0	64.0	**60.8**
quake	54.2	173.4	985.8	36.0	**25.0**	29.0
stock	264.8	80.6	578	36.0	36.0	**8.6**
treasury	197.2	63.6	70.0	49.0	36.0	**35.6**
wankara	458.6	127.8	836.0	25.0	25.0	**22.0**
wizmir	413.8	119.3	189.4	36.0	**25.0**	**25.0**

的数据集上训练时间也较长. 但由表 3 可以看出, PTSK 需要寻优的参数只有 1 个, 而 L2-TSK-FS 需要寻优的参数有 3 个, 因此, 在实际应用中 L2-TSK-FS 在训练时间上并不具备优势. PSVR 是支持向量机算法, 虽然训练时间较短, 但它不具有 FIS 的语义性和解释性. 另外, TSK-IRL-R, MOGUL-TSK-R, WM-R 和 ENSEMBLE-R 通过 KEEL toolbox 平台实现, 该平台使用 Java 软件实现. 由于相同代码在 MATLAB 平台的运行时间比在 Java 平台的时间时间慢大约 6 倍左右[32], 因此, ENSEMBLE-R 和 WM-R 在训练时间上也并不比 PTSK 具有优势.

3) 从表 6 可以看出, 与对比的 5 种 TSK 模糊系统相比, PTSK 仅需少量的模糊规则就能取得良好的分类效果, 说明 PTSK 有较强的解释性. 尽管 B-ZTSK-FS 使用 MH 采样方法能够同时学习模糊规则的前件和后件参数, 但是该方法不能自动学习模糊规则数, 因此 B-ZTSK-FS 构建的模糊规则数多于 PTSK.

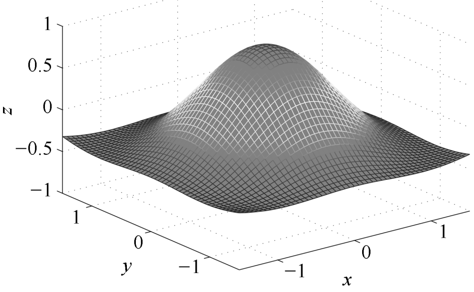
图 2 显示了 PTSK 在 mexihat 数据集上某一折的实验结果, 图 3 对应显示了 PTSK 在 mexihat 数据集上得到的 4 条模糊规则的模糊集示意图. 从图 2 可以看出, PTSK 取得了良好的逼近性能. 我们知道, 规则库的解释性与模糊规则数有关, 另一方面, 规则库的解释性也与模糊子集的清晰度有关. 从图 3 可以看出模糊集具有语义解释性, 由此可得对应的 4 条模糊规则也具有较高的解释性.

3.3 算法分析

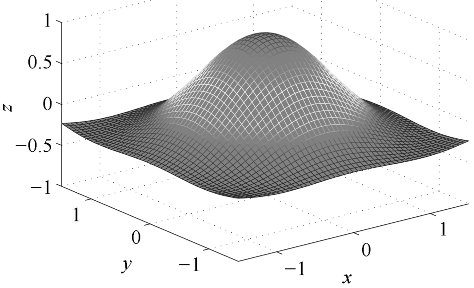
3.3.1 参数敏感性实验和收敛性分析

首先对 PTSK 的参数敏感性进行分析. 稀疏因子 β 使用 5 折交叉验证的方法得到最优值. 实验设置 β 的搜索范围是 $\{1, 2, \cdots, 8\}$. 由于篇幅所

限, 表 7 显示了在 mexihat, elevators, bodyfat 和 wizmir 数据集上 β 参数对 MSE 指标和模糊规则数的影响.



(a) 原始数据集
(a) Original dataset



(b) PTSK 实验结果
(b) Experimental result of PTSK

图 2　PTSK 在 mexihat 数据集上的实验结果

Fig. 2　Experimental results of PTSK on mexihat dataset

由表 7 的结果可以看出: 1) 稀疏因子 β 对模糊规则数起到了决定性作用. 对于绝大多数数据集而言, β 值越大获得的模糊规则数越少; 反之, β 值越小获得的模糊规则数越大. 只有极个别数据集上的模糊规则数受 β 值的影响不大, 其原因是该数据集在模糊空间分布较紧密且聚类划分结果清晰. 2) PTSK 模糊系统的回归性能与模糊规则数密切相关, 模糊规则数较大时 MSE 值达不到最优, 此时易发生过拟合的现象; 模糊规则数较小时 MSE 值也达不到最优. 因此, 稀疏因子 β 起到了平衡系统性能和复杂度的作用, 对 β 在使用交叉验证的方法寻优是必要的.

