

文章编号: 1001-0920(2013)02-0169-08

一种基于模糊规则融合的模糊建模方法及其应用

徐 喆, 毛志忠

(东北大学 a. 信息科学与工程学院, b. 流程工业综合自动化国家重点实验室, 沈阳 110819)

摘 要: 为了有效地利用经验知识, 弥补训练数据覆盖范围不足的问题, 提出一种将经验知识以 TSK (Takagi-Sugeno-Kang) 型模糊规则引入模糊模型的建模方法. 在结构辨识中, 提出了模糊规则融合方法, 用以确定初始模糊规则. 在参数辨识中, 改进了原梯度下降方法中的目标函数, 并引入了经验知识准确性评价参数, 用以平衡样本数据和经验知识对模型的影响. 数值仿真和工程实例应用结果表明, 所提出的方法可以有效地利用经验知识和样本数据, 使预报结果更可靠、更精确.

关键词: 经验知识; 系统辨识; TSK 模糊模型; 模糊规则融合

中图分类号: TP182

文献标志码: A

A fuzzy-rule-fusion based fuzzy modeling method and its application

XU Zhe, MAO Zhi-zhong

(a. College of Information Science and Engineering, b. State Key Laboratory of Integrated Automation for Process Industries, Northeastern University, Shenyang 110819, China. Correspondent: XU Zhe, E-mail: xuzhe83@gmail.com)

Abstract: To effectively use the empirical knowledge to compensate for incomplete training data coverage, a fuzzy modeling method that incorporates empirical knowledge in the form of TSK (Takagi-Sugeno-Kang) fuzzy rules is proposed. In the structure identification process, a fuzzy rule fusion method is proposed to determine the initial fuzzy rules. In the parameter identification process, the original objective function of the gradient descent method is improved and the evaluating parameter of the accuracy of empirical knowledge is introduced to trade off the influence of sample data and empirical knowledge. The numerical simulation and engineering case studies show that the proposed method can offer more reliable and accurate forecasting values.

Key words: empirical knowledge; system identification; TSK fuzzy model; fuzzy rules fusion

0 引 言

随着工业生产对降低生产成本、提高产品质量稳定性要求的不断提升, 人们对关键参数的预报可靠性的要求也逐步提高. 所谓参数的预报可靠性是指参数的预报值与真实值的误差量在允许偏差范围内的概率大小. 传统的以人工经验为主的操作模式已不能满足要求, 人们正寻求使用智能方法建立系统模型来替代人为判断. 自 1992 年 Kosko^[1] 和 Wang 等^[2] 分别证明了模糊系统具有一致逼近性以来, 模糊建模方法便作为一种智能建模方法成功地应用于工业非线性建模领域, 并取得了良好的效果^[3-5]. 模糊建模主要有两种途径: 1) 基于经验知识的方法, 称为“经验方

法”, 又称作模糊专家系统^[6]; 2) 基于数据的方法, 称为“数据驱动方法”. 由于经验方法受经验知识获取以及经验知识不确定性的限制, 其应用范围较为有限; 而数据驱动的建模方法则需要有充足的训练样本作为保证^[7]. 对于钢铁冶金、微生物发酵和石油化工等领域, 一方面, 由于生产周期长以及一些数据的检测成本高而不易获得样本数据; 另一方面, 其生产过程对关键参数预报的可靠性要求较高, 若预报偏差超出允许范围则会造成巨大的经济损失. 受这两方面因素的影响, 仅使用有限的样本数据很难建立能够满足生产要求的预报模型. 为了解决该问题, 可将经验知识与样本数据相结合, 共同建立系统模型. 目前常用

收稿日期: 2011-11-23; 修回日期: 2012-05-15.

基金项目: 国家高技术研究发展计划项目(2007AA041401, 2007AA04Z194); 中央高校基本科研业务费专项资金项目(N100604002).

作者简介: 徐喆(1983-), 男, 博士生, 从事复杂系统建模与优化的研究; 毛志忠(1961-), 男, 教授, 博士生导师, 从事复杂系统建模、优化与控制等研究.

的结合方法是将经验知识以约束条件的形式引入数据驱动模型^[8-10]来降低模型自由度. 该方法是在模型中加入硬约束的正则化方法^[11], 其不足之处在于使用了不确定性较强的经验知识来严格限制模型参数的取值范围^[12].

为了避免上述问题, 本文提出一种基于模糊规则融合的模糊推理系统建模方法 (FRFFIS), 将经验知识以模糊规则的形式引入模型, 并使用软约束^[13]方法在参数优化的目标函数中引入约束项, 使在参数优化过程中不致以训练误差最小化为目标, 还要使模型的预报值尽量接近经验规则的输出值, 达到利用样本数据和经验知识共同指导建模的目的.

1 问题描述

对于多输入单输出的 TSK 型模糊系统^[14], 其模糊规则的一般形式为

$$R_l : \text{If } x_1 \text{ is } F_1^l \text{ and } x_2 \text{ is } F_2^l \text{ and } \dots \text{ and } x_n \text{ is } F_n^l; \\ \text{Then } z = z^l(\mathbf{x}). \quad (1)$$

其中: R_l ($l = 1, 2, \dots, n_r$) 为第 l 条模糊规则, n_r 为规则总数; x_i ($i = 1, 2, \dots, n$) 为输入, n 为输入维数; F_i^l 为用模糊隶属度函数 $\mu_{F_i^l}$ 表示的语言项; z 为模糊规则的输出. 系统的输出函数可表示为^[15-16]

$$f(\mathbf{x}) = \frac{\sum_{l=1}^{n_r} \left(\prod_{i=1}^n \mu_{F_i^l}(x_i, \sigma_i^l, \gamma_i^l) \right) z^l(\mathbf{x})}{\sum_{l=1}^{n_r} \left(\prod_{i=1}^n \mu_{F_i^l}(x_i, \sigma_i^l, \gamma_i^l) \right)}. \quad (2)$$

其中: $f: \mathbf{R}^n \rightarrow \mathbf{R}$, 隶属函数 $\mu_{F_i^l}(x_i, \sigma_i^l, \gamma_i^l)$ 可选用径向基函数 (RBF). 第 l 条规则中第 i 维输入的隶属度计算方法为

$$\mu_{F_i^l} = \exp \left(- \left(\frac{x_i - \gamma_i^l}{\sigma_i^l} \right)^2 \right), \quad (3)$$

其中 γ_i^l 和 σ_i^l 分别为隶属函数的中心点和宽度参数. TSK 型模糊规则的输出函数为输入变量 \mathbf{x} 的一阶函数形式, 即

$$z^l(\mathbf{x}) = f^l(x_1, x_2, \dots, x_n) = w_0^l + \sum_{i=1}^n w_i^l x_i^l, \quad (4)$$

其中 w_i^l ($i = 0, 1, \dots, n$) 为结论参数.

