DOI: 10.7641/CTA.2015.41083

基于子集融合与规则简约的磨矿过程模糊建模

乔 峥, 刘 颖†, 赵 珺, 王 伟, 郭 戈

(大连理工大学 信息与控制研究中心, 辽宁 大连 116023)

摘要: 针对选矿厂磨矿生产过程的模糊建模问题, 本文提出一种基于模糊集融合和规则简约的模糊建模方法. 该方法针对基于数据建立的磨矿过程Takagi-Sugeno模型, 采用模糊C均值聚类方法对同一变量下的隶属度函数参数进行聚类, 得到对不同工况具有代表性的融合后的隶属度函数, 来降低过度拟合的影响. 此外, 本文根据规则库中的规则权值, 对前件相同的冗余规则进行约简, 形成最终的离线模糊规则库, 有效提高了规则库的泛化能力. 为验证本文方法的有效性, 分别采用经典数据与实际工业数据进行了实验论证, 从精度和泛化能力上体现了本文方法的优势.

关键词: 磨矿; Takagi-Sugeno模型; 泛化性; 模糊子集; 模糊规则

中图分类号: TP206+.3 **文献标识码**: A

Grinding process modeling based on fuzzy sets merging and rule simplification

QIAO Zheng, LIU Ying†, ZHAO Jun, WANG Wei, GUO Ge

(Research Center of Information and Control, Dalian University of Technology, Dalian Liaoning 116023, China)

Abstract: In modeling the typical grinding process, we propose a fuzzy modeling method based on fuzzy sets merging and rule simplification. In the fuzzy rule extraction process for the Takagi-Sugeno model, the proposed method adopts fuzzy C mean clustering to partition the fuzzy membership function of every variable to obtain the representative merged membership functions for every working condition to reduce the negative impact of the over-fit phenomenon. Besides, using the weight of each rule, we simplify the fuzzy rule-base by merging the fuzzy rules with the same premises to obtain the final fuzzy model. To validate the proposed approach, a series of comparative experiments are carried out by using classic data and industrial data. The experimental results demonstrate a remarkable generalization ability and practical application potential of the proposed method.

Key words: grinding; Takagi-Sugeno model; generalization; fuzzy sets; fuzzy rules

1 引言(Introduction)

磨矿工业过程复杂, 生产线控制变量多, 且关键变量如给矿, 给水等对磨矿过程的产出有重要影响. 变量间的强非线性动态关系致使磨矿控制模型较难用精确的数学模型来描述. 国内企业普遍需要操作员依据自身经验采用手动方式对控制变量进行实时调控, 然而鉴于操作员的主观经验限制, 工况的复杂和边界条件的多变性, 人工控制方式往往难以达到预期的生产目标, 其控制方案的一致性较差[1–2].

模糊模型是一种描述非线性多变量实际系统的有效方法[3–4]. 早期的Mamdani模糊模型主要用于解决分类问题[5–6], 文[7]提出的Takagi-Sugeno模糊建模方法, 大大提高了模糊模型的推理精度, 使其成为应用较为广泛的模糊模型构建方法[8]. 基于数据的模糊系统建模是通过历史数据来学习, 获得控制规则[9–10], 其中模糊神经网络法利用神经网络的学习特性, 通过对神经网络的训练来形成最终的模糊模型[11–12]. 聚类方法通过分析数据样本间的相似性, 并对其进行聚类分析来建立模糊规则[13]. 此外, 利用遗传算法提取模糊规则的方法可以在降维的同时获得鲁棒性更好的系统[14–15]. 然而以上这些方法对于数据波动频繁的复杂工业过程建模, 往往难以得到很好的效果.

规则提取过程中直接经由数据产生模糊规则[16], 往往存在大量相似模糊集的重叠问题, 易使规则库产

收稿日期: 2014−11−23; 录用日期: 2015−04−03.

†通信作者. E-mail: liuying8227@163.com.

国家自然科学基金项目(61273037, 61304213, 61473056), 国家“863”计划项目(2013AA040703), 中央高校基本科研业务费专项资金项目(DUT13RC203)资助.

Supported by National Natural Science Foundation of China (61273037, 61304213, 61473056), National High-Tech Research and Development Program (2013AA040703) and Fundamental Research Funds for the Central Universities (DUT13RC203).

生针对训练样本的过拟合, 降低模型泛化能力[17]. 针对模糊集重叠问题, 当前较为普遍的方法是从隶属度函数角度计算模糊集之间的相似性[18–19], 通过相似模糊集的融合等处理来消除相似重叠的模糊子集. 文[20]从几何角度描述了模糊集的相似度. 文[21]提出一种模糊集相似指数来分析模糊集的相似程度, 但是相似指数的定义推理过程自身存在误差, 导致后续处理也将包含这些误差. 文[22]提出了基于包含指数的相似性计算方法, 期望得到能够最大包含相似模糊集的新的模糊子集, 代替原来的模糊集, 然而仅用包含程度作为评价指标在处理一些论域范围较小的模糊集时是不合理的, 易造成知识缺失.

