

## 一种基于D-S和ARIMA的多模型软测量方法

王振雷<sup>1</sup>, 唐 苦<sup>1</sup>, 王 昕<sup>2</sup>

(1. 华东理工大学 化工过程先进控制和优化技术教育部重点实验室, 上海 200237; 2. 上海交通大学 电工与电子技术中心, 上海 200240)

**摘 要:** 针对传统软测量方法存在的预测性能差、融合能力低等缺点, 提出一种基于证据理论(D-S)合成规则和差分自回归滑动平均(ARIMA)模型的多模型软测量方法. 首先利用自适应模糊核聚类方法和最小二乘支持向量机建立多个子模型; 然后利用D-S合成规则构造的概率分配函数作为权值因子, 对子模型输出进行融合以得到多模型的输出; 最后结合ARIMA模型对静态多模型输出进行动态校正. 仿真研究与工业应用的结果表明, 所提出的方法具有良好的预测性能和融合能力.

**关键词:** 证据理论合成规则; 差分自回归滑动平均模型; 数据融合; 多模型; 软测量

**中图分类号:** TP273

**文献标志码:** A

### A multi-model soft sensing method based on D-S and ARIMA model

WANG Zhen-lei<sup>1</sup>, TANG Ku<sup>1</sup>, WANG Xin<sup>2</sup>

(1. Key Laboratory of Advanced Control and Optimization for Chemical Processes, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China; 2. Center of Electrical & Electronic Technology, Shanghai Jiaotong University, Shanghai 200240, China. Correspondent: WANG Zhen-lei, E-mail: wangzhen.l@ecust.edu.cn)

**Abstract:** There are some disadvantages in the traditional model algorithm for the soft sensor, such as low predictive accuracy and poor fusion ability. Therefore, a multi-model soft sensor algorithm is proposed based on the D-S rule and difference autoregressive moving average(ARIMA) model. Firstly, the adaptive fuzzy kernel clustering method(AFKCM) and least squares support vector machine(LS-SVM) are used to establish multiple sub-models. Then the output of the soft sensor is obtained through the fusion of the sub-models based on the weight factor calculated by D-S rules. The ARIMA model is used to realize the dynamic correction to the static multi-model output. Simulation results and industry application indicate that, comparing with the traditional soft sensor, the proposed method has better predictive performance and fusion ability.

**Key words:** Dempster-Shafer rule; difference autoregressive moving average; data fuse; multi-model; soft sensor

## 0 引 言

在实际工业生产过程中, 被控对象往往具有强非线性、快时变、多工况等特点, 而生产工艺的复杂多变, 以及现场操作环境复杂等因素都会导致产品质量的波动<sup>[1]</sup>. 在这种情况下, 单一模型的软测量建模方法效果并不理想, 存在过程特性匹配不佳以及预测精度差等问题. 多模型的软测量方法是解决复杂工业系统中关键变量测量问题的有力工具, 能够快速、准确地实现对不可测主导变量的估计, 并且能够有效提高

模型的预测精度和泛化性能.

在多模型软测量建模方法的设计过程中, 文献[2]给出一种多模型软测量方法, 将样本集拆分为多个样本子簇, 并以样本子簇重构线性流形面, 将属于同一流形面且相距较近的子簇进行融合, 再用最小二乘支持向量机(LS-SVM)进行建模. 文献[3-4]利用智能优化算法对LS-SVM模型的超参数进行寻优. 文献[5]提出了一种基于仿射传播聚类 and LS-SVM的多模型建模方法, 先利用仿射传播聚类方法对样本进行聚

收稿日期: 2013-05-17; 修回日期: 2013-08-13.

基金项目: 国家自然科学基金重点基金项目(U1162202); 国家自然科学基金项目(61222303, 61203157); 十二五国家科技支撑计划项目(2012BAF05B00); 上海市重点学科建设项目(B504); 中央高校基本科研业务费专项资金项目; 流程工业综合自动化国家重点实验室开放课题基金项目.

作者简介: 王振雷(1975-), 男, 教授, 博士生导师, 从事智能控制、复杂系统、故障诊断、智能优化算法等研究; 唐苦(1989-), 女, 硕士生, 从事软测量建模的研究.

类,再通过切换策略选择相应的子模型预测变量得到最终输出.然而,上述多模型方法都是针对样本集的聚类和LS-SVM模型参数的优选,忽略了多模型建模中的一个关键问题——如何将多模型的输出结果按照可靠方式融合在一起.基于切换策略的输出融合方式仅适用于局部线性系统,且易产生切换震荡现象.而基于加权策略的输出融合方式<sup>[6-7]</sup>尽管适合非线性系统,但实际应用效果并不理想.因为系统工况的复杂多变、测量对象的强非线性特征、样本划分不精确及聚类结果不佳等因素都会导致样本对各聚类中心的类隶属度准确度大大降低,利用这样的类隶属度作为加权因子,降低了权值的准确度,而且在此基础上融合得到的多模型输出的精度也会降低,造成模型融合能力低、预测效果不理想.