2 TSK 型模糊系统建模方法的改进

改进后的建模方法流程如图 1 所示. 在模糊化过程中, 分别从历史数据和经验知识中提取出 TSK 形式的数据规则和经验规则; 然后使用规则融合方法对经验规则进行简约, 生成模糊模型的初始规则; 最后使用改进的混合参数优化方法对初始规则中的参数进行优化, 得到最终的模糊系统模型. 本文对模糊建模的结构辨识和参数辨识方法进行改进, 下面分别介绍这两种改进方法.

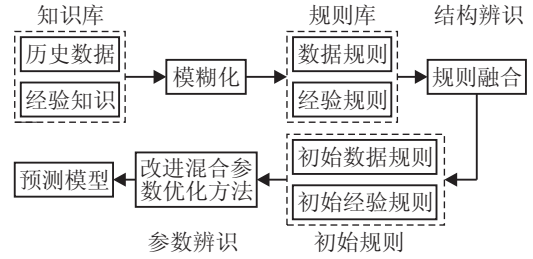


图 1 改进的 TSK 型模糊系统建模方法流程

2.1 结构辨识

2.1.1 初始经验规则的确

经验知识可以用 TSK 型模糊规则表示为

$$\text{RE}_p : \text{If } x_1 \text{ is } \text{FE}_1^p(l_1^p, u_1^p) \text{ and } x_2 \text{ is } \text{FE}_2^p(l_2^p, u_2^p) \text{ and} \\ \dots \text{ and } x_n \text{ is } \text{FE}_n^p(l_n^p, u_n^p); \text{ Then } z = z^p. \quad (5)$$

其中: RE_p 为第 p 条经验规则; $\text{FE}_i^p(l_i^p, u_i^p)$ 为与隶属度 $\mu_{\text{FE}_i^p}$ 对应的语言项; l_i^p, u_i^p 由经验知识确定, 分别为隶属函数作用范围的下限和上限. 隶属函数选取 RBF 函数, 其中心点的初始值按下式计算:

$$\gamma_i^p = \frac{l_i^p + u_i^p}{2}. \quad (6)$$

由于 RBF 函数的实际输入范围为 $(-\infty, +\infty)$, 为了通过经验确定其主要作用范围的上下限, 分别将输入上下限设定为当其输出值为 κ 时的较大输入值和较小输入值, 即当 x_i 为 l_i^p 和 u_i^p 时的隶属度为 κ . 根据式 (3) 和 (6), 可以计算得到相应的 σ_i^p 如下:

$$\sigma_i^p = \frac{(u_i^p - l_i^p)}{2\sqrt{-\ln \kappa}}. \quad (7)$$

经验规则的结论部分设定为一个常数, 即 $z^p = w^p$, 其数值由经验知识确定.

2.1.2 初始数据规则的确

本文使用减聚类方法^[17]从训练样本中提取出有代表性的样本点, 并将其转化为数据规则.

在减聚类中, 第 l 个样本的差异度为

$$D_l = \sum_{j=1}^{n_d} \left[\exp \left(\sum_{i=1}^n \frac{-(x_i^l - x_i^j)^2}{\sigma_i^2} \right) \exp \left(\frac{-(y^l - y^j)^2}{\sigma_y^2} \right) \right]. \quad (8)$$

其中: n_d 为训练样本数; σ_i 和 σ_y 为宽度参数, 其数值如下:

$$\sigma_i = [\max(\mathbf{x}_i) - \min(\mathbf{x}_i)] / K_i, \quad (9)$$

$$\sigma_y = [\max(\mathbf{y}) - \min(\mathbf{y})] / K_y, \quad (10)$$

K_i 和 K_y 用以控制模型复杂度.

首先选择差异度最大的数据 u_1 作为数据规则 R_{u_1} , 其差异度为 D_{u_1} . 每选出一个数据之后, 需要对剩余的数据进行差异度修正. 设已经选出 m ($1 < m < n_d$) 个数据, 剩余数据的修正差异度为

$$D_l = D_l - D_{u_m} \left[\exp \left(\sum_{i=1}^n \frac{-(x_i^l - x_i^{u_m})^2}{(\sigma_i/2)^2} \right) \times \exp \left(\frac{-(y^l - y^{u_m})^2}{(\sigma_y/2)^2} \right) \right]. \quad (11)$$

对所有剩余数据进行差异度修正后,从中选取差异度最大的数据作为下一条数据规则.如此重复该过程,直到剩余数据的最大差异度 $D_{u_m} \leq 0$ 为止.

2.1.3 经验规则与数据规则的融合

为了避免经验规则与数据规则重复,并控制模型的复杂程度,本文提出模糊规则融合方法来确定模型的初始规则.该方法首先选择所有数据规则作为初始规则,由于数据规则由样本生成,与经验规则相比,数据规则具有更高的准确性;然后根据经验规则与数据规则间的重合程度确定所需的经验规则.

定义 1 模糊规则 R_j 和 R_k ($j \neq k$) 在第 i 维输入空间的重合指数 $OL_i^{j,k}$ 为

$$OL_i^{j,k} = \left(\frac{OS}{AS} \right)^a = \left(\frac{\text{MaxSup}_i^j - \text{MinSup}_i^k}{\text{MaxSup}_i^k - \text{MinSup}_i^j} \right)^a, \quad (12)$$

$$a = \text{sgn}(\text{MaxSup}_i^k - \text{MaxSup}_i^j + \text{MinSup}_i^k - \text{MinSup}_i^j). \quad (13)$$

如图2所示, MaxSup 和 MinSup 分别为隶属函数主要作用范围的上限(u)和下限(l).对于经验规则,其上下限为经验设定值;而对于数据规则则是当隶属度为 κ 时的输入值.