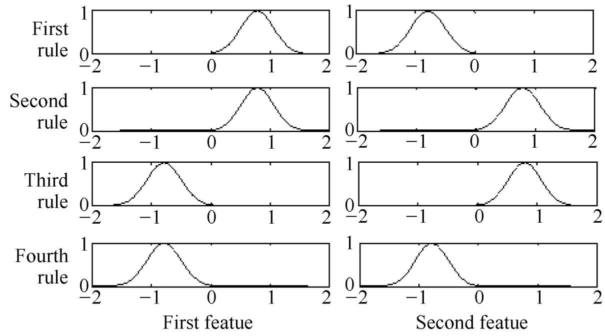


图 3　PTSK 在 mexihat 数据集上得到的模糊集示意图

Fig. 3　Fuzzy sets obtained by PTSK on mexihat dataset

为了考察 PTSK 模糊系统的收敛情况, 图 4 给出了 PTSK 在 mexihat, elevators, bodyfat 和 wizmir 数据集上某次运行的收敛曲线 (固定参数 $\beta = 4$). 从图中曲线可以看出, PTSK 在这 4 个数据集上的迭代次数均小于 200. 此时模糊规则数和求得的规则前件/后件参数趋于最优值, 系统的性能达到稳定.

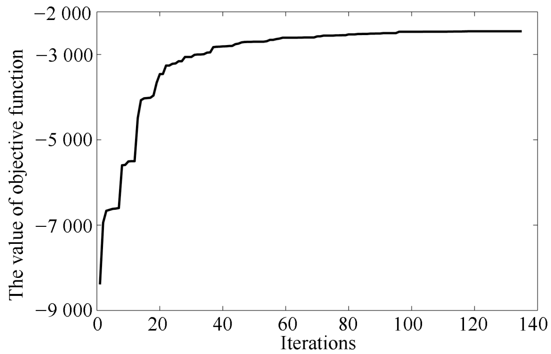
3.3.2　非参数检验

本小节使用非参数检验中的 Friedman 检验和 Holm post-hoc 检验来分析 8 种算法在 28 个数据集上 MSE 值的统计学显著性差异, 设置显著性水平 $\alpha = 0.05$. Friedman 检验是一种利用秩实现对多个总体分布是否存在显著差异的非参数检验方法. 图 5 显示了 8 种算法在 Friedman 检验上 Friedman 秩结果. 实验结果表明本文提出的 PTSK 模糊系统在 25 个数据集上取得了最佳结果.

表 7　mexihat, elevators, bodyfat 和 wizmir 数据集上 β 参数敏感性实验

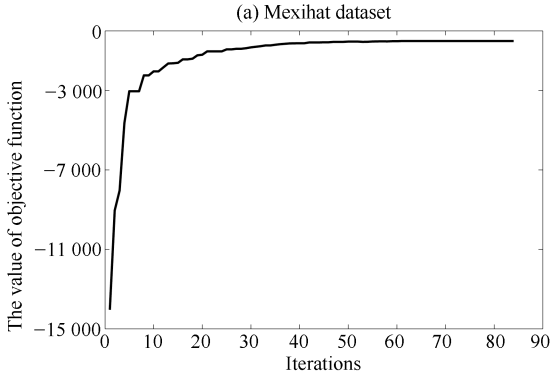
Table 7　Sensitivity experiments of parameter β on mexihat, elevators, bodyfat and wizmir datasets

Datasets		$\beta=1$	$\beta=2$	$\beta=3$	$\beta=4$	$\beta=5$	$\beta=6$	$\beta=7$	$\beta=8$
mexihat	MSE	0.0367	0.0304	0.0257	0.0222	0.0204	0.0213	0.0248	0.0248
	Rules	10.4	8.6	6.6	5.2	4.0	3.8	3.6	3.6
elevators	MSE	5.9476×10^{-4}	5.5531×10^{-4}	5.6307×10^{-4}	5.6948×10^{-4}	5.6281×10^{-4}	5.5845×10^{-4}	6.0934×10^{-4}	6.4256×10^{-4}
	Rules	24.4	23.8	22.6	21.2	20.8	20.2	18.4	16.6
bodyfat	MSE	3.4876×10^{-4}	3.5512×10^{-4}	3.4402×10^{-4}	3.6802×10^{-4}	3.9823×10^{-4}	3.8963×10^{-4}	3.9027×10^{-4}	3.9216×10^{-4}
	Rules	3.5	3.0	3.2	3.0	2.8	2.6	2.6	2.6
wizmir	MSE	1.7657	1.7606	1.6435	1.5245	1.5288	1.7578	1.9244	2.0864
	Rules	28.2	28.0	26.0	25.0	25.0	22.6	20.8	19.8