另一方面,工业过程一般都是动态的,利用静态建模方法建立的软测量模型无法反映工业过程的动态信息和全局特性,造成模型的适应性差,无法长期使用.文献[8]提出利用自回归滑动平均模型(ARMA)来改善系统的动态响应特性,但是该方法只针对单一模型进行动态校正,当操作条件改变时,对系统的真实输出逼近效果不佳.文献[9]将小波分析和相空间重构技术引入多模型建模中,结合ARMA模型对动态误差信息进行建模,但是重构参数的选取复杂且麻烦,参数太小会因相空间被压缩而存在冗余,参数太大则会丢失系统的原始信息,这些都将对多模型的预测结果产生较大的波动.

针对上述问题,本文提出一种基于证据理论(D-S)合成规则和差分自回归滑动平均(ARIMA)模型的多模型软测量方法.首先利用D-S合成规则处理不确定信息的聚焦优势,针对各子模型得到多个证据概率分配函数,考虑到证据源的可靠性和冲突可能性,设计折扣因子并对其进行修正,将修正后的概率分配函数作为权值因子对子模型的输出进行融合,改善了权值的准确度和可靠性,提高了多模型的预测能力和精度;然后结合ARIMA模型良好的逼近能力,对静态条件下得到的多模型输出进行动态校正,提高了模型的动态响应特性;最后,利用本文方法对非线性系统进行软测量,取得了很好的拟合效果,与传统的软测量方法相比,预测性能和精度均有较大改善.实际应用结果表明了该方法的有效性.

## 1 自适应模糊核聚类方法(AFKCM)

多模型建模步骤一般包括子数据集的划分、子模型的建立、多模型的输出3个方面.本节利用自适应模糊核聚类方法(AFKCM)和最小二乘支持向量机(LS-SVM)对样本数据集进行聚类并建立多个子模型.

对于给定的数据样本集  $\mathbf{X} = \{\mathbf{x}_i, i = 1, 2, \dots, n\}$ ,采用自适应模糊核聚类方法,选择聚类目标函数为

$$\begin{cases} J_{\theta}(\mathbf{U}, c) = \sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^m \| [1 - K(\mathbf{x}_i, \mathbf{v}_j)] \|^2, \\ \text{s.t. } \mathbf{U} \in M_{fc}. \end{cases} \quad (1)$$

其中:  $\mu_{ij}$  为第  $i$  个样本  $\mathbf{x}_i$  对应于第  $j$  个聚类中心  $\mathbf{v}_j$  的类隶属度,  $\forall j = 1, 2, \dots, c, \mu_{ij} \in [0, 1]; 0 < \sum_{i=1}^n \mu_{ij} < n, \sum_{j=1}^c \mu_{ij} = 1; m$  为模糊控制指数;  $K(x, y)$  为高斯核函数.

利用拉格朗日乘子法求得式(1)的最小值并迭代求出新的聚类中心  $\mathbf{v}$  和隶属度矩阵  $\mathbf{U}$ , 即

$$\begin{aligned} \mu_{ij}^m &= \frac{(1 - K(\mathbf{x}_i, \mathbf{v}_j))^{-1/(m-1)}}{\sum_{j=1}^c (1 - K(\mathbf{x}_i, \mathbf{v}_j))^{-1/(m-1)}}, \\ \mathbf{v}_j &= \frac{\sum_{i=1}^n \mu_{ij}^m K(\mathbf{x}_i, \mathbf{v}_j) \mathbf{x}_i}{\sum_{i=1}^n \mu_{ij}^m K(\mathbf{x}_i, \mathbf{v}_j)}. \end{aligned} \quad (2)$$

本文采用文献[10]定义的指标,对聚类结果进行有效性评价.考虑到不同样本间的复杂非线性特征,使用高斯核函数对其核化,核化后的有效性指标不仅突出了样本的特征差异,而且还可以自动给出最佳聚类数目.核化后的有效性指标可表示为

$$V_{GK}(c) = \frac{\sum_{j=1}^c \sum_{k=1}^n \mu_{ij}^m (1 - K(\mathbf{v}_k, \mathbf{x}_i)) + \frac{1}{c} \sum_{j=1}^c (1 - K(\mathbf{v}_j, \tilde{\mathbf{v}}))}{\min_{j \neq k} (1 - K(\mathbf{v}_k, \mathbf{v}_j))}. \quad (4)$$

聚类结束后,针对上面得到的  $c$  个子样本数据集  $\text{subset}_1, \dots, \text{subset}_c$ ,采用最小二乘支持向量机方法分别建立  $c$  个子模型  $\text{LS-SVM}_1, \dots, \text{LS-SVM}_c$ .

## 2 基于证据理论(D-S)合成规则的多模型输出融合

多模型的输出需要对各个子模型的输出进行加权融合,因此,各子模型权值因子的合理选择至关重要.针对上节建立的  $c$  个子模型  $\text{LS-SVM}_1, \dots, \text{LS-SVM}_c$ ,为了得到较为准确的子模型权值因子,本节提出了一种基于D-S合成规则的多模型输出融合方法.

