

基于模糊子空间聚类的0阶岭回归TSK模糊系统

邓赵红, 张江滨, 蒋亦樟, 王士同

(江南大学 数字媒体学院, 江苏 无锡 214122)

摘要: 经典数据驱动型 TSK 模糊系统在利用高维数据训练模型时, 由于规则前件采用的特征过多, 导致规则的解释性和简洁性下降. 对此, 根据模糊子空间聚类算法的子空间特性, 为 TSK 模型添加特征抽取机制, 并进一步利用岭回归实现后件的学习, 提出一种基于模糊子空间聚类的 0 阶岭回归 TSK 模型构建方法. 该方法不仅能为规则抽取重要子空间特征, 而且可为不同规则抽取不同的特征. 在模拟和真实数据集上的实验结果验证了所提出方法的优势.

关键词: 解释性; 高维数据; 岭回归; TSK 模糊系统

中图分类号: TP391

文献标志码: A

Fuzzy subspace clustering based 0-order ridge regression TSK fuzzy system

DENG Zhao-hong, ZHANG Jiang-bin, JIANG Yi-zhang, WANG Shi-tong

(School of Digital Media, Jiangnan University, Wuxi 214122, China. Correspondent: DENG Zhao-hong, E-mail: dzh666828@aliyun.com)

Abstract: The classical data-driven Takagi-Sugeno-Kang(TSK) fuzzy system extracts more features for structuring the antecedent of the fuzzy rule when trained by high dimensional data, and the interpretation of system is degenerated and the linguistic interpretation is complex. A fuzzy modeling model for the fuzzy subspace clustering based zero-order ridge regression TSK fuzzy system is proposed, in which the feature extraction mechanism based on the subspace feature of fuzzy subspace clustering is added, and the ridge regression is used to realize the learning of consequent. The proposed method not only can extract important features for structuring fuzzy rules, but also can extract different features for different rules. The experimental results on the synthetic and real-world datasets show the advantage of the proposed method.

Keywords: interpretability; high-dimensional data; ridge regression; Takagi-Sugeno-Kang(TSK) fuzzy system

0 引言

模糊系统是基于规则的一类系统, 其核心部分是由 IF-THEN 规则所组成知识库. 模糊系统的特色是在模糊规则中引入了模糊集合^[1]和模糊逻辑, 能模拟人类的推理方式将语言形式的专家经验转化成精确数学模型的不确定性表达. 在已有的几种模糊系统模型中, TSK 型模糊系统由于其输出的简洁性, 已被很多学者探讨和推广^[2-5].

由于当今社会日益趋于信息化和数据化, 经典数据驱动型 TSK 模糊系统日益成为重要的 TSK 模糊系统构建方法. 经典数据驱动型 TSK 模糊系统缺乏一种特征抽取机制, 即规则抽取训练数据的全部特征.

当用高维数据训练模型时, 经典 TSK 模型存在如下缺陷: 1) 训练数据的维度越高, 隐含的噪音特征越多, 经典 TSK 模型的模糊规则中将会包含噪音特征, 导致模型的泛化性能降低; 2) 当经典 TSK 模型的规则中包含高维数据的全部特征时, 会使规则前件对应的模糊子集个数增多, 规则前件变得复杂, 语言描述性将变差, 规则的解释性和可读性也随之变差. 目前为止, 针对高维数据构建模糊系统的相关工作还比较少. 文献 [6] 针对高维数据模型训练提出了一种 T2HFS 模型. 该模型特征抽取的核心思想是通过 PCA 主成分分析, 从原始训练数据中抽取若干主成分作为新的训练样本. 这种做法在一定程度上改善了经典模糊系统面

收稿日期: 2015-02-04; **修回日期:** 2015-05-29.

基金项目: 国家自然科学基金项目(61170122); 江苏省杰出青年基金项目(BK20140001); 新世纪优秀人才支持计划项目(NCET120882).

作者简介: 邓赵红(1981—), 男, 教授, 博士, 从事人工智能、神经模糊计算等研究; 张江滨(1990—), 男, 硕士生, 从事人工智能、神经模糊计算的研究.

对高维数据泛化性下降的挑战,但是获取的新训练样本的特征失去了原始样本特征的物理意义,导致获取的规则解释性下降.为了解决高维数据分类问题,文献[7]将遗传算法和整数规划融合在一起,提出了一种针对高维数据基于融合遗传算法的模糊规则分类器算法(hGA).该方法的主要任务是有效解决高维数据分类的精度问题和规则数目精简问题,并没有从规则抽取特征的角度去考虑,所以模型获取的规则仍具有语言描述性差和解释性下降的问题.