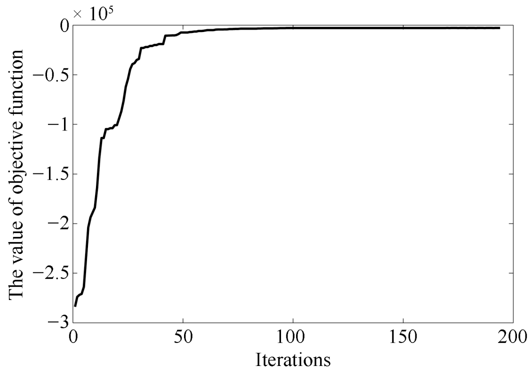
(a) Mexihat 数据集

(a) Mexihat dataset



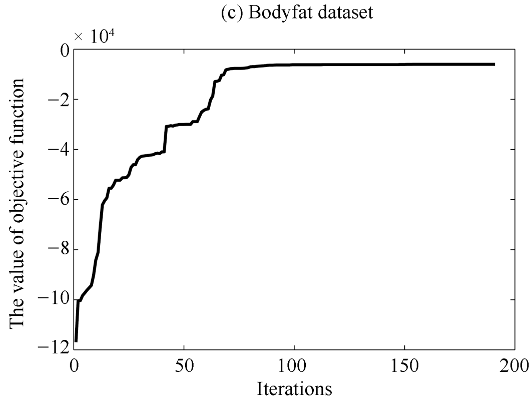
(b) Elevators 数据集

(b) Elevators dataset



(c) Bodyfat 数据集

(c) Bodyfat dataset



(d) Wizmir 数据集

(d) Wizmir dataset

图 4 PTSK 某次运行的收敛曲线

Fig. 4 Convergence curves of PTSK at a certain simulation

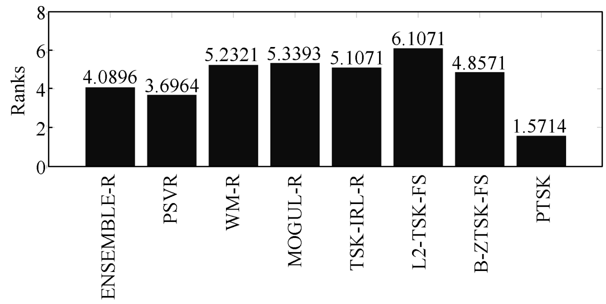


图 5 8 种算法的 Friedman 检验结果

Fig. 5 Friedman results of eight algorithms

Holm post-hoc 检验作为事后分析方法, 常用于根据 Friedman 检验结果进行最优算法与其他算法的两两比较分析. 实验中将 PTSK 与另外 7 种算法进行两两比较, 如果得到的 APV (Adjusted p-value) 值小于显著性水平, 即 $p < \alpha/i$, 则说明 PTSK 模糊系统有显著优势, 反之则说明两个算法的性能间没有显著差异. Holm post-hoc 检验的结果如表 8 所示. 从表 8 中数据可知, 与对比的 7 种算法相比, 所提 PTSK 模糊系统在系统性能上具有显著优势.

表 8 Holm post-hoc 检验结果

Table 8 Holm post-hoc results

Algorithm	z	p	Holm $= a/i$	Hypothesis
L2-TSK-FS	6.9284	0	7.143×10^{-3}	Rejected
WM-R	5.7555	0	8.333×10^{-3}	Rejected
MOGUL-R	5.5918	0	0.0100	Rejected
TSK-IRL-R	5.4009	0	0.0125	Rejected
B-ZTSK-FS	5.0190	1.0×10^{-6}	0.0167	Rejected
ENSEMBLE-R	3.8461	1.2×10^{-5}	0.0250	Rejected
PSVR	3.2460	0.00117	0.0500	Rejected

4 结论

本文使用概率模型构建了一种新的概率 TSK 模糊系统 PTSK. 在模糊和概率理论的协同工作模式下, PTSK 建立了结构辨识和参数优化的协同学习机制. 该学习机制将 TSK 模糊系统的构建视为一个整体, 能充分挖掘输入空间和输出空间之间的内在联系. PTSK 基于最大后验概率估计, 使用粒子滤波同时求得模糊规则数和前后件参数的最优解, 解决了传统聚类法 TSK 模糊系统分阶段求解参数和模糊规则数需预先设定的问题. 实验结果表明 PTSK 的逼近性能和模糊规则数均取得了令人满意的结果. 应当指出, 本文算法仍存在一些不足之处, 例如, 在大规模样本的回归问题中, PTSK 的时间效率还有待提高; 另外, PTSK 能否有效处理带噪声的回归数据亦没有进行探讨, 这将作为我们近期的研

究重点.