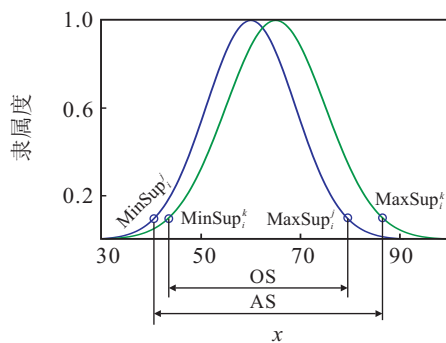


图2 相邻模糊规则重合区域示意图

定义 2 经验规则 RE_p 在第 i 维输入空间与数据规则的重合度为

$$OLe_i^p = \max(OL_i^{p,l}), \quad l = 1, 2, \dots, n_t, \quad (14)$$

其中 n_t 为数据规则的数量.

在确定了经验规则与数据规则的重合度之后,即可进行经验规则的选取,选取方法如下:首先,计算每条经验规则 RE_p 在每一维输入空间与数据规则的重合度 $OLe_i^p, i = 1, 2, \dots, n$;然后,选取至少有一维输入空间的重合度 $OLe_i^p < OLeM$ 的经验规则,将其与所有已选数据规则一起作为系统的初始规则.其中 $OLeM$ 为重合度控制参数.

2.2 参数辨识

对于仅使用数据规则的TSK型模糊系统,在参数辨识阶段使用梯度下降与最小二乘相结合的迭代方法优化模糊模型中的前提参数和结论参数^[15].

梯度下降方法用于确定前提参数 $\gamma_i^l, \sigma_i^l, i = 1, 2, \dots, n, l = 1, 2, \dots, n_t$, 其目标函数为

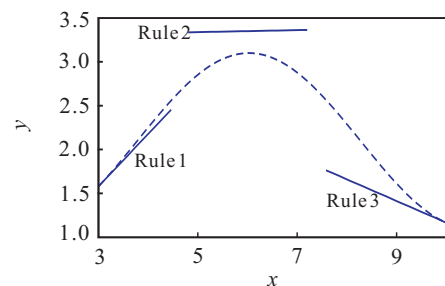
$$J_1(k) = \frac{1}{n_t} \sum_{p=1}^{n_t} e_p(k)^2 = \frac{1}{n_t} \sum_{p=1}^{n_t} (\hat{y}_p(k) - y_p)^2. \quad (15)$$

其中: $\hat{y}_p(k)$ 为第 k 次迭代时对于输入 x_p 的预测值, y_p 为与 x_p 对应的实际输出值.最小二乘算法用于优化数据规则结论部分的参数 $w_i^l, i = 0, 1, \dots, n, l = 1, 2, \dots, n_t$.由于经验规则的后件参数是由经验设定的常数,在参数优化时不对其进行自适应更新.

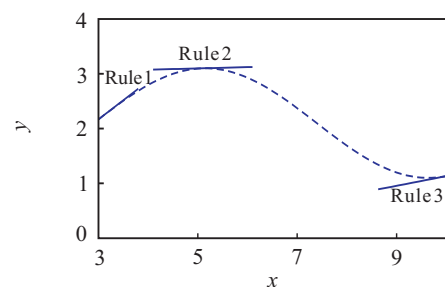
在上述混合迭代方法的基础上,为了结合经验知识,本文对前提参数的求解方法进行如下改进.

2.2.1 梯度下降方法中目标函数的改进

如果将模糊模型看作由模糊规则组成的全局模型,则模糊规则可看作在其作用范围内的局部模型.如果仅以训练误差最小化(J_1)为目标函数,则所建模型会出现局部模型与全局模型偏离的情况.如图3(a)中的全局模型是由3个局部模型构成的,为使全局模型的误差最小,Rule2与Rule1和Rule3的作用范围存在较大的重合,使得Rule2在另两个局部模型的影响下与全局模型产生较大偏离;而图3(b)中的3条规则相互重合程度较小,从而使局部模型更贴近全局模型.



(a) 局部模型 Rule2 与全局模型有较大偏离



(b) 局部模型 Rule2 与全局模型有较小偏离

图3 局部模型与全局模型偏离程度示意图

由此可见,为了使全局模型更接近由经验规则确定的局部模型,增强经验规则在其作用范围内的指导

作用,需降低其他数据规则对它的影响.为此,将原目标函数修改为

$$\begin{aligned}
 J(k) = & \frac{(1-\lambda)}{n_t} \sum_{p=1}^{n_t} (\hat{y}_p(k) - y_p)^2 + \\
 & \frac{\lambda}{n_e} \sum_{j=1}^{n_e} (\hat{y}(\gamma^{d_j}(k)) - z^{d_j}(\gamma^{d_j}(k)))^2 = \\
 & (1-\lambda)J_1(k) + \lambda J_2(k). \quad (16)
 \end{aligned}$$

改进的目标函数在原函数 J_1 的基础上增加了用于约束模型假设空间的正则化项 J_2 . 其中: z^{d_j} 表示第 j 条经验规则的输出,也表示第 d_j 条模糊规则的输出,即第 j 条经验规则对应第 d_j 条融合后的模糊规则; γ^{d_j} 表示第 j 条经验规则前提部分的中心点; J_2 表示经验规则在其中心点处的局部模型与全局模型输出值的偏差,偏差值越小,表示数据规则对经验规则的影响程度越小,从而在经验规则作用范围之内模糊模型输出值更接近经验规则输出值,也就是使得模型输出值更符合经验知识; $\lambda \in [0, 1]$ 为折衷系数,用于平衡数据与经验知识对模型的影响程度.