### 2.1 基于D-S合成规则的概率分配函数

为了利用D-S合成规则,将聚类所得的所有  $c$  个子模型构成证据理论中的辨识框架  $\theta$ ,则可将任一子模型视为焦点  $C_j (j = 1, 2, \dots, c)$ .对于样本  $\mathbf{x}_1$ ,

首先根据式(2)求出其对应于第1个子模型(也即第1个焦点 $C_1$ )的模糊类隶属度,并根据证据理论<sup>[11-13]</sup>,将其作为一条证据,记该证据的概率分配函数为 $m(C_1|\mathbf{x}_1) = \mu_{11}$ .而对于所有 $n$ 个样本数据 $\mathbf{X} = \{\mathbf{x}_i, i = 1, 2, \dots, n\}$ ,同理可得到 $n$ 条证据,其概率分配函数记为 $m(C_1|\mathbf{x}_i) = \mu_{i1} (i = 1, 2, \dots, n)$ .然后,使用D-S合成规则对这些概率分配函数进行融合,将融合后的概率分配函数作为第1个子模型的概率分配函数,即

$$\begin{cases} m(C_1|\mathbf{X}) = \frac{\sum_{C_1|\mathbf{x}_1 \cap \dots \cap C_1|\mathbf{x}_n = C_1|\mathbf{X}} m_1(C_1|\mathbf{x}_1) \cdots m_n(C_1|\mathbf{x}_n)}{1 - k}, \\ m(\emptyset) = 0. \end{cases} \quad (5)$$

其中矛盾因子

$$k = \sum_{C_1|\mathbf{x}_1 \cap \dots \cap C_1|\mathbf{x}_n = \emptyset} m_1(C_1|\mathbf{x}_1) \cdots m_n(C_1|\mathbf{x}_n),$$

$k$ 的大小反映了证据间的冲突程度.依此类推,对于所有的 $c$ 个子模型,按照上述方法,可以得到 $c$ 个证据概率分配函数,分别记为 $m(C_1|\mathbf{X}), \dots, m(C_c|\mathbf{X})$ ,将其作为各子模型的权值因子,从而获得多模型的输出.

## 2.2 折扣因子修正概率分配函数

当证据信息源出现不可靠情况或者证据之间冲突时,证据合成规则在合成过程中出现的不稳定性会给融合效果带来一定的影响.为了避免这种情况出现,专家学者们基于不同角度提出了各种改进方法来修正原始证据.如文献[14]基于证据距离,文献[15]基于矛盾因子,文献[16]引入证据权来修正原始证据.本节综合各类改进方法,在确保之前得到的概率分配函数不会产生大的波动前提下设计折扣因子,对得到的概率分配函数进行小范围的平稳修正,以改进权值因子的准确度.具体修正步骤如下.

1) 对于分为 $c$ 类的样本集,每类的样本数是已知的,令 $\omega_j$ 为第 $j$ 个概率分配函数的权重,则有

$$\omega_j = n_j/n. \quad (6)$$

其中: $n_j$ 为第 $j$ 类的样本数, $n$ 为总样本数.

2) 计算所有概率分配函数的权重向量

$$W = (\omega_1, \omega_2, \dots, \omega_c). \quad (7)$$

3) 设最大权重 $\omega_{\max} = \max(\omega_1, \omega_2, \dots, \omega_c)$ ,则其相对权重向量为

$$W' = (\omega_1, \omega_2, \dots, \omega_c)/\omega_{\max}. \quad (8)$$

4) 根据式(6)~(8),可以得到证据概率分配函数的“折扣因子” $\alpha$ 为

$$\alpha_j = \omega_j/\omega_{\max}. \quad (9)$$

5) 修正后的概率分配函数值可表示为

$$m^*(C_j|\mathbf{X}) = \alpha_j m(C_j|\mathbf{X}). \quad (10)$$

## 2.3 多模型输出融合

将式(10)得到的 $c$ 个证据概率分配函数作为子模型的权值因子,对子模型的输出进行融合,得到的多模型输出为

$$\hat{y} = m^*(C_1|\mathbf{X})\hat{y}_1 + \dots + m^*(C_c|\mathbf{X})\hat{y}_c. \quad (11)$$

利用折扣因子对原来的概率分配函数进行小幅度的修正,修正后的概率分配函数综合考虑了各样本的类隶属度信息和每类样本数构成的权重信息,通过准确融合这些信息获得各子模型的概率分配函数,将其作为权值因子对子模型进行有效融合,得到了多模型输出.相比于传统的直接简单地将模糊隶属度作为权值因子融合子模型输出,本文方法充分利用了证据理论在处理不确定信息时的聚焦优势和决策能力,构造出的权值因子用于加权输出,不但可以降低复杂工况、样本划分不精确或者聚类结果不佳等情况对后续多模型输出精度的影响,而且有效提高了模型的整体预测能力.