本文针对磨矿工业过程模糊控制规则库构建过程中, 存在大量相似模糊子集重叠, 导致规则库针对训练样本产生过拟合的现象, 提出一种模糊集融合方法, 通过聚类分析的方法得到相似隶属度函数的集合, 用聚类中心的相应值代替原隶属度函数, 使原相似的隶属度函数簇融合为较少的、有代表性的新隶属度函数. 针对模糊隶属度函数替换后的规则库含有冗余规则的情况, 本文根据各隶属度函数的权值, 计算各个规则的权重, 根据规则权重对规则库原规则参数进行加权平均, 得到新规则, 消除冗余, 最终得到拥有较好泛化性的规则库.

为了验证本文所提方法的有效性, 分别对Mackey-glass时间序列数据与磨矿实际生产数据进行了实验分析, 并采用多种模糊建模方法作为对比实验, 从实验结果的精度和模糊模型的结构方面体现出本文方法的优势.

2　问题描述(Description of the problem)

选矿厂磨矿过程作为重要的环节, 向浮选过程提供单体解离度尽可能高的矿石, 对最后的选矿品位以及矿石的回收率有十分重要的影响. 图1为某铜钼矿选矿厂磨矿过程流程图, 铜钼原矿经粗碎、半自磨和筛分后, 筛上物料经顽石破碎回半自磨, 筛下物料进入渣浆泵池, 泵池内矿浆通过渣浆泵进入水力旋流器, 经水力旋流器分级, 溢流部分进入浮选作业, 沉砂则送球磨机再磨, 简称半自磨–球磨–破碎(SABC)流程.

SABC流程的控制可以划分为半自磨机回路与旋流器回路的控制. 半自磨机回路中主要需控制给矿量和半自磨机给水, 通过调节给矿量和给水量的比例, 调节半自磨机的工作浓度, 以适应不同的生产需要. 旋流器回路主要需控制渣浆泵频率和球磨机排矿水, 一方面调节渣浆泵频率, 保证旋流器工作在正常的压力和浓度下, 另一方面通过调节球磨机排矿水维持整个磨矿系统的浓度稳定.

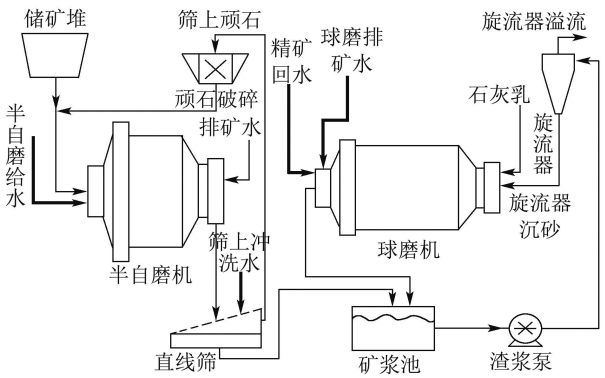


图 1　磨矿过程流程图

Fig. 1　Flow chart of the grinding process

由于生产线较长, 工况和边界条件复杂多变, 人工调整难以保证生产过程的持续稳定. 此外, 对于非线性复杂系统, 控制往往难以达到预期效果. 而采用模糊控制方法, 进行基于工业数据的模糊规则提取过程, 由于是完全基于样本数据的, 训练过程会尽可能产生针对样本数据误差最小的规则库, 而导致许多隶属度函数会在一定区域内密集分布, 但又难以完全重叠. 这些分布密集的隶属度函数实际上是拟合了样本集中同一工况下的由于噪声信号产生的扰动数据, 导致了过拟合现象的产生, 即针对样本中的数据可能有较高的精度, 但却失去了一般性, 从而使得在处理实际数据时精度较低.

3　模糊规则库构建与约简(Construction and simplification of the rule base)

3.1　初始规则库(Initial rule base)

本文采用模糊模型建模方法对磨矿生产过程相关量的控制规则进行建模, 得到规则形式如下的初始规则库, 其后件为一阶线性函数, 即

$$R_i: \text{If } x_1 \text{ is } A_{i1} \text{ and } \cdots \text{ and } x_p \text{ is } A_{ip}, \quad (1)$$

$$\text{Then } y_i = b_{i0} + b_{i1}x_1 + \cdots + b_{ip}x_p, \quad (2)$$

其中: $i = 1, \cdots, M$, $A_{ij}(j = 1, \cdots, p)$为前件因素的模糊子集, p为前件因素数, x_j代表磨矿模糊系统中的前件变量值, 如给矿量、泵池液位、溢流粒度等, y为磨矿模糊系统的后件变量值, 即需要调整的球磨排矿水、半自磨机给水、渣浆泵频率, $b_{il}(l=0,1,\cdots,p)$为模糊后件参数.

3.2　基于模糊 C 均值聚类的模糊子集融合 (Fuzzy sets merging based on fuzzy C-means (FCM) clustering)

经过规则提取生成的磨矿过程规则库中, 同一变量下的隶属度函数分布, 大体分为如下几种情况, 如图2所示.

1) 位置不同: 如$mf1$与$mf6$等, 位置不同的隶属度函数代表了不同的工况;

2) 位置接近, 形状不同: 如$mf4$与$mf5$, 即使位置接近, 形状不同也要分作不同的工况;

3) 位置接近, 形状相似: 如$mf1$与$mf2, mf6$与$mf7$等, 此类相似的隶属度函数代表了同一种工况;

4) 针状集: 如$mf8$, 针状集代表了数据的异常.