虽然在降低 J_2 的同时, J_1 可能会有所增加,但 J_1 的最小化与 J_2 的最小化之间并不矛盾.这是因为最小化 J_1 的目的在于降低训练数据的拟合误差,以提高模型的预测精度.但是,在训练数据不充足的情况下, J_1 的最小化只代表了在训练数据覆盖区域内的拟合误差最小,此时已不能表示预测模型的整体预测精度,而 J_2 项的引入则弥补了该不足,它表示在训练数据未覆盖区域的预测精度.因此,只有 J_1 与 J_2 的结合项 J 才能表示模型在整体输入空间的预测精度.

改进后模型中第 l 条规则的前提参数 $\gamma_i^l(k+1)$ 和 $\sigma_i^l(k+1)$ 的求取方法如下:

$$\begin{aligned}
 \gamma_i^l(k+1) = & \gamma_i^l(k) - \eta \left((1-\lambda) \frac{\partial J_1(k)}{\partial \gamma_i^l} \Big|_{\gamma_i^l = \gamma_i^l(k)} + \right. \\
 & \left. \lambda \frac{\partial J_2(k)}{\partial \gamma_i^l} \Big|_{\gamma_i^l = \gamma_i^l(k)} \right), \quad (17)
 \end{aligned}$$

$$\begin{aligned}
 \sigma_i^l(k+1) = & \sigma_i^l(k) - \eta \left((1-\lambda) \frac{\partial J_1(k)}{\partial \sigma_i^l} \Big|_{\sigma_i^l = \sigma_i^l(k)} + \right. \\
 & \left. \lambda \frac{\partial J_2(k)}{\partial \sigma_i^l} \Big|_{\sigma_i^l = \sigma_i^l(k)} \right). \quad (18)
 \end{aligned}$$

其中

$$\frac{\partial J_2(k)}{\partial \gamma_i^l} \Big|_{\gamma_i^l = \gamma_i^l(k)} = \frac{2}{n_e} \sum_{j=1}^{n_e} (e^{d_j}(k) A_j(k)). \quad (19)$$

$$\frac{\partial J_2(k)}{\partial \sigma_i^l} \Big|_{\sigma_i^l = \sigma_i^l(k)} = \frac{2}{n_e} \sum_{j=1}^{n_e} (e^{d_j}(k) C_j(k)). \quad (20)$$

$$A_j(k) = \begin{cases} \frac{\prod_{q=1}^n \mu_{F_q^g}(\gamma_q^{d_j}(k), \gamma_q^g(k), \sigma_q^g(k))}{\sum_{g=1, g \neq d_j}^{n_r} \left(\frac{\prod_{q=1}^n \mu_{F_q^g}(\gamma_q^{d_j}(k), \gamma_q^g(k), \sigma_q^g(k))}{\sum_{g'=1}^{n_r} \prod_{q=1}^n \mu_{F_q^{g'}}(\gamma_q^{d_j}(k), \gamma_q^{g'}(k), \sigma_q^{g'}(k))} \right)} \times \\ \left(\frac{\gamma_i^{d_j}(k) - \gamma_i^g(k)}{\sigma_i^g(k)^2} z^g(\gamma^{d_j}(k)) + w_i^g \right), \quad l = d_j; \\ \frac{\prod_{q=1}^n \mu_{F_q^l}(\gamma_q^{d_j}(k), \gamma_q^l(k), \sigma_q^l(k))}{\sum_{g=1}^{n_r} \prod_{q=1}^n \mu_{F_q^g}(\gamma_q^{d_j}(k), \gamma_q^g(k), \sigma_q^g(k))} \times \\ \frac{\gamma_i^{d_j}(k) - \gamma_i^l(k)}{(\sigma_i^l(k))^2} z^l(\gamma^{d_j}(k)), \quad l \neq d_j. \end{cases} \quad (21)$$

$$C_j(k) = \begin{cases} 0, \quad l = d_j; \\ \frac{\prod_{q=1}^n \mu_{F_q^l}(\gamma_q^{d_j}(k), \gamma_q^l(k), \sigma_q^l(k))}{\sum_{g=1}^{n_r} \prod_{q=1}^n \mu_{F_q^g}(\gamma_q^{d_j}(k), \gamma_q^g(k), \sigma_q^g(k))} \times \\ \frac{(\gamma_i^{d_j}(k) - \gamma_i^l(k))^2}{(\sigma_i^l(k))^3} z^l(\gamma^{d_j}(k)), \quad l \neq d_j. \end{cases} \quad (22)$$

$$\begin{aligned}
 \frac{\partial J_1(k)}{\partial \gamma_i^l} \Big|_{\gamma_i^l = \gamma_i^l(k)} = & \sum_{p=1}^{n_t} \left[\frac{2e_p(k)}{n_t} \frac{\prod_{q=1}^n \mu_{F_q^l}(x_q^p(k), \gamma_q^l(k), \sigma_q^l(k))}{\sum_{g=1}^{n_r} \left(\prod_{q=1}^n \mu_{F_q^g}(x_q^p(k), \gamma_q^g(k), \sigma_q^g(k)) \right)} \times \right. \\
 & \left. \frac{x_i^p(k) - \gamma_i^l(k)}{(\sigma_i^l(k))^2} \right] z^l(\mathbf{x}^p). \quad (23)
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial J_1(k)}{\partial \sigma_i^l} \Big|_{\sigma_i^l = \sigma_i^l(k)} = & \sum_{p=1}^{n_t} \left[\frac{2e_p(k)}{n_t} \frac{\prod_{q=1}^n \mu_{F_q^l}(x_q^p(k), \gamma_q^l(k), \sigma_q^l(k))}{\sum_{g=1}^{n_r} \left(\prod_{q=1}^n \mu_{F_q^g}(x_q^p(k), \gamma_q^g(k), \sigma_q^g(k)) \right)} \times \right. \\
 & \left. \frac{(x_i^p(k) - \gamma_i^l(k))^2}{(\sigma_i^l(k))^3} \right] z^l(\mathbf{x}^p). \quad (24)
 \end{aligned}$$

$$e^{d_j}(k) = \hat{y}(\gamma^{d_j}(k)) - z^{d_j}(\gamma^{d_j}(k)). \quad (25)$$

η 为学习速率因子, $z^l(\mathbf{x})$ 为第 l 条模糊规则的输出值.