## 3 差分自回归滑动平均(ARIMA)模型

利用上述方法得到多模型的输出后,采用ARIMA模型对其进行动态校正,以改善静态多模型的动态响应特性.ARIMA模型描述系统当前时刻 $t$ 的响应 $\hat{y}_t$ ,即 $\hat{y}_t$ 不仅在时间上与其以前的观测值有关,而且还与系统扰动的现值和滞后值存在一定的依存关系.ARIMA( $p, d, q$ )模型可表示为

$$\begin{aligned} \hat{y}_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t - \\ \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}. \end{aligned} \quad (12)$$

其中:AR为自回归, $p$ 为自回归项;MA为移动平均, $q$ 为移动平均项数; $d$ 为将非平稳时间序列转成平稳序列所作的差分次数.

引入线性推移算子 $B$ ,则有 $B\hat{y}_t = y_{t-1}, B^2\hat{y}_t = y_{t-2} \dots$ ,故式(12)可变换为

$$\phi(B)\hat{y}_t = \theta(B)\hat{y}_t. \quad (13)$$

其中: $\varepsilon_t$ 为满足 $N(0, \sigma^2)$ 的白噪声序列, $\phi(B)$ 和 $\theta(B)$ 为推移算子 $B$ 的 $m$ 阶和 $n$ 阶多项式,即

$$\begin{cases} \phi(B) = (1 - \phi_1 B - \dots - \phi_m B^m), \\ \theta(B) = (1 - \theta_1 B - \dots - \theta_n B^n). \end{cases} \quad (14)$$

当 $\hat{y}_t$ 为非平稳的时间序列时,采用差分方法对其进行平稳化处理. $\nabla$ 表示差分,用下述步骤进行差分变换:

$$\nabla \hat{y}_t = \hat{y}_t - \hat{y}_{t-1}, \nabla^2 \hat{y}_t = \nabla \hat{y}_t - \nabla \hat{y}_{t-1}. \quad (15)$$

经过 $d$ 次差分处理后, $\hat{y}_t$ 便转化为平稳、正态、零均值的平稳序列 $\hat{y}'_t$ ,且 $\hat{y}'_t = \nabla^d \hat{y}_t$ .然后根据希尔伯

特空间上线性算子的基本理论, 可用一个 ARIMA( $n, n - 1$ ) 模型以任意精度逼近该时间序列, 采用递增法逐步逼近,  $n$  以 2 为增量, 直到找到合适的  $n$  值.

#### 4 基于 D-S 和 ARIMA 的多模型软测量建模步骤

基于 D-S 和 ARIMA 的多模型软测量建模系统结构如图 1 所示, 算法的实现步骤如下.

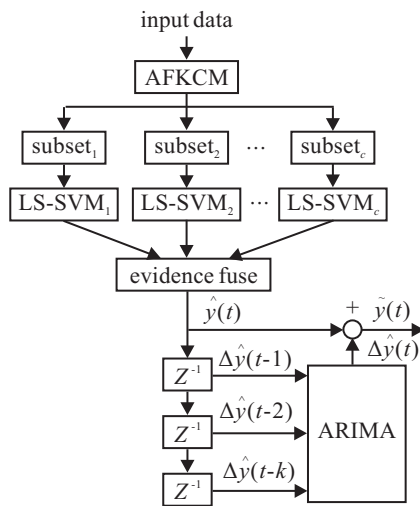


图 1 基于 D-S 和 ARIMA 的多模型软测量方法结构框图

1) 数据预处理. 选择  $m \times n$  维的输入输出变量  $\mathbf{X}$ ,  $m$  为样本维数,  $n$  为样本个数, 利用“ $3\sigma$ ”法剔除异常数据, 并对数据进行归一化处理.

2) 自适应聚类分析. 采用 AFKCM 对训练样本集进行聚类, 得到每个样本的模糊类隶属度和各聚类中心, 并自动确定出最佳聚类数目  $c$ .

3) 建立子模型. 对于各子训练样本集, 采用 LS-SVM 对其进行训练学习并确定模型参数. 选择高斯核函数作为 LS-SVM 的核函数, 通过交叉验证法确定各子模型的参数  $C$  和  $\sigma$ , 并得出样本对应于各子模型的子输出  $\hat{y}_1, \dots, \hat{y}_c$ .

4) 基于 D-S 规则的子模型输出融合. 按照式 (5) ~ (10) 的方法得到各子模型的证据概率分配函数值, 并将其作为子模型的权值因子, 然后利用式 (11) 对各子模型的输出进行证据融合, 得到多模型的输出  $\hat{y}$ .