References

1 Zhang Bi-Shan, Ma Zhong-Jun, Yang Mei-Xiang. Robust $H\infty$ fuzzy output-feedback control with both general multiple probabilistic delays and multiple missing measurements and random missing control. *Acta Automatica Sinica*, 2017, **43**(9): 1656−1664
(张必山, 马忠军, 杨美香. 既含有一般多个随机延迟以及多个测量丢失和随机控制丢失的鲁棒 $H\infty$ 模糊输出反馈控制. *自动化学报*, 2017, **43**(9): 1656−1664)

2 Gu Xiao-Qing, Jiang Yi-Zhang, Wang Shi-Tong. Zero-order TSK-type fuzzy system for imbalanced data classification. *Acta Automatica Sinica*, 2017, **43**(10): 1773−1788
(顾晓清, 蒋亦樟, 王士同. 用于不平衡数据分类的 0 阶 TSK 型模糊系统. *自动化学报*, 2017, **43**(10): 1773−1788)

3 Luo M, Sun F C, Liu H P. Joint block structure sparse representation for multi-input-multi-output (MIMO) T-S fuzzy system identification. *IEEE Transactions on Fuzzy Systems*, 2014, **22**(6): 1387−1400

4 Garcia A M, Carmona C J, Gonzalez P, Jesus M J. MOEA-EFEP: Multi-objective evolutionary algorithm for the extraction of fuzzy emerging patterns. *IEEE Transactions on Fuzzy Systems*, 2018, **3**(9): 2861−2872

5 Deng Z H, Choi K S, Wang S T. Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation. *IEEE Transactions on Fuzzy Systems*, 2011, **19**(4): 210−226

6 Leski J M. SparseFIS: Data-driven learning of fuzzy systems with sparsity constraints. *IEEE Transactions on Fuzzy Systems*, 2010, **18**(2): 396−411

7 Jiang Y Z, Deng Z H, Wang S T. Mamdani-Larsen type transfer learning fuzzy system. *Acta Automatica Sinica*, 2012, **38**(9): 1393−1409
(蒋亦樟, 邓赵红, 王士同. ML 型迁移学习模糊系统. *自动化学报*, 2012, **38**(9): 1393−1409)

8 Pal N R, Mudi R, Pal K, Rule extraction through exploratory data analysis for self-tuning fuzzy controllers. *International Journal of Fuzzy Systems*, 2004, **6**(2): 71−80

9 Juang C F, Hsieh C D. A fuzzy system constructed by rule generation and iterative linear SVR for antecedent and consequent parameter optimization. *IEEE Transactions on Fuzzy Systems*, 2012, **20**(2): 372−384

10 Liu J F, Chung F L, Wang S T. Bayesian zero-order TSK fuzzy system modeling. *Applied Soft Computing*, 2017, **55**(6): 253−264

11 Zadeh L A. Discussion: Probability theory and fuzzy logic are complementary rather than competitive. *Technometrics*, 1995, **37**(3): 271−276

12 Hardy A. The Poisson processes in cluster analysis, classification and multivariate analysis for complex data structures: studies in classification, data analysis, and knowledge organization. Springer-Verlag Berlin Heidelberg. July 2011

13 Puri C, Kumar N. Type-2 projected Gustafson-Kessel clustering algorithm. *International Journal of Computer Applications*, 2017, **167**(14): 1−6

14 Agounad S, Aassif E H, Khandouch Y, Maze G, Décultot D. Characterization and prediction of the backscattered form function of an immersed cylindrical shell using hybrid fuzzy clustering and bio-inspired algorithms. *Ultrasonics*, 2018, **83**(2): 222−235

15 Cheng W Y, Juang C F. A fuzzy model with online incremental SVM and margin-selective gradient descent learning for classification problems. *IEEE Transactions on Fuzzy System*, 2012, **20**(2): 372−384

16 Lin C J, Chen C H. A self-constructing compensatory neural fuzzy and its applications. *Mathematical and Computer Modelling*, 2000, **42**(3): 339−351

17 Chao C, Zare A, Trinh H N, Omotara G O, Cobb J T, Lagaunne T A. Partial membership latent Dirichlet allocation for soft image segmentation. *IEEE Transactions on Image Processing*, 2017, **26**(12): 5590−5602

18 Mitianoudis N. A generalized directional Laplacian distribution: Estimation, mixture models and audio source separation. *IEEE Transactions on Audio Speech and Language Processing*, 2012, **20**(9): 2397−2408