其中, 情况1)和情况2)由于存在位置或者形状上较为明显的区别, 可通过隶属度函数的参数进行区分. 情况3)代表了基于数据提取的模糊规则库普遍存在的问题, 即相似隶属度函数的密集分布. 情况4)所描述的针状集对应的论域范围极窄, 产生针状集的原因往往是数据中存在明显的异常点, 因此在实际应用中, 其所对应的规则几乎不可能触发. 泛化性较好的模糊规则库, 能够区分出1)−2)两种情况的分布, 同时对于3)类隶属度函数分布, 能减轻相似隶属度函数导致的规则库推理结果过拟合, 所以对规则库的处理可以概括为区分出不同工况, 然后对相同工况下的隶属度函数进行处理的过程.

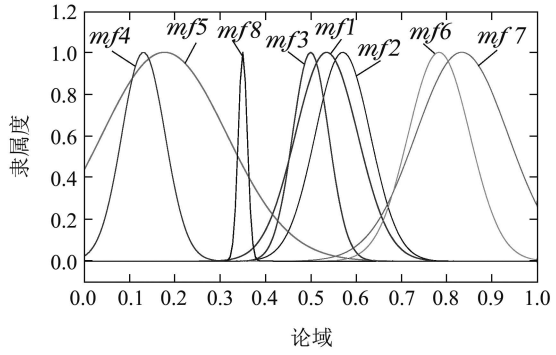


图 2 模糊子集的分布类型

Fig. 2 Distribution pattern of the fuzzy sets

考虑到隶属度函数参数可用于描述系统运行的不同工况. 其位置由均值代表, 其形状则由标准差表示. 为此, 本文采用均值m和方差σ表示隶属度函数特征, 采用模糊C均值聚类[23], 对规则库中每个前件因素x_i $(i=1,\cdots,p)$的隶属度函数参数$[m,\sigma]$进行聚类, 将数据集划分为重叠的子集, 其聚类结果通过最小化如下目标函数来实现:

$$\begin{aligned} &J(U,v_1,\cdots,v_c)=\sum_{i=1}^{n}\sum_{j=1}^{c}u_{ij}^{\alpha}d_{ij}^{2},\\ &\text{s.t.}\ \sum_{j=1}^{c}u_{ij}=1, \end{aligned} \tag{3}$$

其中: c为聚类个数, $v_j=[vm_j,v\sigma_j](j=1,\cdots,c)$代表聚类中心, d_{ij}表示数据点$[m_i,\sigma_i]$到聚类中心v_j的欧氏距离, 模糊参数$\alpha>1$. 权值矩阵U由隶属度值u_{ij}组成, u_{ij}代表了第i个数据样本$[m_i,\sigma_i]$对聚类中心v_j的隶属度值. 经过对样本数据进行规则提取, 规则库中的规则数目为M, 则每个输入因素都对应有M个模糊子集. 对于聚类个数的确定, 由于实际现场情况复杂多变, 数据波动大, 每个因素对应的工况数量不同, 此处采用人工经验指定每个因素的聚类个数.

考虑到同一聚类下的隶属度函数在位置和形状上都较为相似, 可做同一工况处理. 为提高模糊规则库的泛化性, 本文提出一种隶属度函数融合方法, 产生对特定工况最有代表性的隶属度函数, 表征一种最一般的情况, 以提高模糊规则库的泛化性. 经过聚类过程后, 相似的隶属度函数会被聚为同一簇, 该簇的聚类中心代表了一般情况, 则可通过如下方式获得对该工况有代表性的新隶属度函数参数: 聚类之后, 根据每个U中的隶属度值, 得到x_i中隶属度函数A_{ir}对应的聚类中心, 采用下式得到新的隶属度函数均值参数, 即

$$m_{ir(\text{new})}=\{v(\text{mean})_{ij}|\max(u_{rj})\}, \tag{4}$$

其中: u_{rj}代表聚类得到的隶属度值, $v(\text{mean})_{ij}$代表聚类中心v_{ij}的均值参数. 该式先得到A_{ir}对应隶属度最大的聚类, 然后取该聚类中心中的均值参数值代替原来的隶属度函数的均值参数. 融合后的隶属度函数要尽量多的包含原隶属度函数信息, 为此, 可以用该聚类中最大的标准差作为融合后的隶属度函数的标准差, 对应到图2情况, 用$mf7$的标准差作为所有隶属于该聚类的隶属度函数的新的标准差. 即

$$\sigma_{ir(\text{new})}=\max(\sigma_{vi}), \tag{5}$$

其中σ_{vi}为属于该聚类的隶属度函数的标准差组成的集合. 经过隶属度替换后, 规则库变为如下形式:

$$\begin{aligned} R_i:\ &\text{If } x_1 \text{ is } \mathrm{A}_{\text{new-}1j} \text{ and } \cdots \text{and } x_p \text{ is } \mathrm{A}_{\text{new-}pj},\\ &\text{Then } y_i=\ b_{i0}+\ b_{i1}x_1+\cdots+\ b_{ip}x_p, \end{aligned} \tag{6}$$

其中$A_{\text{new-}ij}$为经过替换后的新隶属度函数. 此时的规则库更能反映出实际情况下的操作调整规则, 因此具有更好的泛化性.