5) 模型输出的动态化. 利用上面的静态模型得到样本的多模型输出  $\hat{y}$  之后, 使用 ARIMA 模型对当前时刻  $t$  的多模型输出  $\hat{y}$ , 即对  $\hat{y}_t$  进行动态调整. 首先将  $\hat{y}_t$  与真实测量值  $y$  相减, 得到一个关于输出值误差  $\Delta y$  的时间序列; 然后判断该时间序列是否是平稳序列, 若不是平稳序列, 则对  $\Delta y$  进行差分处理, 将其转换为平稳序列; 接着利用基本的 ARMA( $n, n - 1$ ) 模型对该时间序列进行建模, 即可得到关于预测误差的 ARIMA 模型, 进而求出  $\Delta \hat{y}_t$ ; 最后, 将以上两模型相结合进行模型预测, 得到样本的最终输出为  $\tilde{y}_t =$

$\hat{y}_t - \Delta \hat{y}_t$ .

#### 5 仿真研究

为了验证本文算法, 利用文献 [17] 中的复杂非线性系统进行仿真研究, 该系统可用如下非线性方程表示:

$$y(t) = \frac{x_1(t)}{1 + 0.5 \sin(x_2(t))} + \frac{x_2(t)}{1 + 0.5 \sin(x_1(t))} + \varepsilon(t),$$

$$\begin{cases} x_1(t+1) = \left( \frac{x_1(t)}{1 + x_1^2(t)} + 1 \right) + \sin(x_2(t)), \\ x_2(t+1) = \\ x_2(t) \cos(x_2(t)) + \exp\left(-\frac{x_1^2(t) + x_2^2(t)}{8}\right) \times \\ x_1(t) + \frac{u^3(t)}{1 + u^2(t) + 0.5 \cos(x_1(t) + x_2(t))}. \end{cases} \quad (16)$$

其中:  $x_1(t)$  和  $x_2(t)$  为系统的状态,  $u(t)$ 、 $y(t)$  和  $\varepsilon(t)$  分别为系统的输入、输出和白噪声.

假定系统状态不可测, 通过已知的输入输出信息来预测系统的输出  $y(t)$ . 选择软测量建模的输入为

$$\phi(t-1) = [y(t-1), y(t-2), y(t-3), u(t-1), u(t-2), u(t-3)]^T.$$

以  $u(t) \in [-2.5, 2.5]$  的随机信号与  $\varepsilon(t) \in N(0, 0.1)$  的白噪声作用于系统, 得到 5 000 组时间序列训练样本. 以测试信号

$$u(t) = \sin(0.5\pi t) + \sin(0.08\pi t) \quad (17)$$

作用于该系统, 产生了 200 组测试样本, 分别用来检验 BP、LS-SVM、AFKCM-LS-SVM、AP-LS-SVM 和 D-S-ARIMA 这 5 种方法的性能.

仿真结果如图 2 所示. BP 方法是一种局部搜索方法, 在求解复杂非线性函数的全局极值时易陷入局部最优, 导致训练失败, 因而它的局部拟合效果好但整体拟合效果差. LS-SVM 方法由于模型过于单一而导致整体精度不高. AP-LS-SVM 在对样本进行 AP 聚类后建立若干个 LS-SVM 子模型, 使用切换策略得到多模型输出, 对于某类样本, 用其所对应的子模型预测样本输出作为样本的最终输出. 但是对于该非线性系统, 该方法在最高点和最低点附近切换时, 出现了震荡现象, 过渡不平缓, 所以预测效果不理想. 基于 AFKCM-LS-SVM 的方法, 对样本进行自适应模糊核聚类后建立若干个子模型, 直接将类隶属度作为权值因子对子模型输出进行融合, 总体效果不错, 但是对于最低点附近的样本, 由于类隶属度准确度不够, 导致权值因子准确度不高, 影响了多模型输出的精度, 使得对这部分样本的拟合效果不理想.

相比于上述几种方法, 本文方法兼顾了局部拟合效果和在最高点、最低点附近的拟合效果, 使用加权融合, 避免了出现过渡不平缓; 在类隶属度的基础上, 将修正后的证据概率分配函数作为权值因