本文以0阶TSK模糊系统为研究对象,以模糊子空间聚类为规则前件学习方法和特征抽取机制,以岭回归优化算法为规则后件参数学习方法,构造出一种基于模糊子空间聚类的0阶岭回归TSK模糊系统模型(FSC-0-RR-TSK-FS).与经典TSK模型相比,该模型增加了特征抽取机制,可为每条规则抽取所需的重要特征,去除掉对建模有干扰作用的噪音特征.在面对高维数据建模时,该方法增强了模糊规则的解释性,提高了模型的泛化性能.

1 相关工作

1.1 岭回归

岭回归法又称脊回归法、吉洪诺夫正则化法,是由美国学者Hote1^[8]于1962年提出的一种改进最小二乘估计法.随着近几十年的探讨和研究,岭回归算法越来越趋于完善,出现了各种形式的岭回归方法^[9-11].根据文献[12],基本岭回归的优化过程如下.

对于给定的回归任务和数据集 $D_{reg} = \{x_i, y_i\}$, $x_i \in R^d, y_i \in R, i = 1, 2, \dots, N$,岭回归的基本思想是获取一个线性回归模型,即

$$y = f(x) = x^T w, \quad (1)$$

使得优化目标函数

$$\min_w J = \frac{1}{2} \|Xw - y\|^2 + \frac{\lambda}{2} \|w\|^2 \quad (2)$$

的值最小.其中: $X = (x_1, x_2, \dots, x_N)^T$ 为输入的 $N \times d$ 矩阵; $y = (y_1, y_2, \dots, y_N)^T$; λ 为人工可调节的岭参数.根据优化理论对式(2)求偏导可得

$$\frac{\partial J}{\partial w} = 0, \quad (3)$$

可求解出

$$w^* = (X^T X + \lambda I_d)^{-1} X^T y, \quad (4)$$

其中 I_d 为 $d \times d$ 的单位矩阵.

1.2 TSK型模糊系统简介及高维数据之挑战

TSK模糊系统是由Takagi等^[13-14]提出的一类模糊系统.对于一个含有 d 维输入、1维输出和 K 条规则的TSK模糊系统,其模糊规则表示如下:

$$\text{If } x_1 \text{ is } A_1^k \wedge x_2 \text{ is } A_2^k \wedge \dots \wedge x_d \text{ is } A_d^k;$$

$$\text{Then } f^k(x) = p_0^k + p_1^k x_1 + \dots + p_d^k x_d. \quad (5)$$

其中: K 为模糊规则数; k 为模糊规则序号; $x = (x_1, x_2, \dots, x_d)^T$ 为输入的 d 维真值变量; \wedge 为模糊与操作符; A_i^k 为与 x_i 对应的第 k 条规则的模糊子集; $f^k(x)$ 为第 k 条规则输出的真值,等于 x 的线性组合; p_i^k 为规则后件的实数值参数.

对于一个真值输入向量 $x = (x_1, x_2, \dots, x_d)^T$,TSK模糊系统的输出 y^0 是对模糊规则后件 $f^k(x)$ 的加权平均,即

$$y^0 = \frac{\sum_{k=1}^K \omega^k f^k(x)}{\sum_{k=1}^K \omega^k}. \quad (6)$$

其中:加权系数 ω^k 表示第 k 条模糊规则作用于输入 x 的贡献程度,即

$$\omega^k = \prod_{i=1}^d \mu_{A_i^k}(x_i), \quad (7)$$

这里 $\mu_{A_i^k}(x_i)$ 表示 x 的第 i 维对应于模糊集 A_i^k 的隶属程度.

由式(5)可知,规则前件包含训练数据的全部特征维度.经典TSK模型对高维数据进行训练时,获取的模糊规则会出现解释性和简洁性下降的问题.若高维训练数据中含有很多噪音特征,则经典TSK模型会出现泛化性能下降的问题.具体表现在如下几方面.

1) 去噪能力差.若高维数据中存在噪音特征,则会使训练出来的模型缺乏稳定性,导致模型的泛化性能下降.

2) 对高维数据解释性下降.若训练样本的维数过高,则获得的规则前件包含的模糊子集 A_i^k 过多,导致规则前件过于复杂、系统解释性下降.

3) 经典TSK模型的模糊规则采用数据的全部特征,规则全部从同一角度对问题进行推理,与人类的实际推理机制不相符.