3.3 规则冗余约简(Redundancy reduction)

经过隶属度函数的融合替换后, 由于对每个变量进行了聚类, 使每个变量对应的隶属度函数个数减少为对应的聚类个数, 而规则数没有变化, 则可能会出现规则前件相同的模糊规则. 前件相同而后件不同的规则, 即可认为是冗余规则, 对冗余规则需要进行相应的处理, 在消除冗余的同时也能对规则数量进行优化. 为此, 本文提出一种基于规则权重的冗余规则约简方法. 对于前件相同的如下两条规则, 表示形式如下:

$$\begin{aligned} R_i:\ &\text{If } x_1 \text{ is } A_{1n} \text{ and } \cdots \text{ and } x_p \text{ is } A_{pm},\\ &\text{Then } \ y_i=\ b_{i0}+\ b_{i1}x_1+\cdots+\ b_{ip}x_p; \end{aligned} \tag{7}$$

$$\begin{aligned} R_j:\ &\text{If } \ x_1 \text{ is } A_{1n} \text{ and } \cdots \text{ and } x_p \text{ is } A_{pm},\\ &\text{Then } \ y_j=\ b_{j0}+\ b_{j1}x_1+\cdots+\ b_{jp}x_p. \end{aligned} \tag{8}$$

对于如上两条规则, 在其隶属度函数融合之前, 前件是不同的, 经过上文中的聚类分析, 每个原隶属度函数对于新隶属度函数中心都存在一个隶属度值, 该隶

属度值代表原隶属度函数与新的隶属度函数的相似程度. 通过该值可计算得到规则R_i对于融合后规则的相似度, 即

$$s_i = \prod_{j=1}^{p} u_{jn}, \tag{9}$$

其中u_{jn}为原隶属度函数对于新隶属度函数均值的隶属度值. 根据此值, 计算出规则R_i的权重系数

$$w_i = \frac{s_i}{\sum\limits_{i=1}^{l} s_i}, \tag{10}$$

其中l代表与R_i前件相同的规则数. 融合后的规则, 其后件直接与原规则库中前件一致的规则相关. 新的规则后件可以表示为原参数的加权和, 即

$$b_{(\mathrm{new})n} = b_{in}w_i + b_{jn}w_j,\ n = 0, 1, \cdots, p. \tag{11}$$

以上2条规则可以融合为1条新规则. 对于3条或者3条以上前件相同的规则, 也可以按照相同的方法进行处理.

通过模糊子集的融合以及规则的约简, 对规则提取之后的模型进行了进一步的处理, 简化了模型结构和冗余的同时, 处理了过拟合现象, 提高了泛化性. 为了提高模型的精度, 可利用模糊神经网络的训练特性对模型的参数进行优化调整.

完整模糊模型的构建过程如下:

Step 1　基于样本数据, 提取产生初始规则库;

Step 2　根据规则库的每个前件因素的实际分布情况, 分别确定其对应的工况个数, 即模糊集融合阶段的聚类个数;

Step 3　对规则库每条规则中的前件隶属度函数参数(均值和方差), 按输入因素分别进行FCM聚类;

Step 4　计算融合后的隶属度函数参数: 根据式(3)和式(4), 分别计算新隶属度函数的均值和标准差参数;

Step 5　根据每个隶属度函数所属的聚类中心, 进行隶属度函数的融合替换;

Step 6　对前件相同的多条规则进行约简: 计算前件相同的规则的权重, 根据权重计算约简后的规则后件参数;

Step 7　整定规则库参数.

4　实验应用效果分析 (Experiment and analysis)

为了验证本文方法的可行性, 首先采用经典数据Mackey-glass时间序列进行仿真实验, 然后介绍了文中的方法在磨矿系统中的应用, 并详细描述规则库的生成过程以及实验效果. 采用不同的建模方法进行了多组对比实验, 包括自适应模糊神经推理系统Anfis[24]、基于误差的数据驱动建模方法[16]、基于特征选择的T–S模型建模方法[8]. 这3种方法属于典型的基于数据提取的模糊模型建模方法, 且均未经过模糊集融合与规则约简. 实验采用MAPE和RMSE作为误差度量标准, 规则数量和模糊集数量来代表模型的复杂程度.

$$\mathrm{MAPE} = \frac{1}{S}\sum_{k=1}^{S}\frac{|y(k) - \hat{y}(k)|}{y(k)}, \tag{12}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{S}\sum_{k=1}^{S}\left(y(k) - \hat{y}(k)\right)^2}, \tag{13}$$

其中: S为样本数量, $\hat{y}(k)$为k时刻的模型输出.

4.1　含高斯白噪声的Mackey-glass时间序列预测 (Prediction of the Mackey-glass time series with Gaussian white noise)

Mackey-glass是经典时间序列预测实验的数据样本之一, 为模仿实际的含噪声信号, 在Mackey-glass时间序列样本中加入高斯白噪声, 来产生训练样本. 实验数据采用时滞参数为17的1000组连续的时间序列样本, 加入方差为0.001的高斯白噪声, 其中前500组为训练样本, 后500组作为验证样本. 取$x(t-30)$, $x(t-18), x(t-12), x(t)$作为4个模型输入量, 预测$x(t+1)$. 模糊集融合过程设置每个变量点的聚类数为3.