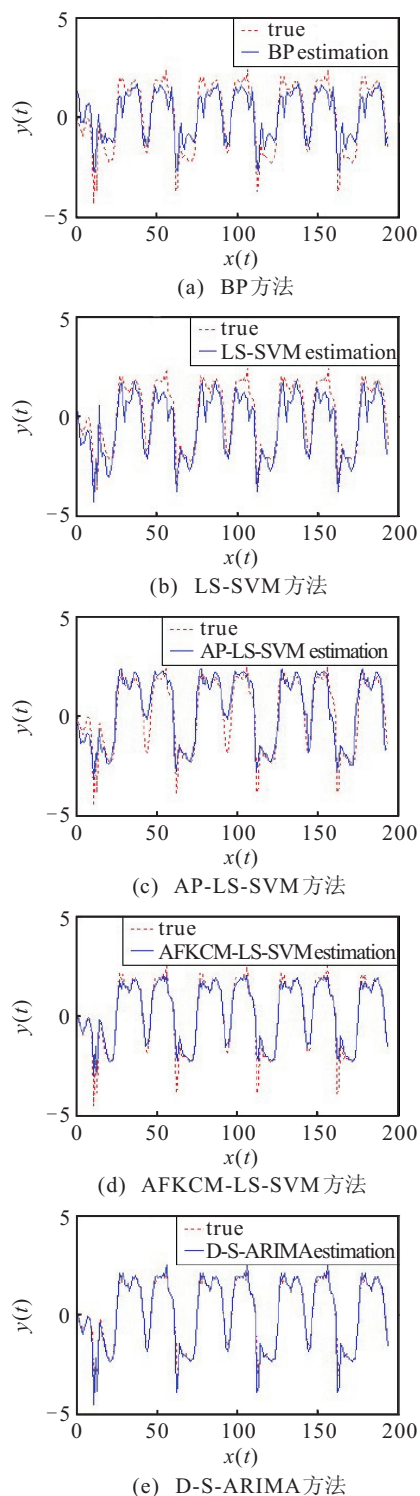


图2 5种不同方法对该系统的实际输出与估计输出

子,改善了权值的准确度;对多模型输出加以动态校正,改善了全局动态特性,提高了预测精度,拟合结果如图2(e)所示.对比图2(c)和图2(d)可以看出,本文方法能较好地逼近系统的实际输出,预测效果优于不加证据融合和动态校正的其他两种多模型软测量方法.采用相对均方误差(RMSE)和最大相对误差(MAXE)来评估模型的预测能力,这几种方法的性能比较结果如表1所示,从中也可看出本文方法的效果优于其他几种方法.

表1 不同方法的性能比较

Method	RMSE / $10^{-1}$	MASE / $10^{-1}$
BP	0.3948	1.3839
LS-SVM	0.3462	1.3659
AP-LS-SVM	0.3664	1.3481
AFKCM-LS-SVM	0.3628	1.3462
D-S-ARIMA	0.3162	1.2868

## 6 工业应用

酯化反应作为整个聚酯生产工艺的关键环节,对稳定聚酯生产起决定性作用.而其第1酯化釜出口的主要质量指标酯化率的高低直接影响后续反应的进行和聚酯产品的结晶性能,因此,常常通过控制酯化率来控制整个生产过程.图3为酯化反应过程示意图.将对苯二甲酸(PTA)和乙二醇(EG)配比成一定浓度的浆料后,送入第1酯化反应器进行酯化反应,反应生成的乙二醇送回酯化釜再次参与反应,水蒸汽排出系统,酯化釜主产物BHET进入预缩聚反应器.

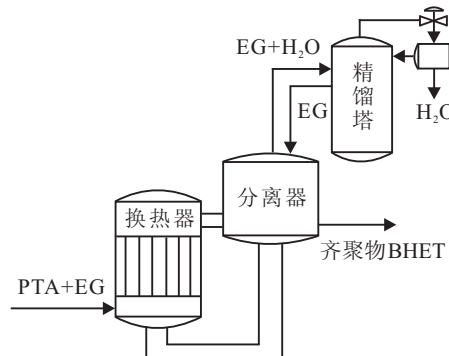


图3 酯化反应过程示意图

酯化反应的主要目标是达到缩聚工艺要求的酯化率,而不同的缩聚工艺对酯化率有不同的要求,所以必须通过调整反应压力和原料量比等操作条件来达到所需的酯化率.但是操作条件的突然改变会引起酯化率的质量波动,不利于整个生产过程的实时控制.另一方面,酯化过程大多采用两个酯化反应器来达到工艺要求的酯化率,而反应系统的高度非线性、时变性和不确定性加大了酯化率在线测量的难度.因此,采用本文提出的基于D-S和ARIMA的多模型软测量建模方法对酯化率进行软测量.

根据对酯化反应过程的工艺机理分析,选择停留时间( $r$ )、酯化反应温度( $T$ )、压力( $P$ )和EG/PTA摩尔配比作为辅助变量,建立基于D-S和ARIMA的多模型软测量模型,用来对酯化率进行预测.

以某化工厂聚酯生产过程为例,具体的建模步骤如下.

1) 数据采集和处理.对现场采集到的数据进行处理,得到1000组标准数据.将其中的900组数据作为训练数据集,用于模型的建立;剩下的100组作为测试数据集,用于检验模型的预测能力.

2) 利用 AFKCM 对训练数据进行聚类. 设定最大迭代次数为  $t_{\max} = 1000$ , 允许误差为  $\delta = 0.0001$ , 模糊控制指数为  $m = 2$ . 当聚类数目  $c = 4$  时, 聚类有效性指标取极小值 89.1069, 并得到对应的聚类中心  $\mathbf{v}$  和隶属度矩阵  $U$ .

3) LS-SVM 子模型建立. 对于得到的 4 个子样本集, 利用 LS-SVM 方法建立 4 个子模型并训练学习, 经交叉验证法确定 LS-SVM 的参数, 如表 2 所示.