2 基于模糊子空间聚类的0阶岭回归TSK模糊系统

本节以模糊子空间聚类(FSC)算法^[15]作为模糊规则前件特征提取方法,以岭回归算法作为模糊规则后件参数学习方法,构造一种适宜于高维数据的新型TSK模糊系统模型——FSC-0-RR-TSK-FS,其模糊规则表示如下:

$$\text{If } x_1^k \text{ is } A_1^k \wedge x_2^k \text{ is } A_2^k \wedge \dots \wedge x_{m_k}^k \text{ is } A_{m_k}^k;$$

$$\text{Then } f_k(x) = p_0^k. \quad (8)$$

其中: K 为模糊规则数; k 为模糊规则序数; $\mathbf{x}^k = (x_1^k, x_2^k, \dots, x_{m_k}^k) \subseteq \mathbf{x}$ 为从 \mathbf{x} 中提取出的对第 k 条模糊规则贡献程度大的 m_k 个特征组成的向量; A_i^k 为与 x_i^k 相对应的模糊子集.

上述规则克服了经典 TSK 模型的规则之不足, 既剔除了影响系统性能的噪音特征, 又降低了规则前件模糊子集的个数, 同时不同的规则可从不同的视角进行推理, 符合人类的推理机制.

2.1 基于模糊子空间聚类的模糊规则前件学习

模糊子空间聚类是聚类算法的分支之一, 它能把高维数据转化到相关的子空间中进行聚类, 并且拥有良好的去噪能力. 根据文献 [16], 模糊子空间聚类的目标函数可以表示为如下形式:

$$J_{\text{FSC}} = \sum_{k=1}^K \sum_{i=1}^d w_{ki}^T \left(\sum_{j=1}^N u_{kj} (x_{ji} - v_{ki})^2 + \varepsilon_0 \right),$$

s.t.

$$u_{kj} \in \{0, 1\}, \sum_{k=1}^K u_{kj} = 1, 0 < \sum_{j=1}^N u_{kj} < N,$$

$$0 \leq w_{ki} \leq 1, \sum_{i=1}^d w_{ki} = 1. \quad (9)$$

其中: \mathbf{V} 和 \mathbf{U} 分别表示聚类中心和硬划分矩阵; K 、 N 和 d 分别表示聚类个数、样本个数和样本特征维数; ε_0 为一个小的正则化常数. 为了体现子空间聚类特性, 为每个聚类设定权值向量 \mathbf{w}_k , $k = 1, 2, \dots, K$, 其中 $w_{ki} \in \mathbf{w}_k$ 表示第 i 个特征为第 k 个类簇的贡献程度. FSC 算法的子空间特性与 FSC-0-RR-TSK-FS 模糊规则前件提取具有一致性, 可用 FSC 算法为每条模糊规则选取前件特征, 具体选取策略如下.

给定训练数据集 $D_{\text{tr}} = \{\mathbf{x}_i, y_i\}$, $\mathbf{x}_i \in R^d$, $y_i \in R$, $i = 1, 2, \dots, N$, 通过 FSC 算法将 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ 划分为 K 类, 与 K 条模糊规则相对应; 权值向量 \mathbf{w}_k 对应于全部特征对第 k 条模糊规则的贡献程度. 给定阈值 $\eta = \gamma/d$, 第 k 条规则遍历权值向量 \mathbf{w}_k , 从中选取贡献程度不小于 η 的特征, 即 $\mathbf{x}^k = (x_1^k, x_2^k, \dots, x_{m_k}^k) \subseteq \mathbf{x}$. 其中 $0 < \gamma \leq d$ 为人工设置的参数, 可根据 γ 动态调节每条规则选取特征的个数.

每条模糊规则选取的每个特征对应于一个模糊子集 A_i^k . 若采用高斯函数作为隶属度函数, 则 A_i^k 所对应的高斯函数参数可用如下方式估计:

$$c_i^k = \frac{\sum_{j=1}^N u_{kj} x_{ji}^k}{\sum_{j=1}^N u_{kj}}, \quad (10)$$

$$\delta_i^k = \frac{h \sum_{j=1}^N u_{kj} (x_{ji}^k - c_i^k)^2}{\sum_{j=1}^N u_{kj}}. \quad (11)$$

其中: x_{ji}^k 表示第 j 个样本 $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jd})^T$ 在第 k 条模糊规则中选取的第 i 个特征, u_{kj} 表示第 j 个样本是否属于第 k 个聚类.

2.2 基于岭回归优化算法的模糊规则后件学习

根据 2.1 节中的特征选取策略和 TSK 模型的输出式 (6), 可得 FSC-0-RR-TSK-FS 的模型输出如下:

$$f(\mathbf{x}) = \sum_{k=1}^K \frac{\mu^k(\mathbf{x}^k)}{\sum_{k'=1}^K \mu^{k'}(\mathbf{x}^{k'})} f^k(\mathbf{x}^k) = \sum_{k=1}^K \tilde{\mu}^k(\mathbf{x}^k) p_0^k. \quad (12)$$

其中

$$\mu^k(\mathbf{x}^k) = \prod_{i=1}^{m_k} \mu_{A_i^k}(x_i^k), \quad (13)$$

$$\tilde{\mu}^k(\mathbf{x}^k) = \frac{\mu^k(\mathbf{x}^k)}{\sum_{k'=1}^K \mu^{k'}(\mathbf{x}^{k'})}. \quad (14)$$