为了对模型精度以及泛化性改善效果进行比较, 应用了Anfis和基于特征选择的T–S模型建模方法, 以及数据驱动的方法进行对比实验. 每种方法分别生成两组规则数不同的规则库进行比较. 各实验中的参数设置尽量使最终规则库的规则数目或者隶属度函数数目相同, 对于文[8]中的方法, 为了获得有比较性的精度, 规则数目和隶属度函数数量设定值都要大于其他几种方法. 进行多组实验, 各方法的实验指标统计结果如表1所示, 其中, 本文方法拥有8条、11条规则的模型, 初始规则库规则数量分别为11, 17, 初始隶属度函数数量分别为44和68. 图3中列出了每个模型对于验证样本的预测效果.

从实验结果可以看出, 在与其他规则库拥有相同数目的规则时, 尽管本文的训练精度低于某些方法, 但是预测精度最高. 同时本文方法所建立的模糊模型中, 隶属度函数数量得到了大幅度减少, 模型结构最为简单, 对于海量数据建模, 可以有效地避免规则爆炸, 提高规则库的运行效率. 另一方面, 其他某些的方法中, 随着规则数的提升, 模糊子集数量也会增加, 训练效果都有所提升, 但相应的验证样本精度则产生了一定程度的下降, 说明随着规则数的增多, 各模型都对含噪声的训练样本产生了过拟合. 而本文的建模方法, 在规则数增加时, 由于限制了模糊子集融合过程的聚类数, 因此没有增加隶属度函数的数量, 而且预测精度有所提升, 说明本方法有效地解决了模型对训练样本的过拟合, 提高了泛化性.

表 1 含高斯白噪声的Mackey-glass时间序列模糊建模结果

Table 1 Fuzzy modeling results of the Mackey-glass time series with Gaussian white noise

方法	训练误差		测试误差		规则数量	隶属度函数数量
	MAPE	RMSE	MAPE	RMSE		
数据驱动的T–S模型	3.44E–02	3.71E–02	4.34E–02	4.58E–02	8	32
	2.59E–02	3.03E–02	4.49E–02	4.98E–02	16	64
Anfis	3.35E–02	3.70E–02	4.26E–02	4.57E–02	8	32
	2.71E–02	3.03E–02	4.50E–02	4.88E–02	16	64
基于特征选择的T–S模型	4.05E–02	4.41E–02	4.30E–02	4.69E–02	20	80
	3.61E–02	3.89E–02	4.32E–02	4.46E–02	30	120
本文方法	3.68E–02	4.01E–02	4.03E–02	4.40E–02	8	12
	3.57E–02	3.85E–02	3.85E–02	4.17E–02	11	12

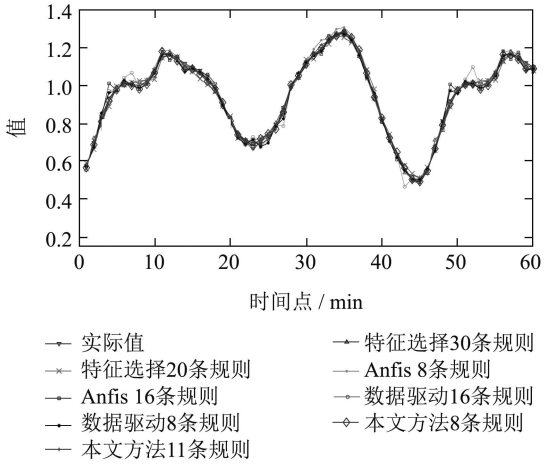


图 3 含高斯白噪声的Mackey-glass模糊建模预测效果对比图

Fig. 3 Prediction results of the Mackey-glass time series with Gaussian white noise

4.2 球磨机排矿水控制应用 (Application to the control of the discharge water of ball mill)

为验证本文方法在磨矿过程的实际建模效果, 本节以磨矿过程球磨排矿水的建模过程为例进行验证. 球磨排矿水作为磨矿系统的重要环节, 主要用于控制磨矿过程的水循环. 球磨排矿水的水量直接影响磨矿系统内的总水量, 进而对整个系统的浓度产生影响. 系统内浓度过大, 会严重影响磨机的正常工作以及旋流器分级效果, 严重的可能导致磨机涨肚, 而浓度过低, 会降低磨机的利用率, 影响整个系统的生产效率. 经过工艺论证, 选定泵池液位、给矿量、磨矿总耗水量、溢流流量作为该规则库的输入, 球磨排矿水作为输出.

根据本文中所提出的方法构建球磨排矿水的规则库. 训练数据选取2013年12月上旬一段连续的9000组数据样本, 每条数据为一分钟内采样点的均值. 为了验证本文方法的泛化性和精度, 进行了多组验证实验. 每种工况随机选取训练样本之外的数据作为验证实验样本, 共选取20组, 每组样本包含100条数据. 对训练后的模型, 用该20组测试样本分别进行测试, 取测试结果的均值作为总的测试结果. 本实验同样采用Anfis方法, 基于特征选择的T–S模型建模方法, 以及数据驱动的方法进行对比实验. 对于本文方法中模糊子集融合阶段各个输入变量的聚类个数的选择, 主要通过经验分析估计各点的实际生产状态, 此处将泵池液位、给矿量、磨矿总耗水量、溢流流量的聚类数分别设置为3, 5, 5, 6. 实验结果如表2所示. 图4中列出了部分验证实验的效果图.