2.2.2 对经验规则中心点的限定

经验规则中隶属函数的中心点参数在参数寻优过程中自适应调整.由于 J_2 只考虑了模糊模型与经验规则在输入为中心点时输出值的偏差程度,调整后的中心点参数可能会出现远离经验设定值的情况,即

所引入模型的经验规则中包含的信息与经验知识存在较大偏差. 为此, 提出如下公式对经验规则中心点的可调整范围进行限定:

$$\gamma_i^p(k+1)^* = \begin{cases} \gamma_i^p + h_i^p(u_i^p - \gamma_i^p), & \gamma_i^p(k+1) \geq \gamma_i^p + h_i^p(u_i^p - \gamma_i^p); \\ \gamma_i^p - h_i^p(\gamma_i^p - l_i^p), & \gamma_i^p(k+1) \leq \gamma_i^p - h_i^p(\gamma_i^p - l_i^p); \\ \gamma_i^p(k+1), & \text{other.} \end{cases} \quad (26)$$

其中: γ_i^p 、 u_i^p 和 l_i^p 分别为经验规则 RE_p 在第 i 维输入的初始中心点以及作用范围的上下限; $h_i^p \in (0, 1]$ 表示对于经验知识的确定程度, 确定程度越大, h_i^p 值可设定得越小, 从而保证所引入模型中的经验知识的准确性.

2.3 建模步骤

Step 1: 设定参数 $\kappa, K_i, K_y, \lambda, OLeM, h_i^p, i = 1, 2, \dots, n, p = 1, 2, \dots, n_e$, 以及参数迭代算法的最大迭代次数 n_e 和目标训练误差 E_r .

Step 2: 结构辨识. 根据式 (5)~(7) 将经验知识表达为 TSK 型经验规则; 使用减聚类方法, 按式 (8)~(11) 确定数据规则; 按式 (14) 计算经验规则重合指数, 并按规则融合方法确定初始模糊规则.

Step 3: 参数辨识. 使用改进的前提参数辨识方法, 按式 (16)~(26) 优化前提参数, 使用最小二乘方法确定数据规则中的结论参数.

Step 4: 判断是否满足迭代停止条件 (达到最大迭代次数或满足目标训练误差). 如果没有满足条件, 则转 Step 3 继续迭代, 否则保留优化结果并停止.

3 实验与分析

3.1 数值仿真实验

使用的测试函数为

$$f(x) = \cos(0.0035x^2)x^3 \cdot 10^{-4}. \quad (27)$$

在 [30, 100] 范围内随机选取 25 个输入值, 将与其对应的 $(x, f(x))$ 数据对作为训练样本. 在输入范围内以 0.5 为间隔选取 141 个输入值, 并将与其对应的数据对作为测试样本. 根据函数拐点处的相关知识确定参数 l^p, u^p 和 $z^p, p = 1, 2, \dots, 10$, 并对每个经验参数加入 $(-1, 1)$ 范围的噪声, 最终生成的经验规则如下:

- Rule 1: If x is $FE_1^1(37.9, 47.8)$, Then $z^1 = 7.8$;
- Rule 2: If x is $FE_2^2(47.9, 58.5)$, Then $z^2 = -13.7$;
- Rule 3: If x is $FE_3^3(54.6, 65.6)$, Then $z^3 = 21.6$;
- Rule 4: If x is $FE_4^4(61.9, 71.1)$, Then $z^4 = -30.1$;
- Rule 5: If x is $FE_5^5(68.4, 79.4)$, Then $z^5 = 39.4$;
- Rule 6: If x is $FE_6^6(73.9, 83.3)$, Then $z^6 = -48.8$;
- Rule 7: If x is $FE_7^7(79.7, 90.2)$, Then $z^7 = 60.7$;

Rule 8: If x is $FE_1^8(87.4, 92.5)$, Then $z^8 = -72.7$;

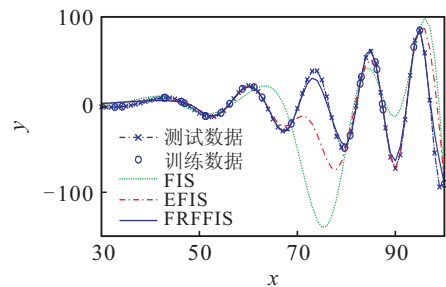
Rule 9: If x is $FE_1^9(92.5, 97.8)$, Then $z^9 = 84.5$;

Rule 10: If x is $FE_1^{10}(96.5, 99.5)$, Then $z^{10} = -93.9$.

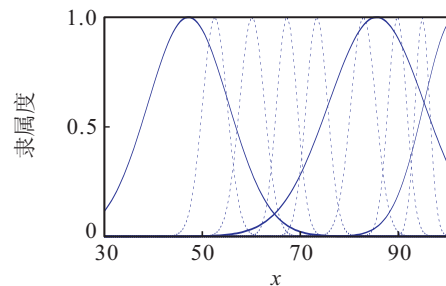
实验对只使用训练数据的模糊推理系统 (FIS) 方法和只使用本文的结构辨识方法结合经验知识的模糊推理系统 (EFIS) 方法, 以及本文提出的 FRFFIS 方法进行比较. 共重复 50 次仿真比较, 每次随机生成一组训练数据. 参数设定如下: $\kappa = 0.1$; $K_i = K_y = 8$; $OLeM = 0.7$; $h^p = 0.8, p = 1, 2, \dots, 10$; $n_e = 100$; $E_r = 1$. 为比较 FRFFIS 方法中 λ 值对预报结果的影响, 使用 $[0, 1]$ 范围内的不同 λ 值建立多个 FRFFIS 模型 ($\lambda = 0$ 时的 FRFFIS 方法即为 EFIS 方法). 预报结果如表 1 和图 4 所示.