表 2 LS-SVM 模型的参数

Model	model 1	model 2	model 3	model 4
$C$	150.0006	150.00	150.0	150.001
$\sigma$	10.5091	11.006	10.008	9.8384

4) D-S 融合子模型输出. 按照式 (5)~(10) 计算得到各个子模型的权值因子, 然后利用式 (11) 对各子模型的输出进行证据融合, 得到多模型的输出  $\hat{y}_t$ .

5) ARIMA 动态建模. 将静态多模型的输出  $\hat{y}$  与人工分析值  $y$  相减, 得到输出值误差  $\Delta y$  的时间序列值, 对该时间序列进行 ARIMA 建模, 检验得到当差分次数  $d = 1$ , 最佳阶数  $n = 5$  时, ARIMA 模型的逼近性能最好. 将得到的  $\Delta \hat{y}_t$  与多模型输出相减, 则样本的最终输出为  $\tilde{y}_t = \hat{y}_t - \Delta \hat{y}_t$ .

6) 模型预测输出. 对于测试数据集, 采用上面的方法预测其输出.

图 4 为 4 种不同软测量方法的预测结果, 表 3 比较了这些方法的性能参数.

表 3 不同方法的性能比较

Method	RMSE / $10^{-1}$	MASE / $10^{-1}$
LS-SVM	0.4101	1.4235
AFKCM-LS-SVM	0.4092	1.4167
AP-LS-SVM	0.3656	1.2167
D-S-ARIMA	0.3426	1.1983

从上述结果可以看出, 基于 D-S 和 ARIMA 的多模型方法相比于单模型及传统的多模型方法, 预测性能有了较大改善, 具有更好的回归精度. 这是因为酯化反应具有较高的非线性和复杂工况的特点, 而单模型建模时需要考虑全部训练样本, 这限制了模型的精度. 传统的多模型方法在建模时尽管对训练数据进行了聚类划分, 分别建立了不同的子模型, 但是并未考虑到聚类划分的准确度、权值因子的选择方法和过程的动态变化对多模型输出结果的精度影响, 因此预测性能没有显著改善. 本文方法不但充分考虑了权值对多模型输出结果的影响, 利用折扣因子改进的 D-S 合成规则对多模型输出进行有效融合, 而且考虑了实际工业过程的动态特性, 结合 ARIMA 模型对静态多模型的输出进行动态校正, 显著改善了系统的动态响应特性, 因而具有更好的适用性, 在对酯化率的软测量建模中取得了较好的拟合效果.

### 7 结 论

本文针对实际工业生产过程中关键变量的软测量, 设计了一种基于 D-S 和 ARIMA 模型的多模型软测量建模方法. 该方法考虑了 D-S 合成规则的聚焦特性, 针对自适应模糊核聚类得到的各子模型成功构造了多条证据, 并设计折扣因子对其概率分配函数进行修正, 将修正后的概率分配函数作为权值因子对多模型的输出进行了有效融合, 改善了权值的准确度和可靠性; 然后利用 ARIMA 模型对多模型的输出进行动态校正, 实现了对静态多模型的动态校正, 改善了系统的动态响应特性, 提高了模型的精度. 通过对非线性系统的仿真验证和酯化率的软测量建模应用, 表明了本文提出的方法具有较好的预测性能和适用性, 是一种有效的多模型建模方法.