由表中结果可知, 3种对比方法中, 数据驱动的建模方法在规则数较少的情况下建模精度优于本文方法. 这是由于规则和隶属度函数较少的模型对于该建模过程尚未达到过拟合, 而通过本文方法得到的同一规则量级下的模型, 由于隶属度函数的数量减少, 可能会出现一定程度的知识缺失, 导致预测精度的下降. 这种规则数较少的规则库由于本身精度的限制不能作为最终模型. 本文方法随着模糊规则数量的提高, 训练精度与预测精度都有所提高, 且明显优于其他方法.同时, 前两种对比方法都一定程度呈现出了过拟合现象, 即训练精度与预测精度成反比, 而本文所提的方法则有效地避免了过拟合. 此外, 本文方法所建立的模糊模型, 在拥有相同数目的规则时, 模糊子集数量最少, 简化了规则库的结构. 图5和图6分别列出了应用本文方法建模, 模糊子集融合前后的隶属度函数分布情况, 表3列出了实验中本文方法3个模型对应的初始规则库以及约简后的规则和隶属度函数数量, 其中规则库a, b, c, 分别对应包含11条、14条、16条规则的模型, 可以明显看出简化的效果.

表 2 球磨排矿水模糊建模结果

Table 2 Fuzzy modeling results of the discharge water of ball mill

方法	训练误差		测试误差		规则数量	隶属度函数数量
	MAPE	RMSE	MAPE	RMSE		
数据驱动的T–S模型	2.74E–02	39.8467	2.89E–02	40.5157	11	44
	2.32E–02	37.6884	2.82E–02	41.7308	16	64
Anfis	2.41E–02	35.3239	3.27E–02	45.0919	11	44
	2.31E–02	33.9631	3.39E–02	48.2650	18	72
	2.16E–02	32.0821	3.52E–02	49.8670	22	88
基于特征选择的T–S模型	3.82E–02	56.7715	4.83E–02	62.4544	40	160
	3.39E–02	51.8504	4.54E–02	59.6709	80	320
本文方法	2.66E–02	40.6645	3.01E–02	43.0195	11	19
	2.63E–02	36.9737	2.91E–02	40.5672	14	19
	2.53E–02	36.0076	2.67E–02	37.2149	16	19

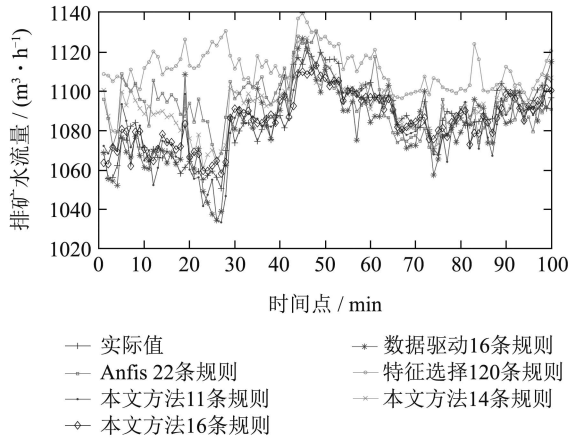
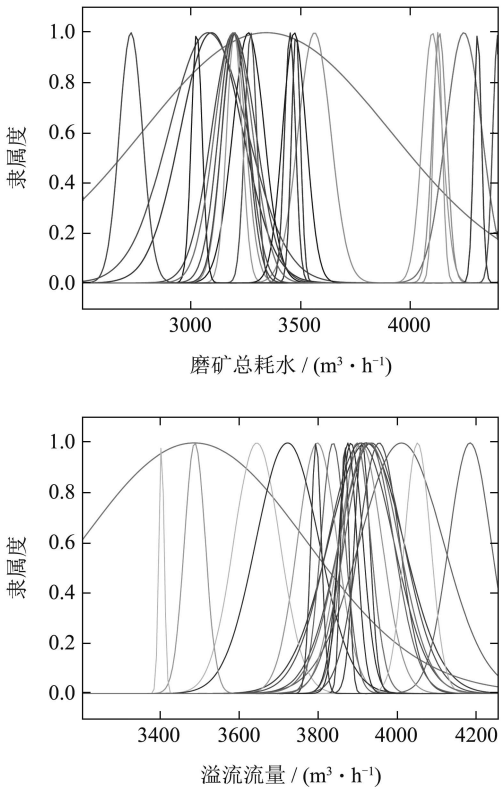
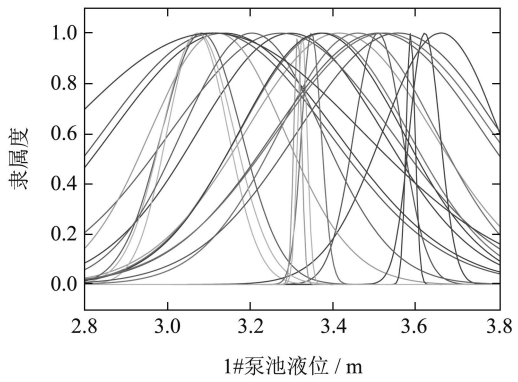