表 1 不同模型多次建模的预报误差比较

方法	误差	方法	误差
FIS	31.89 ± 25.38	FRFFIS ($\lambda = 0.5$)	14.08 ± 4.22
EFIS	24.92 ± 21.16	FRFFIS ($\lambda = 0.6$)	14.59 ± 3.04
FRFFIS ($\lambda = 0.1$)	14.10 ± 3.72	FRFFIS ($\lambda = 0.7$)	13.04 ± 2.58
FRFFIS ($\lambda = 0.2$)	12.82 ± 3.02	FRFFIS ($\lambda = 0.8$)	14.19 ± 2.83
FRFFIS ($\lambda = 0.3$)	14.53 ± 3.67	FRFFIS ($\lambda = 0.9$)	14.46 ± 2.66
FRFFIS ($\lambda = 0.4$)	14.32 ± 2.54	FRFFIS ($\lambda = 1.0$)	14.24 ± 3.15



(a) FIS、EFIS 与 FRFFIS($\lambda=0.2$) 预测输出值比较



(b) FRFFIS($\lambda=0.2$) 预测模型中的隶属函数

图 4 不同模型的预报结果比较

表 1 中的误差包含实际值和预报值的均方根误差 (RMSE) 的平均值和方差. 由表 1 可以看出: 只使用样本数据建模 (FIS 方法) 的方法存在较大的预报误差; 结合经验规则 (EFIS 方法) 后的预报误差均值和方差都有所减小; 在结合经验规则并使用改进的参数辨识方法 (FRFFIS 方法) 后, 预报误差的均值和方差进一步降低; 当 $\lambda = 0.2$ 时, FRFFIS 方法具有最小的预

报误差均值和较小的误差方差;当 λ 的数值远离0.2时,预报误差的均值随之增大。

图4为不同方法预报结果的比较.其中的隶属函数曲线由FRFFIS($\lambda = 0.2$)方法生成,虚线表示经验规则的隶属度函数,实线表示数据规则的隶属度函数.由图4可以看出:在输入为(63, 69)、(69, 80)和(87, 94)这3个训练样本没有覆盖的区间内,规则融合方法选择了经验规则Rule 4、Rule 5和Rule 7进行弥补时,与FIS方法相比,提高了模型在该区间的预报精度,从而在整体上提高了模型的预报可靠性;FRFFIS方法还通过引入Rule 2、Rule 3、Rule 8和Rule 9对函数的细节进行描述,也起到了提高模型预报精度的作用;EFIS和FRFFIS两种方法虽然选取了相同的数据规则和经验规则,但这两种方法的预报精度存在一定的差别,尤其在(69, 80)和(95, 100)输入空间内差别较大,这是因为FRFFIS方法将正则化约束项引入目标函数,使得经验知识的指导作用更为明显。

FRFFIS($\lambda = 0.2$)方法最终得到的数据规则为

If x is $MF_1^1(47.14, 11.64)$,

Then $z^1 = -0.1121x + 2.7519$;

If x is $MF_1^7(85.55, 13.81)$,

Then $z^7 = -20.7612x + 9.1806$;

If x is $MF_1^{10}(101.6, 9.051)$,

Then $z^{10} = -256.5138x + 19753.1268$.

其中语言项 $MF_i^j(\gamma_i^j, \sigma_i^j)$ 的隶属函数为

$$MF_i^j(\gamma_i^j, \sigma_i^j) = \exp\left(\frac{-(x - \gamma_i^j)^2}{(\sigma_i^j)^2}\right). \quad (28)$$

参数调整后得到的最终经验规则为

If x is $FE_1^2(47.18, 57.94)$, Then $z^2 = -13.7$;

If x is $FE_1^3(54.69, 65.61)$, Then $z^3 = 21.6$;

If x is $FE_1^4(62.74, 71.78)$, Then $z^4 = -30.1$;

If x is $FE_1^5(68.68, 78.06)$, Then $z^5 = 39.4$;

If x is $FE_1^6(77.42, 88.32)$, Then $z^6 = 60.7$;

If x is $FE_1^8(85.74, 93.86)$, Then $z^8 = -72.7$;

If x is $FE_1^9(91.73, 98.06)$, Then $z^9 = 84.5$.

3.2 工程实例研究

元素收得率是钢水精炼过程中计算合金配料方案的重要参数.元素收得率定义为钢水中某元素含量的变化量与该元素总加入量的比值,其值越大表示元素的利用率越高.对于该参数的确定,目前以人为判断为主,但为了提高钢水质量、降低生产成本,需要更为精确的估计值.对于钢包炉(LF)精炼的合金化过程而言,它包含复杂的物理化学反应,无法确定准确的收得率机理模型,因此只能通过建立数据驱动模型进行预报.因为精炼过程的周期较长(20~30 min),而且炉况复杂多变,所以可用的建模数据较少,往往不能

满足预报可靠性的要求.为此,本文使用FRFFIS方法以硅(Si)和锰(Mn)两种元素为例建立其收得率预报模型,并与FIS、EFIS方法进行比较。

精炼操作流程如图5所示.其中: s_0 和 T_0 分别为进入LF工位之前的氩后钢水成分样和钢水温度; s_1 和 T_1 分别为在LF工位第1次采集的钢水成分样和钢水温度; s_2 和 T_2 分别为在LF工位第2次采集的钢水成分样和钢水温度,如果此时成分不合格,则需要进行第2次合金化; s_3 和 T_3 分别为在精炼即将结束,钢水离开LF工位前采集的成品钢水成分样和钢水出钢温度,其中钢水成分样中包含多种元素在钢水中的质量百分含量数据。

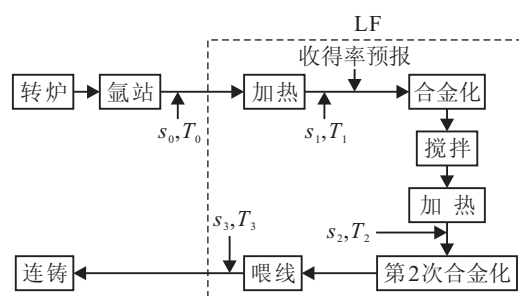


图5 精炼流程示意图

根据冶炼机理,模型选用的输入如下:钢水初始碳含量($w[C_0]$),钢水初始硫含量($w[S_0]$),钢水碳元素含量变化率($\Delta w[C]$),钢水锰元素含量变化率($\Delta w[Mn]$),钢水硅元素含量变化率($\Delta w[Si]$),钢水硫元素含量变化率($\Delta w[S]$),吹氩搅拌强度(D_P),加渣量(m_S),钢水温度(T)和钢水质量(m_t)。