为了便于优化模糊规则后件参数, 根据文献 [2] 的相关策略, 令

$$\mathbf{x}_g = (\tilde{\mu}^1(\mathbf{x}^1), \tilde{\mu}^2(\mathbf{x}^2), \dots, \tilde{\mu}^K(\mathbf{x}^K))^T, \quad (15)$$

$$\mathbf{p}_g = (p_0^1, p_0^2, \dots, p_0^K)^T, \quad (16)$$

可将式 (12) 转化为

$$f(\mathbf{x}) = \mathbf{p}_g^T \mathbf{x}_g. \quad (17)$$

根据 1.1 节岭回归优化算法, 利用训练数据集 $D_{\text{tr}} = \{\mathbf{x}_i, y_i\}$, $\mathbf{x}_i \in R^d$, $y_i \in R$, $i = 1, 2, \dots, N$, 构造 FSC-0-RR-TSK-FS 模型的优化目标函数如下:

$$\min_{\mathbf{p}_g} J = \frac{1}{2} \|\tilde{\mathbf{X}} \mathbf{p}_g - \mathbf{y}\|^2 + \frac{\lambda}{2} \|\mathbf{p}_g\|^2. \quad (18)$$

其中: $\tilde{\mathbf{X}} = (\mathbf{x}_{g1}, \mathbf{x}_{g2}, \dots, \mathbf{x}_{gN})^T$, \mathbf{x}_{gj} 表示第 j 个样本通过式 (15) 转化而来的一个向量, λ 为岭参数. 通过求解 (18), 可得到 \mathbf{p}_g 的最优值为

$$\mathbf{p}_g^* = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}} + \lambda \mathbf{I}_d)^{-1} \tilde{\mathbf{X}}^T \mathbf{y}. \quad (19)$$

至此, 整合 2.1 节规则前件学习和 2.2 节规则后件学习, 可得到 FSC-0-RR-TSK-FS 模型.

2.3 FSC-0-RR-TSK-FS 模型性能分析

1) 基于 FSC 特征自适应学习的规则前件获取.

传统的聚类技术, 如模糊 C -均值算法 (FCM), 没有考虑到数据中的子空间特性, 在对高维数据进行聚类过程中, 容易受到噪声特征的影响, 导致聚类划分

出现偏差. 假设给定的数据集具有 d 维特征 $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$, 其特征子集 $\mathbf{x}' = (x_1, x_2, \dots, x_j)^T \subseteq \mathbf{x}$ 包含有效的特征, 在特征子集 $\mathbf{x}'' = (x_{j+1}, x_{j+2}, \dots, x_d)^T \subseteq \mathbf{x}$ 中包含噪声特征, 则 FCM 算法在聚类过程中, 样本集在整个空间可能会导致样本点之间的差异度变小, 易出现错误划分的情况, 导致相应的模糊规则的空间划分不合理, 影响训练所得系统的泛化性能. 而 FSC 算法为每一类中的每个特征添加了权重 w_{ij} , 在权重的影响下, 有效消除了噪声特征的影响, 使聚类的结果更加容易接近真实聚类, 使得相应的模糊规则中对应的空间划分更合理. 因此, TSK 模型采用 FSC 构造规则前件, 更能准确地表述真实的建模任务.

2) 岭回归规则后件优化.

学习模型的风险是由经验风险和置信风险组成的. 最小二乘法的主要思想是求解最优向量 \mathbf{w} , 使样本数据上的模型输出结果与真实结果之间的差值最小, 即经验风险最小化. 经验风险最小化的缺点是, 训练出来的模型在样本集上可达到很高的正确率, 在样本之外的数据正确率很低, 导致模型过拟合, 泛化能力差. 为了求解得到最优化参数, 需要同时最小化经验风险和置信风险. 根据统计学习理论^[17-18], 为最小二乘法添加结构风险正则化项, 构造出岭回归方法, 使模型得到较好的泛化性能.