图 4 排矿水预测实验结果对比

Fig. 4 Prediction results of the discharge water of ball mill





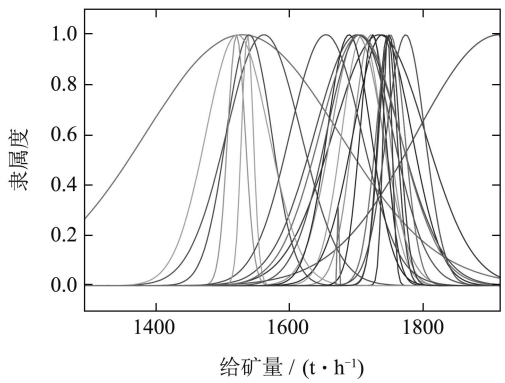
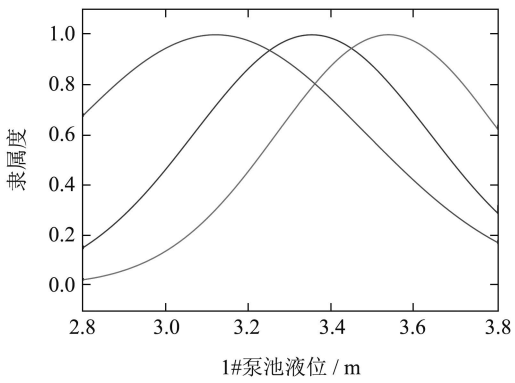
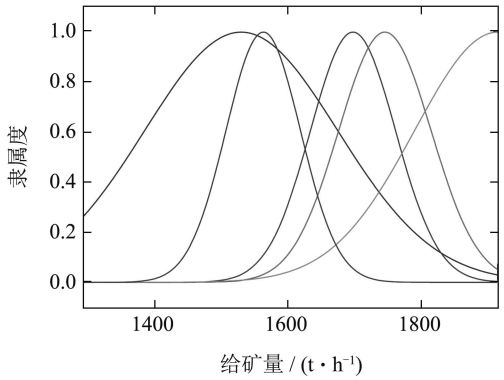
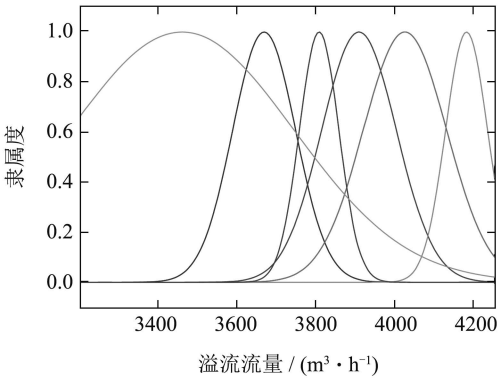


图 5 模糊子集融合前的隶属度函数分布

Fig. 5 Distribution of membership functions before fuzzy sets merging







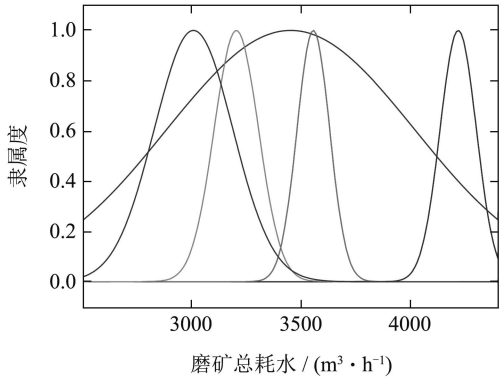


图 6 模糊子集融合后的隶属度函数分布

Fig. 6 Distribution of membership functions after fuzzy sets merging

由图5–6和表3可以看出, 经过融合处理后, 变量拥有更清晰的模糊子集划分, 去除了重叠模糊子集带来影响, 对于绝大部分的数据, 模糊子集融合后的规则库都拥有更高的精度, 提高了泛化性, 同时增强了可解释性.

表 3 本文方法规则库模糊子集融合及规则约简效果

Table 3 Results of fuzzy sets merging and rule simplification of the proposed method

规则库	规则库a		规则库b		规则库c	
	隶属度函数	规则数	隶属度函数	规则数	隶属度函数	规则数
初始规则库	60	15	84	21	96	24
最终规则库	19	11	19	14	19	16

5 结论(Conclusions)

本文提出了一种针对工业数据模糊建模的模糊集融合和规则约简方法. 在该方法中, 针对基于数据提取的模糊规则, 分析了重叠的模糊子集对规则库推理结果的影响, 并对这部分模糊子集进行聚类分析, 根据聚类中心对原隶属度函数进行融合, 得到新的隶属度函数, 最后, 对模糊子集融合后的规则库进行规则约简, 减少了模糊子集和规则的数量, 从整体上简化了规则库的结构, 更加容易通过规则库直观地观测出系统变量间的联系, 而且能使模型获得更好的泛化性. 最后, 应用本文方法对标准数据集和实际磨矿工业数据分别进行了建模分析, 并与传统建模方法进行了多组对比试验, 从实验的预测精度和最终规则库结构可以看出, 本文的方法能够有效地降低规则数量, 有效提高了基于数据的模糊模型的泛化性,改善了模型精度.

参考文献(References):

[1] JAMSA-JOUNELA S L. Current status and future trends in the automation of mineral and metal processing [J]. *Control Engineering Practice*, 2001, 9(9): 1021 – 1035.