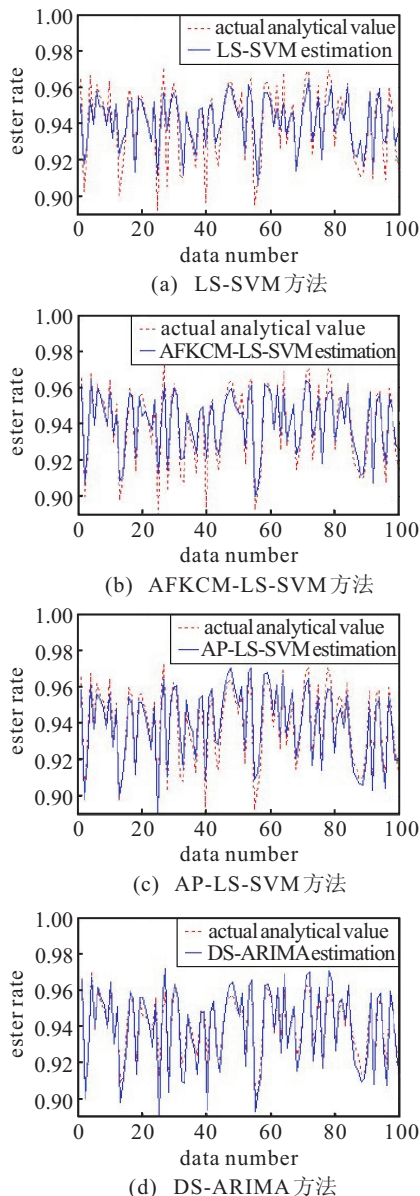


图 4 4 种不同的软测量方法对酯化率测试样本的预测值与人工值对比结果

## 参考文献(References)

- [1] Lin B, Recke B, Knudsen J K H, et al. A systematic approach for soft sensor development[J]. *Computers and Chemical Engineering*, 2007, 31(5/6): 419-425.
- [2] 陈定三, 杨慧中. 基于局部重构融合流形聚类的多模型软测量建模[J]. *化工学报*, 2011, 62(8): 2281-2285. (Chen D S, Yang H Z. Multiple model soft sensor based on local reconstruction and fusion manifold clustering[J]. *J of Chemical Industry and Engineering*, 2011, 62(8): 2281-2285.)
- [3] Liao R, Zheng H, Grzybowski S, et al. Particle swarm optimization-least squares support vector regression based forecasting model on dissolved gases in oil-filled power transformers[J]. *Electric Power Systems Research*, 2011, 81(12): 2074-2080.
- [4] Rubio G, Pomares H, Rojas I, et al. A heuristic method for parameter selection in LS-SVM: Application to time series prediction[J]. *Int J of Forecasting*, 2011, 27(3): 725-739.
- [5] 李丽娟, 宋坤, 赵英凯. 基于仿射传播聚类的ARA发酵过程建模[J]. *化工学报*, 2011, 62(8): 2116-2122. (Li L J, Song K, Zhao Y K. Modeling of ARA fermentation based on affinity propagation clustering[J]. *J of Chemical Industry and Engineering*, 2011, 62(8): 2116-2122.)
- [6] 李卫, 杨煜普, 王娜. 基于核模糊聚类的多模型LSSVM回归建模[J]. *控制与决策*, 2008, 23(5): 560-564. (Li W, Yang Y P, Wang N. Multi-model LSSVM regression modeling based on kernel fuzzy clustering[J]. *Control and Decision*, 2008, 23(5): 560-564.)
- [7] Li X, Su H, Chu J. Multiple models soft-sensing technique based on online clustering arithmetic[J]. *J of Chemical Industry and Engineering*, 2007, 58(11): 2834-2839.
- [8] 杜文莉, 官振强, 钱锋. 一种基于时序误差补偿的动态软测量建模方法[J]. *化工学报*, 2010, 61(2): 339-443. (Du W L, Guan Z Q, Qian F. Dynamic soft sensor modeling based on timeseries error compensation[J]. *J of Chemical Industry and Engineering*, 2010, 61(2): 339-443.)
- [9] 唐志杰, 唐朝晖, 朱红求. 一种基于多模型融合软测量建模方法[J]. *化工学报*, 2011, 62(8): 2248-2252. (Tang Z J, Tang Z H, Zhu H Q. A multi-model fusion soft sensor modeling method[J]. *J of Chemical Industry and Engineering*, 2011, 62(8): 2248-2252.)
- [10] 唐苦, 王昕, 王振雷, 等. 一种基于高斯核化有效性指标的自适应优选聚类数的模糊核聚类方法[J]. *计算机与应用化学*, 2012, 29(10): 1199-1203. (Tang K, Wang X, Wang Z L, et al. An adaptive optimal clustering number of FKCM base on Gaussian kernel validity index[J]. *Computers and Applied Chemistry*, 2012, 29(10): 1199-1203.)
- [11] Zhang S G, Li W H, Ding K. A novel approach to evidence combination based on the evidence credibility[J]. *Control Theory & Applications*, 2009, 26(7): 812-814.
- [12] Li X, Dezert J, Smarandache F, et al. Evidence supporting measure of similarity for reducing the complexity in information fusion[J]. *Information Sciences*, 2011, 181(10): 1818-1835.
- [13] Su Z, Wang P, Yu X, et al. Maximal confidence intervals of the interval-valued belief structure and applications[J]. *Information Sciences*, 2011, 181(9): 1700-1721.
- [14] 刘准钊, 程咏梅, 潘泉, 等. 基于证据距离和矛盾因子的加权证据合成法[J]. *控制理论与应用*, 2009, 26(12): 1439-1442. (Liu Z G, Cheng Y M, Pan Q, et al. Combination of weighted belief functions based on evidence distance and conflicting belief[J]. *Control Theory & Applications*, 2009, 26(12): 1439-1442.)
- [15] Su Z, Wang P, Shen J, et al. Multi-model strategy based evidential soft sensor model for predicting evaluation of variables with uncertainty[J]. *Applied Soft Computing*, 2011, 11(2): 2595-2610.
- [16] 叶清, 吴晓平, 宋业新. 基于权值系数与冲突概率重新分配的证据合成方法[J]. *系统工程与电子技术*, 2006, 28(7): 1014-1016. (Ye Q, Wu X P, Song Y X. Evidence combination method based on the weight coefficients and the confliction probability distribution[J]. *Systems Engineering and Electronics*, 2006, 28(7): 1014-1016.)
- [17] Zeng J, Wang J, Guo J. Local multi-model method based on similarity of vector[J]. *Application Research of Computers*, 2012, 29(5): 1631-1634.

(责任编辑: 李君玲)