另外, 选择岭回归进行模糊规则后件参数优化, 相对于经典的基于结构风险最小化的 LS-SVM 等方法, 可获得更宽松的优化条件. 如文献 [20] 所分析, 本文提出的方法通过化解式 (18) 可得

$$\begin{aligned} \min_{\mathbf{p}_g} J &= \frac{\lambda}{2} \|\mathbf{p}_g\|^2 + \frac{1}{2} \sum_{i=1}^N \xi_i^2; \\ \text{s.t. } \mathbf{p}_g^T \mathbf{x}_{gi} &= y_i - \xi_i, \quad i = 1, 2, \dots, N. \end{aligned} \quad (20)$$

对比 LS-SVM^[19]的优化函数为

$$\begin{aligned} \min_{\mathbf{w}} J &= \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{2} \sum_{i=1}^N \xi_i^2; \\ \text{s.t. } y_i(\mathbf{w}\phi(\mathbf{x}_i) + b) &= 1 - \xi_i, \\ i &= 1, 2, \dots, N. \end{aligned} \quad (21)$$

根据 KKT 理论, 式 (20) 和 (21) 的优化问题分别对应于如下对偶优化问题:

$$\begin{aligned} \max L_{RR-TSK} &= -\frac{1}{2\lambda} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \mathbf{x}_{gi}^T \mathbf{x}_{gj} - \\ &\frac{1}{2} \sum_{i=1}^N (\alpha_i)^2 + \sum_{i=1}^N \alpha_i y_i; \\ \text{s.t. } \alpha_i &\geq 0, \quad \forall i; \end{aligned} \quad (22)$$

$$\begin{aligned} \max L_{LS-SVM} &= \\ &-\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \phi^T(\mathbf{x}_i) \phi(\mathbf{x}_j) - \\ &\frac{1}{2C} \sum_{i=1}^N (\alpha_i)^2 - \sum_{i=1}^N (y_i b - 1) \alpha_i; \\ \text{s.t. } \sum_{i=1}^N \alpha_i y_i &= 0, \quad \alpha_i \geq 0, \quad \forall i. \end{aligned} \quad (23)$$

对比式 (22) 和 (23) 可发现, 将岭回归运用到模糊系统中, 会比运用到 LS-SVM 中少一个约束条件, 即

$$\sum_{i=1}^N \alpha_i y_i = 0. \quad (24)$$

因此, 将岭回归运用到模糊系统建模时, 会具有更加宽松的优化条件, 能够在整个的特征空间中寻找到更优解.

3 实验结果与分析

3.1 实验设置

1) 对比算法.

为了论证 FSC-0-RR-TSK-FS 模型, 在展示其在高维数据方面解释性增强的特色时, 本文选取了 L2 型 TSK 模型 (L2-TSK-FS)、RBF 神经网络 (RBF-NN)、LibSVM 工具箱中的 SVM 算法和 SVR 算法作为对比实验. 同时, 为了验证新模型特征选取的优势, 为各对比算法添加 PCA 技术, 作为具有特征选取功能的对比算法.

2) 评价标准.

在模拟数据集实验部分进行的是分类实验, 分类实验的评价标准为正确分类率 J_{cla} . J_{cla} 越高, 说明分类效果越好. 在真实数据集实验部分进行的是回归实验, 可采用下式作为评价标准^[2]:

$$J_{reg} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y'_i - y_i)^2 / \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}. \quad (25)$$

其中: N 为测试样本数, y_i 为第 i 个测试样本的期望输出, y'_i 为第 i 个测试样本的模糊系统输出, $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$. J_{reg} 的数值越小, 说明测试效果越好.

3) 实验环境.

实验硬件平台: Intel(R) Core(TM) i3-3240 CPU; 主频: 3.40 GHz; 内存: 4 GB; 系统类型: Win7 32 位操作系统; 编程环境: Matlab 7.10.0.499 (R2010a).

3.2 模拟数据集实验

为了体现 FSC-0-RR-TSK-FS 模型的优势, 本文构造了 300 组 30 维的数据用于分类任务. 该组数据集共有 3 类. 其中: 标签 1 类在 $\{x_3, x_7, x_{12}, x_{18}, x_{25}\}$ 具有子空间特征; 标签 2 类在 $\{x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}$

表 1 模拟数据集实验性能对比

算法	规则数/隐节点数	采用的特征数目	$J_{cla}/\%$
L2-TSK-FS	3条规则	3条规则均采用30个特征	74.00
RBF-NN	6个隐节点	6个隐节点均采用30个特征	53.67
PCA+L2-TSK-FS	9条规则	9条规则均抽取PCA变换后得到的19个新特征	83.67
PCA + RBF-NN	5个隐节点	5个隐节点均抽取PCA变换后得到的22个新特征	68.00
FSC-0-RR-TSK-FS	3条规则	3条规则从原始特征中分别抽取的特征数目为4、5、5	99.67
SVM		采用训练数据中的30个特征	64.33
PCA+ SVM		采用PCA变换后得到的22个新特征	68.67