19 Gu X Q, Chung F L, Ishibuchi H, Wang S T. Imbalanced TSK fuzzy classifier by cross-class Bayesian fuzzy clustering and imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017, **47**(8): 2005−2020

20 Gu X Q, Wang S. Bayesian Takagi-Sugeno-Kang fuzzy model and its joint learning of structure identification and parameter estimation. *IEEE Transactions on Industrial Informatics*, 2018, **47**(8): 5327−5337

21 Chopin N. A sequential particle filter method for static models. *Biometrika*, 2002, **89**(3): 539−551

22 Cheng W C. PSO algorithm particle filters for improving the performance of lane detection and tracking systems in difficult roads. *Sensors*, 2012, **12**(12): 17168−17185

23 LIBSVM Datasets [Online], available: https://www.csie.ntu.edu.tw/ cjlin/libsvmtools/datasets, November 11, 2017

24 KEEL Software and KEEL Datasets [Online], available: http://sci2s.ugr.es/keel, November 11, 2017

25 Cordón O, Herrera F. A two-stage evolutionary process for designing TSK fuzzy rule-based systems. *IEEE Trans. Systems, Man and Cybernetics, Part B: Cybernetics*, 1999, **29**(6): 703−715

26 Alcalá R, Alcala-Fdez J, Casillas J, Cord ó n O, Herrera F. Local identification of prototypes for genetic learning of accurate TSK fuzzy rule-based systems. *International Journal of Intelligent Systems*, 2007, **22**(9): 909−941

27 Wang L X, Mendel J M. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics*, 1992, **22**(6): 1414−1427

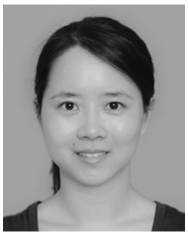
28 Pedrajas N G, Osorio C G, Fyfe C. Nonlinear boosting projections for ensemble construction. *Journal of Machine Learning Research*, 2007, **8**(1): 1−33

29 Peng X J, Xu D. Projection support vector regression algorithms for data regression. *Knowledge-Based Systems*, 2016, **112**(11): 54−66

30 Demsar J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 2006, **7**(1): 1−30

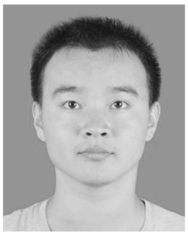
31 Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 1979, **6**(2): 65−70

32 Eichhorn H, Cano J L, McLean F, Anderl R. A comparative study of programming languages for next-generation astrodynamics systems. *CEAS Space Journal*, 2018, **10**(3): 115−123

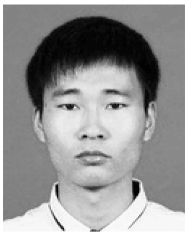


顾晓清 常州大学信息科学与工程学院副教授. 2017 年获江南大学博士学位. 主要研究方向为模式识别和机器学习. E-mail: czxqgu@163.com
(**GU Xiao-Qing** Associate professor at the School of Information Science and Engineering, Changzhou University. She received her Ph. D. degree from Jiangnan University in 2017. Her research interest covers pattern recognition and machine learning.)

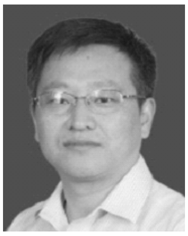
倪彤光 常州大学信息科学与工程学院副教授. 2014 年获江南大学博士学位. 主要研究方向为人工智能. 本文通信作者. E-mail: hbxtntg-12@163.com
(**NI Tong-Guang** Associated professor at the School of Information Science and Engineering, Changzhou University. He received his Ph. D. degree from Jiangnan University in 2014. His main research interest is artificial intelligence. Corresponding author of this paper.)



张 聪 常州大学信息科学与工程学院硕士研究生. 主要研究方向为模糊理论. E-mail: 18000052@smail.cczu.edu.cn
(**ZHANG Cong** Master student at the School of Information Science and Engineering, Changzhou University. His main research interest is fuzzy theory.)



戴臣超 常州大学信息科学与工程学院硕士研究生. 主要研究方向为机器学习. E-mail: 17000138@smail.cczu.edu.cn
(**DAI Chen-Chao** Master student at the School of Information Science and Engineering, Changzhou University. His main research interest is machine learning.)



王洪元 常州大学信息科学与工程学院教授. 2004 年获南京理工大学博士学位. 主要研究方向为人工智能. E-mail: hywang@cczu.edu.cn
(**WANG Hong-Yuan** Professor at the School of Information Science and Engineering, Changzhou University. He received his Ph. D. degree from Nanjing University of Science and Technology in 2004. His main research interest is artificial intelligence.)