$\Delta w[C]$ 的计算方法如下:

$$\Delta w[C] = \frac{w[C_1] - w[C_0]}{t_s}. \quad (29)$$

其中: $w[C_1]$ 和 $w[C_0]$ 分别为 s_1 和 s_0 样中的碳元素质量百分含量; t_s 为从冶炼开始到第1次采样的间隔时间。 $\Delta w[Mn]$, $\Delta w[Si]$ 和 $\Delta w[S]$ 的计算方法与 $\Delta w[C]$ 相同。

D_P 用以评价渣面的翻腾对电流的影响,从而可以间接反映吹氩强度,其计算方法为

$$D_P = \frac{1}{3}(D_A + D_B + D_C), \quad (30)$$

$$D_A = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_A^i - \bar{x}_A)^2}. \quad (31)$$

其中: D_A 、 D_B 和 D_C 为A、B、C三相电流的波动情况; D_A 的计算方法如式(31)所示; x_A^i 表示A相在 i 时刻的电流值, \bar{x}_A 为A相的平均电流值; D_B 、 D_C 与 D_A 的计算方法相同。

m_S 为在精炼工位加入的总渣量之和,对于硅镇静钢, m_S 为加入的石灰与萤石的渣量之和。

$$T = (T_1 + T_E)/2.$$

其中: T_1 为进站后第 1 次钢水测温温度, T_E 为钢水出站的目标温度, T 用于间接反应精炼过程中的整体温度情况. 以 Q345B 钢种为例, 各输入的数值范围如表 2 所示.

表 2 模型输入变量的数值范围

输入	范围
$w[C_0]$	[0.08%, 0.15%]
$w[S_0]$	[0.018%, 0.051%]
$\Delta w[C] / \text{min}^{-1}$	[0%, 0.004%]
$\Delta w[Mn] / \text{min}^{-1}$	[0%, 0.03%]
$\Delta w[Si] / \text{min}^{-1}$	[0%, 0.009%]
$\Delta w[S] / \text{min}^{-1}$	[-0.0029%, -0.0007%]
D_P / kA	[0, 6]
m_S / t	[0.2, 1.2]
$T / ^\circ\text{C}$	[1 535, 1 569]
m_t / t	[95, 100]

根据冶炼经验确定如下 4 种情况下硅 (Si)、锰 (Mn) 元素收得率预报的经验规则.

Rule 1 : If $w[C_0]$ is FP(0.11%, 0.12%) and $w[S_0]$ is FP(0.032%, 0.037%) and $\Delta w[C]$ is FP(0.001 7%, 0.002 3%) and $\Delta w[Mn]$ is FP(0.012 5%, 0.017 5%) and $\Delta w[Si]$ is FP(0.005%, 0.006%) and $\Delta w[S]$ is FP(-0.002%, -0.001 6%) and D_P is FP(2.5, 3.5) and m_S is FP(0.616 7, 0.783 3) and T is FP(1 549, 1 555) and m_t is FP(97, 98); Then $z_{Si} = 0.990$, $z_{Mn} = 0.980$.

Rule 2 : If $w[C_0]$ is FP(0.12%, 0.13%) and $w[S_0]$ is FP(0.032%, 0.037%) and $\Delta w[C]$ is FP(0.002 2%, 0.002 8%) and $\Delta w[Mn]$ is FP(0.012 5%, 0.017 5%) and $\Delta w[Si]$ is FP(0.003%, 0.005%) and $\Delta w[S]$ is FP(-0.002%, -0.001 6%) and D_P is FP(2.5, 3.5) and m_S is FP(0.616 7, 0.783 3) and T is FP(1 549, 1 555) and m_t is FP(97, 98); Then $z_{Si} = 0.945$, $z_{Mn} = 0.900$.

Rule 3 : If $w[C_0]$ is FP(0.11%, 0.12%) and $w[S_0]$ is FP(0.036%, 0.041%) and $\Delta w[C]$ is FP(0.001 7%, 0.002 3%) and $\Delta w[Mn]$ isFP(0.012 5%, 0.017 5%) and $\Delta w[Si]$ isFP(0.003%, 0.005%) and

$\Delta w[S]$ is FP(-0.002%, -0.001 6%) and D_P is FP(2.5, 3.5) and m_S is FP(0.616 7, 0.783 3) and T is FP(1 549, 1 555) and m_t is FP(97, 98); Then $z_{Si} = 0.950$, $z_{Mn} = 0.955$.

Rule 4 : If $w[C_0]$ is FP(0.11%, 0.12%) and $w[S_0]$ is FP(0.032%, 0.037%) and $\Delta w[C]$ is FP(0.001 7%, 0.002 3%) and $\Delta w[Mn]$ is FP(0.012 5%, 0.017 5%) and $\Delta w[Si]$ is FP(0.003%, 0.005%) and $\Delta w[S]$ is FP(-0.001 7%, -0.001 3%) and D_P is FP(2.5, 3.5) and m_S is FP(0.616 7, 0.783 3) and T is FP(1 549, 1 555) and m_t is FP(97, 98); Then $z_{Si} = 0.985$, $z_{Mn} = 0.980$.

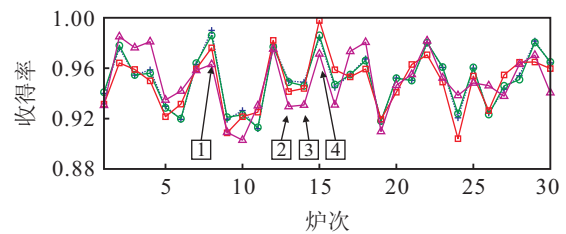
在 4 组经验规则覆盖区域之外随机选取 370 组训练数据, 26 组测试数据. 在测试数据中再加入 4 个在经验规则覆盖区域之内的样本数据, 样本数据与经验规则一一对应, 以此模拟训练数据覆盖范围不足的情况. 相关参数设定如下:

$$\lambda = 0.2; \kappa = 0.1; \text{OLeM} = 0.65;$$

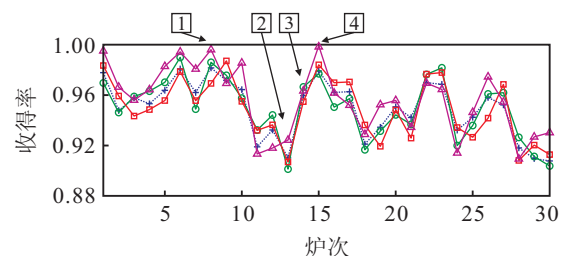
$$h_i^p = 0.5; p = 1, 2, \dots, 4; i = 1, 2, \dots, 10;$$

$$n_e = 100; E_r = 0.003 5.$$

在使用减聚类方法选取初始数据规则时, 对于 Si 元素, 参数 $K_i = K_y = 11.36$; 对于 Mn 元素, 参数 $K_i = K_y = 11.12$. 使用不同方法所建模型的预报情况如图 6 和表 3 所示.