表 2 FSC-0-RR-TSK-FS 方法为模拟数据构造的模糊规则库

规则号	规则所抽取的特征	规则前件	规则后件
R_1	15, 17, 23, 24	$c_1 = [0.0309, 0.0300, 0.0394, 0.0330]$; $\delta_1 = [-0.7328, -0.7272, 0.3465, 0.6507]$	$p_0^1 = 2.9268$
R_2	7, 9, 10, 11, 12	$c_2 = [0.0320, 0.0329, 0.0345, 0.0347, 0.0345]$; $\delta_2 = [-0.4305, 0.1626, -0.2213, 0.5784, 0.4031]$	$p_0^2 = 1.9511$
R_3	3, 7, 12, 18, 25	$c_3 = [0.1274, 0.3390, 0.6619, 0.1795, -0.3558]$; $\delta_3 = [0.0401, 0.0404, 0.0385, 0.0403, 0.0362]$	$p_0^3 = 0.9756$

具有子空间特征; 标签3类在 $\{x_{15}, x_{17}, x_{19}, x_{23}, x_{24}\}$ 具有子空间特征. 其他的维度为噪音特征.

表1描述的是各对比算法的性能对比. 由表1可得如下结论:

1) 由于模拟数据中存在大量的噪音特征, 对模型训练具有强烈的干扰作用, 导致L2-TSK-FS模型、RBF-NN模型和LS-SVM模型的分类效果很差.

2) 在各经典方法与PCA技术相结合之后, 利用PCA对数据进行特征变换降低了特征维数, 经典模型的分类精度具有一定的提升.

3) RBF-NN模型和SVM模型的性能虽与L2-TSK-FS模型相当, 但在进行模型训练时, 仍采用模拟数据的30个特征, 这样增加了模型的训练复杂度. 另外, 由于这些模型的本身特性, 它们无法拥有TSK-FS模型的解释性.

4) L2-TSK-FS模型在对样本进行训练时, 采用样本的30个特征构造模糊规则, 规则前件有30个模糊子集, 导致规则的前件变得复杂, 语言描述性和规则的解释性变差. 与PCA技术相结合后, L2-TSK-FS模型为每条规则抽取出19个特征. 该方法虽然在一定程度上增强了规则的语言性描述, 使得规则变得简洁, 但是采用PCA技术获取的新样本特征失去了原有样本特征的物理意义, 导致模型的解释性下降.

5) FSC-0-RR-TSK-FS模型采用FSC算法为规则抽取出重要特征, 去掉了训练数据中的噪音特征, 使得模型分类精度达到99.67. 另外, 该方法只采用了较少的重要特征来构建模糊规则, 使得规则前件变

得简洁, 具有良好的语言描述性, 从而增强了规则的解释性.

图1为FSC算法对模拟数据集聚类得到的每类数据特征的分布情况. FSC-0-RR-TSK-FS模糊规则 R^k 描述如下.

规则前件: If x_1^k is $(c_1^k, \delta_1^k) \wedge x_2^k$ is $A_2^k(c_2^k, \delta_2^k) \wedge \dots \wedge x_{m_k}^k$ is $A_{m_k}^k(c_{m_k}^k, \delta_{m_k}^k)$;

规则后件: Then $y^k = f^k(x^k) = p_0^k$.

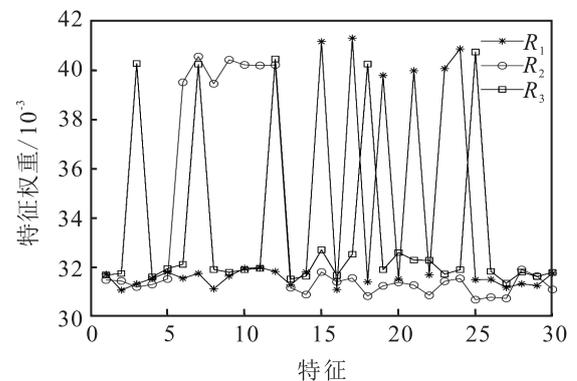


图 1 FSC 算法对模拟数据集聚类得到的每条规则的特征权重

表2描述的是FSC-0-RR-TSK-FS为模拟数据集所构造的模糊规则库. 由图1和表2可见, 模型为构造的3条规则抽取出不同的特征, 使3条规则可从不同的视角进行推理, 更加符合人类的推理机制. 3条规则所抽取的特征与图1中权重较高的特征相对应.