[2] ZHU H P, ZHOU Z Y, YANG R Y, et al. Discrete particle simulation of particulate systems: a review of major applications and findings [J]. *Chemical Engineering Science*, 2008, 63(23): 5728 – 5770.

[3] BOUCHE C, BRANDT C, BROUSSAUD A, et al. Advanced control of gold ore grinding plants in South Africa [J]. *Minerals Engineering*, 2005, 18(8): 866 – 876.

[4] ZHOU P, CHAI T, SUN J. Intelligence-based supervisory control for optimal operation of a DCS-controlled grinding system [J]. *IEEE Transactions on Control Systems Technology*, 2013, 21(1): 162 – 175.

[5] CHANG X, LILLY J H. Evolutionary design of a fuzzy classifier from data [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2004, 34(4): 1894 – 1906.

[6] CORDON O. A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: designing interpretable genetic fuzzy systems [J]. *International Journal of Approximate Reasoning*, 2011, 52(6): 894 – 913.

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

[8] PAL N R, SAHA S. Simultaneous structure identification and fuzzy rule generation for Takagi-Sugeno models [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2008, 38(6): 1626 – 1638.

[9] JIN Y. Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement [J]. *IEEE Transactions on Fuzzy Systems*, 2000, 8(2): 212 – 221.

[10] GUO F, LIU B, SHI X, et al. T–S fuzzy model identification of MIMO nonlinear systems based on data-driven [C] *//2011 International Conference on Electronics, Communications and Control (ICECC)*. Ningbo: IEEE, 2011: 1186 – 1189.

[11] CETISLI B. Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 1 [J]. *Expert Systems with Applications*, 2010, 37(8): 6093 – 6101.

[12] CETISLI B. The effect of linguistic hedges on feature selection: Part 2 [J]. *Expert Systems with Applications*, 2010, 37(8): 6102 – 6108.

[13] CHEN Y C, PAL N R, CHUNG I F. An integrated mechanism for feature selection and fuzzy rule extraction for classification [J]. *IEEE Transactions on Fuzzy Systems*, 2012, 20(4): 683 – 698.

[14] MANSOORI E G, ZOLGHADRI M J, KATEBI S D. SGERD: a steady-state genetic algorithm for extracting fuzzy classification rules from data [J]. *IEEE Transactions on Fuzzy Systems*, 2008, 16(4): 1061 – 1071.

[15] GACTO M J, ALCALA R, HERRERA F. A multi-objective evolutionary algorithm for an effective tuning of fuzzy logic controllers in heating, ventilating and air conditioning systems [J]. *Applied Intelligence*, 2012, 36(2): 330 – 347.

[16] REZAEE B, ZARANDI M H. Data-driven fuzzy modeling for Takagi-Sugeno-Kang fuzzy system [J]. *Information Sciences*, 2010, 180(2): 241 – 255.

[17] ZHOU S M, GAN J Q. Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modeling [J]. *Fuzzy Sets and Systems*, 2008, 159(23): 3091 – 3131.

[18] MENCAR C, CASTELLANO G, FANELLI A M. Distinguishability quantification of fuzzy sets [J]. *Information Sciences*, 2007, 177(1): 130 – 149.

[19] ZARANDI M H F, NESHAT E, TURKSEN I B. A new cluster validity index for fuzzy clustering based on similarity measure [C] *//The 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*. Toronto: Springer Science and Business Media, 2007, 4482: 127 – 128.

[20] CROSS V, SUN Y. Semantic, fuzzy set and fuzzy measure similarity for the gene ontology [C] *//Fuzzy Systems Conference*. London: IEEE, 2007: 1 – 6.

[21] WU D, MENDEL J M. A vector similarity measure for linguistic approximation: Interval type–2 and type–1 fuzzy sets [J]. *Information Sciences*, 2008, 178(2): 381 – 402.

[22] NEFTI S, OUSSALAH M, KAYMAK U. A new fuzzy set merging technique using inclusion-based fuzzy clustering [J]. *IEEE Transactions on Fuzzy Systems*, 2008, 16(1): 145 – 161.

[23] BEZDEK J C, EHRLICH R, FULL W. FCM: The fuzzy c-means clustering algorithm [J]. *Computers and Geosciences*, 1984, 10(2): 191 – 203.

[24] DENAI M A, PALIS F, ZEGHBIB A. ANFIS based modeling and control of non-linear systems: a tutorial [C] *//IEEE International Conference on Systems, Man and Cybernetics*. Hague: IEEE, 2004: 3433 – 3438.

作者简介:

乔　峥　(1988−), 男, 硕士研究生, 主要研究方向为流程工业建模与优化, E-mail: qiaozhjo@gmail.com;

刘　颖　(1982−), 女, 讲师, 主要研究方向为一体化生产计划与调度、智能优化算法及应用, E-mail: liuying8227@163.com;

赵　珺　(1981−), 男, 副教授, 博士生导师, 主要研究方向为生产计划与调度、现代集成制造系统, E-mail: zhaoj@dlut.edu.cn;

王　伟　(1955−), 男, 教授, 博士生导师, 主要研究方向为自适应控制、现代集成制造系统和流程工业过程控制, E-mail: wangwei@dlut.edu.cn;

郭　戈　(1972−), 男, 教授, 博士生导师, 主要研究方向为工业过程建模与控制, E-mail: geguo@dlut.edu.cn.