(a) 不同建模方法对 Si 元素收得率预报结果比较



(b) 不同建模方法对 Mn 元素收得率预报结果比较

图 6 不同建模方法对元素收得率预报结果比较

---+--- 真实值 -○- FRFFIS -□- EFIS -△- FIS

表 3 不同方法的误差比较 10^{-2}

误差统计(元素)	建模方法		
	FRFFIS	EFIS	FIS
RMSE(Si)	0.36	0.99	1.50
RMSE(Mn)	0.99	1.13	1.39
MAE(Si)	0.40	1.66	2.70
MAE(Mn)	1.44	1.77	2.29

从预报结果可以看出, FRFFIS 方法对 4 个特殊样本点的预报精度均高于其他方法, 而且对于两种元素的预报值均有最低的均方根误差. 从冶炼工艺分析, Q345B 钢种对于元素收得率估计的允许最大绝对偏差(MAE)为 1.5×10^{-2} , 3 种方法中只有 FRFFIS 方法对于两种元素的预报误差均满足该要求.

4 结 论

本文提出了一种融合训练数据与经验知识的模糊建模方法, 用于解决训练数据不充足, 但具有较多领域经验知识的非线性回归问题. 通过理论分析和仿真实验, 可以得出以下结论:

1) 本文提出的规则融合方法可以有效选取经验规则, 弥补训练数据的不足.

2) 改进的梯度下降方法中的目标函数可以平衡数据和经验知识两种信息对于模型性能的影响; 对于经验规则中心点自适应调整范围的限定, 保证了经验知识可以较准确地融入预测模型中.

3) FRFFIS 方法可以有效地结合经验知识和样本数据, 提高训练数据没有覆盖的输入区域的预报精度, 从而使所建模型具有较高的预报可靠性.

综上所述, 本文提出的 FRFFIS 方法具有理论可行性, 在对可靠性要求较高的离线建模领域具有一定的应用潜力.

参考文献(References)

- [1] Kosko B. Fuzzy systems as universal approximators[C]. Proc of the IEEE Int Conf on Fuzzy Systems. San Diego: IEEE, 1992: 1153-1162.
- [2] Wang L X, Mendel J M. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning[J]. IEEE Trans on Neural Networks, 1992, 3(5): 807-814.
- [3] Zhou Q, Wu Y X, Chan C W, et al. From neural network to neuro-fuzzy modeling: Applications to the carbon dioxide capture process[J]. Energy Procedia, 2011, 4(0): 2066-2073.
- [4] Mehrabi M, Pesteei S M, Pashae G T. Modeling of heat transfer and fluid flow characteristics of helicoidal double-pipe heat exchangers using adaptive neuro-fuzzy inference system(ANFIS)[J]. Int Communications in Heat and Mass Transfer, 2011, 38(4): 525-532.
- [5] Gregor G, Gordon L. Nonlinear system identification: From multiple-model networks to Gaussian processes[J]. Engineering Applications of Artificial Intelligence, 2008, 21(7): 1035-1055.
- [6] Jukka K, Nurminen O K, Kimmo H. What makes expert systems survive over 10 years-empirical evaluation of several engineering applications[J]. Expert Systems with Applications, 2003, 24(2): 199-211.
- [7] Petr K, Bogdan G, Sibylle S. Data-driven soft sensors in the process industry[J]. Computers & Chemical Engineering, 2009, 33(4): 795-814.
- [8] Tewari A, Macdonald M U. Knowledge-based parameter identification of TSK fuzzy models[J]. Applied Soft Computing, 2010, 10(2): 481-489.
- [9] Eyke H. Fuzzy sets in machine learning and data mining[J]. Applied Soft Computing, 2011, 11(2): 1493-1505.
- [10] Chen X F, Gui W H, Wang Y L, et al. An integrated modeling method for prediction of sulfur content in agglomerate[J]. J of Central South University of Technology, 2003, 10(2): 145-150.
- [11] Polak A G. An error-minimizing approach to regularization in indirect measurements[J]. IEEE Trans on Instrumentation and Measurement, 2010, 59(2): 379-386.
- [12] Tewari A. Prior knowledge based identification of Takagi-Sugeno-Kang fuzzy models for static nonlinear systems[D]. Pennsylvania: The Pennsylvania State University, 2009: 36-37.
- [13] Benavoli A, Chisci L, Farina A, et al. Hard-constrained versus soft-constrained parameter estimation[J]. IEEE Trans on Aerospace and Electronic Systems, 2006, 42(4): 1224-1239.
- [14] Miguel A V, Omar S, Sixto R, et al. A new methodology to improve interpretability in neuro-fuzzy TSK models[J]. Applied Soft Computing, 2010, 10(2): 578-591.
- [15] Jang R. ANFIS: Adaptive-network-based fuzzy inference system[J]. IEEE Trans on System, Man and Cybernetics, 1993, 23(3): 665-685.
- [16] Shooehdeli M A, Teshnehlab M, Sedigh A K, et al. Identification using ANFIS with intelligent hybrid stable learning algorithm approaches and stability analysis of training methods[J]. Applied Soft Computing, 2009, 9(2): 833-850.
- [17] Eftekhari M, Katebi S D. Extracting compact fuzzy rules for nonlinear system modeling using subtractive clustering, GA and unscented filter[J]. Applied Mathematical Modelling, 2008, 32(12): 2634-2651.