3.3 发酵数据集仿真实验

为了进一步验证FSC-0-RR-TSK-FS模型解释性

表 3 发酵数据集实验性能对比 (菌体浓度 $X(k+1)$)

算法	规则数/隐节点数	采用的特征数目	J_{reg}
L2-TSK-FS	11 条规则	11 条规则均均采用 6 个特征	0.4029
RBF-NN	9 个隐节点	9 个隐节点均采用 6 个特征	0.3994
PCA+L2-TSK-FS	5 条规则	5 条规则均抽取 PCA 变换后得到的 3 个新特征	0.4746
PCA + RBF-NN	7 个隐节点	7 个隐节点均抽取 PCA 变换后得到的 2 个新特征	0.4684
FSC-0-RR-TSK-FS	9 条规则	9 条规则从原始特征中分别抽取的特征数目为 1、4、3、3、3、2、2、1、3	0.3925
SVR		采用训练数据中的 6 个特征	0.6029
PCA+ SVR		采用 PCA 变换后得到的 2 个新特征	0.6227

增强的特色, 下面采用谷氨酸发酵过程建模进行回归仿真实验. 该发酵数据集是采集、筛选自工厂发酵数据, 共有 294 组数据. 该发酵数据集有 6 维输入, 分别为发酵时间 k , 葡萄糖浓度 $S(k)$, 菌体浓度 $X(k)$, 谷氨酸浓度 $P(k)$, 搅拌速度 $R(k)$ 和空气流通速度 $Q(k)$; 输出为下一时刻的菌体浓度 $X(k+1)$. 为了对输出进行预测, 需构建一个 TSK 模糊系统模型. 各对比算法对菌体浓度 $X(k+1)$ 的预测实验结果如表 3 所示.

由表 3 可得出如下结论:

1) 因为训练样本只含有 6 维特征, 其所包含的噪音特征很少, 所以各算法的回归性能相差不大.

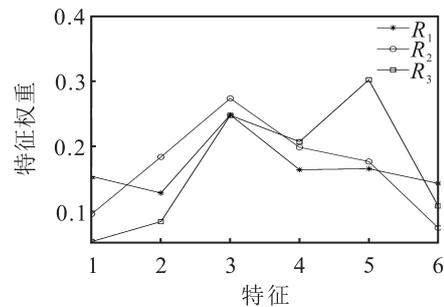
2) 添加 PCA 技术之后, 发酵数据集浓缩为新的训练样本. 因为原样本只含有 6 维特征, 故浓缩后得到的新样本损失了部分信息, 导致添加了 PCA 技术的对比算法性能稍差一些.

3) RBF-NN 模型和 SVR 模型虽然得到了与其他算法可比较的回归性能, 但由于其模型的特性, 无法具有较好的模型解释性.

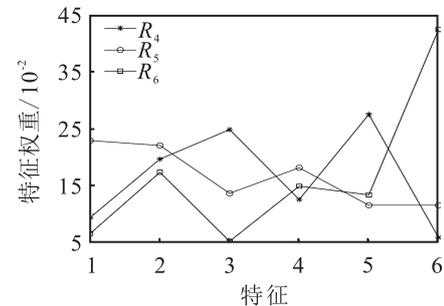
4) 相对于 RBF-NN 模型和 SVR 模型, L2-TSK-FS 模型因其对训练数据构造了模糊规则, 具有较好的解释性. 但是, L2-TSK-FS 模型的模糊规则采用了数据的 6 个特征, 每条规则对应于 6 个模糊子集的语言描述, 使得规则的语言描述性稍差, 规则较复杂. 若利用 PCA 技术降低每条规则的模糊子集, 获取的新特征则会失去原始特征的物理意义.

5) 本文方法在不损失回归性能的基础上, 为每条规则只选取较少的特征, 使每条规则对应的模糊子集个数很少, 从而易于进行语言描述, 增强了规则的简洁性和解释性.

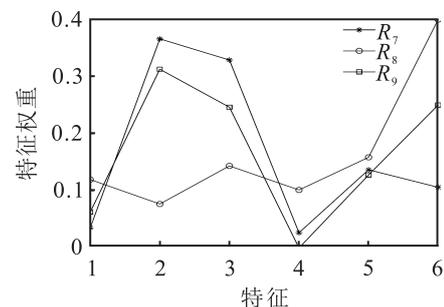
图 2 为 FSC-0-RR-TSK-FS 模型 9 条规则对应的特征权值分布情况. 由图 2 可见, 9 条规则均具有各自不同的重要特征, 对应于不同的特征子空间. FSC-0-RR-TSK-FS 模型构造出来的 9 条规则可分别从不同的视角进行推理, 更加符合人类的推理机制.



(a) 规则 R_1 - R_3 的特征权重



(b) 规则 R_4 - R_6 的特征权重



(c) 规则 R_7 - R_9 的特征权重

图 2 FSC 算法对发酵数据集聚类得到的每条规则的特征权重

4 结 论

传统 TSK 型模糊系统在对高维数据进行训练时, 得到的模糊规则所对应的模糊子集个数增多, 导致规则的语言描述性差, 解释性下降. 本文将 FSC 算法用于规则前件的划分和特征提取, 构造了一种 FSC-0-RR-TSK-FS 方法. 与传统 TSK 模型相比, 该方法可选取重要的特征来构建模糊规则, 去掉了样本中的噪音特征. 在保证模型泛化性能的同时, 可使规则对应的模糊子集个数减少, 易于语言描述, 增强了规则的简洁性和解释性, 适宜于高维数据建模. 同时, 该方法

获取的不同规则对应于不同的特征子空间,使得不同的规则可从不同的视角进行推理,更加符合人类的推理机制。

目前, FSC-0-RR-TSK-FS 仍具有一定的缺陷. 若给定的训练数据集的特征权值分布均匀, 则会使系统因丢失特征而导致系统性能下降. 如何从子空间特征不明显的高维数据中获取具有更强解释性的模糊系统是今后将进行的工作.

参考文献(References)

- [1] Zadeh L A. Fuzzy sets[J]. *Information and Control*, 1965, 8(3): 338-353.
- [2] Deng Z H, Choi K S, Chung F L, et al. Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation[J]. *IEEE Trans on Fuzzy Systems*, 2011, 19(2): 210-226.
- [3] 蒋亦樟, 邓赵红, 王士同. 0阶L2型TSK迁移学习模糊系统[J]. *电子学报*, 2013, 41(5): 897-904.
(Jiang Y Z, Deng Z H, Wang S T. 0-order L2-norm Takagi-Sugeno-Kang type transfer learning fuzzy system[J]. *Acta Electronica Sinica*, 2013, 41(5): 897-904.)
- [4] Lin Y Y, Chang J Y, Lin C T. A TSK-type-based self-evolving compensatory interval type-2 fuzzy neural network(TSCIT2FNN) and its applications[J]. *IEEE Trans on Industrial Electronics*, 2014, 61(1): 447-459.
- [5] Deng Z H, Choi K S, Cao L B, et al. Type-2 fuzzy extreme learning algorithm for fast training of interval type-2 TSK fuzzy logic system[J]. *IEEE Trans on Neural Networks and Learning Systems*, 2014, 25(4): 664-676.
- [6] Liu Z, Chen C L P, Zhang Y, et al. Type-2 hierarchical fuzzy system for high-dimensional data-based modeling with uncertainties[J]. *Soft Computing*, 2012, 16(11): 1945-1957.
- [7] Aydogan E K, Karaoglan I, Pardalos P M. Hybrid genetic algorithm in fuzzy rule-based classification systems for high-dimensional problems[J]. *Applied Soft Computing*, 2012, 12(2): 800-806.
- [8] Hoerl A E. Application of ridge analysis to regression problems[J]. *Chemical Engineering Progress*, 1962, 58(3): 54-59.
- [9] Hoerl A E, Kennard R W. Ridge regression: Biased estimation for nonorthogonal problems[J]. *Technometrics*, 1970, 12(1): 55-67.
- [10] Saunders C, Gammerman A, Vovk V. Ridge regression learning algorithm in dual variables[C]. *Proc of the 15th Int Conf on Machine Learning*. Madison, 1998: 515-521.
- [11] Jang J S R, Sun C T, Mizutani E. Neuro-fuzzy and soft computing — A computational approach to learning and machine intelligence[J]. *IEEE Trans on Automatic Control*, 1997, 42(10): 1482-1484.
- [12] Deng Z H, Choi K S, Jiang Y Z, et al. Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods[J]. *IEEE Trans on Cybernetics*, 2014, 44(12): 2585-2599.
- [13] 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.
- [14] Sugeno M, Kang G T. Structure identification of fuzzy model[J]. *Fuzzy Sets and Systems*, 1988, 28(1): 15-33.
- [15] Gan G, Wu J, Yang Z. A fuzzy subspace algorithm for clustering high dimensional data[C]. *Advanced Data Mining and Applications*. Berlin: Springer Heidelberg, 2006: 271-278.
- [16] Deng Z H, Cao L B, Jiang Y Z, et al. Minimax probability TSK fuzzy system classifier: A more transparent and highly interpretable classification model[J]. *IEEE Trans on Fuzzy Systems*, 2014, 23(4): 813-826.
- [17] 贾东立, 张家树. 基于混沌变异的小生境粒子群算法[J]. *控制与决策*, 2007, 22(1): 117-120.
(Jia D L, Zhang J S. Niche particle swarm optimization combined with chaotic mutation[J]. *Control and Decision*, 2007, 22(1): 117-120.)
- [18] Coello C A C, Lechuga M. A proposal for multiple objective particle swarm optimization[C]. *Proc of IEEE Congress on Evolutionary Computation*. Hawaii, 2002: 1051-1056.
- [19] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. *Neural Processing Letters*, 1999, 9(3): 293-300.
- [20] Huang G B, Zhou H M, Ding X J, et al. Extreme learning machine for regression and multiclass classification[J]. *IEEE Trans on Systems, Man, and Cybernetics, Part B*, 2012, 42(2): 513-529.

(责任编辑: 滕蓉)