

有偏最小最大概率模型及在汽油属性预测中的应用

贺凯迅^{1,2†}, 刘晶晶¹, 王小邦¹, 苏照阳¹

(1. 山东科技大学 电气与自动化工程学院, 山东 青岛 266590;

2. 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)

摘要: 汽油属性的在线预测多采用无偏估计方法建立的近红外定量分析模型实现, 累积预测误差的正负偏差范围难以控制, 这会严重影响汽油调合优化控制的投运效果. 针对这一问题, 本文提出了一种采用有偏估计实现油品属性在线预测的方法. 首先从最小最大概率学习机出发, 提出了有偏最小最大概率回归模型. 然后利用即时学习方法设计了有偏回归模型的局部建模与更新策略, 用以提高回归模型的自适应能力. 最后在国内某炼厂汽油调合过程中采集的工业数据上进行实验, 结果表明该方法与传统方法相比具有明显优势, 有利于大幅度提高调合优化控制的投运率.

关键词: 汽油调合; 最小最大概率机; 动态建模; 机器学习; 过程系统

引用格式: 贺凯迅, 刘晶晶, 王小邦, 等. 有偏最小最大概率模型及在汽油属性预测中的应用. 控制理论与应用, 2020, 37(8): 1799 – 1807

DOI: 10.7641/CTA.2020.90913

Biased minimax probability model and its application in prediction of gasoline properties

HE Kai-xun^{1,2†}, LIU Jing-jing¹, WANG Xiao-bang¹, SU Zhao-yang¹

(1. College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao Shandong 266590, China;

2. Key Laboratory of Advanced Control and Optimization for Chemical Processes, East China University of Science and Technology, Shanghai 200237, China)

Abstract: The online prediction of gasoline properties is mostly realized by the near-infrared quantitative analysis model which established by the unbiased estimation method. However, the range of the positive and negative deviations of the cumulative prediction error is difficult to control, which will seriously affect the operation of gasoline blending optimization control. To deal with this issue, a biased estimation method is proposed for the online prediction of gasoline properties. Firstly, a biased minimax probability regression model is proposed based on minimax probability machine. Then, based on just-in-time learning approach, a local modeling and updating strategy is developed for the biased regression model to improve its adaptive ability. Finally, experiments are carried out with the gasoline data collected from a domestic oil refinery. The results show that the present method has obvious advantages compared with traditional algorithms, and it is beneficial to improve the operation rate of the optimal control system of gasoline blending.

Key words: gasoline blending; minimax probability machine; dynamic modeling; machine learning; process systems

Citation: HE Kaixun, LIU Jingjing, WANG Xiaobang, et al. Biased minimax probability model and its application in prediction of gasoline properties. *Control Theory & Applications*, 37(8): 1799 – 1807

收稿日期: 2019-11-03; 录用日期: 2020-04-08.

†通信作者. E-mail: kaixunhe@sdust.edu.cn.

本文责任编辑: 王伟.

国家自然科学基金项目(61803234, 61873149, 61751307), 山东省自然科学基金项目(ZR2017BF026), 中国博士后科学基金项目(2018M632691), 山东省泰山学者项目研究基金项目资助.

Supported by the National Natural Science Foundation of China (61803234, 61873149, 61751307), the Shandong Provincial Natural Science Foundation of China (ZR2017BF026), the China Postdoctoral Science Foundation (2018M632691) and the Research Fund for the Taishan Scholar Project of Shandong Province of China.

1 引言

汽油调合是汽油出厂前的最后一道工序,油品的质量主要取决于该阶段.为了保证调合成功,并最大程度的减少油品属性的质量过剩,需要对调合过程进行优化控制^[1].调合优化控制的实施需要对汽油属性进行实时检测,而研究法辛烷值(research octane number, RON)、马达法辛烷值(motor octane number, MON)等关键属性的在线检测显得尤为重要.目前辛烷值的检测主要依赖美国材料实验协会(American Society of Testing Materials, ASTM)标准的汽油辛烷值试验机;该方法只能用于离线检测,检测费用高、检测周期长,难以为调合优化控制提供及时的属性值反馈^[2].

自1989年Jeffery等人^[3]首次将近红外技术引入到汽油属性的分析以来,在国内外炼厂中汽油辛烷值的在线预测普遍依赖近红外分析仪.该仪表是二次测量仪器,需要预先建立汽油属性与其近红外光谱(near infrared spectrum, NIR)之间的软测量模型(近红外模型)^[4].为了解决这一问题,很多基于统计学习的方法被提出和应用.其中在工业上应用最为广泛的方法有多元线性回归(multiple linear regression, MLR)、主成分回归(principal component regression, PCR)和偏最小二乘回归(partial least squares regression, PLSR)^[5-6].此外,机器学习领域的一些新方法也不断被应用于近红外模型的建立上:如支持向量机回归(support vector machine regression, SVR)、神经网络(artificial neural networks, ANN)、随机森林(random forest, RF)等^[7-9].

近红外模型的输出特性决定了调合优化控制的效果:若调合头后汽油属性的预测值偏高,优化控制器会根据出厂指标调整调合配方、拉低该属性值,这容易引起优化失败并造成产品不合格;反之则抬高相应的属性值,而属性值升高太多,又会造成质量过剩^[10-11].一个调合批次若调合失败,则需要将该批次汽油重新调合直至合格,重调不仅会造成巨大的油品挥发损失还会延误订单交付、占用储罐资源^[12].因此调合优化控制首先要保证调合成功,然后再优化配方、减少质量过剩.要保证调合优化控制的顺利进行,需要近红外模型给出精确的无偏估计.现有的建模方法虽然在离线应用验证中可以取得较为满意的结果,但是它们都难以在较长时间周期内提供精确无偏估计,在一个批次内累积误差的正负也难以把握,模型的维护工作也耗时耗力^[13-15].解决上述问题的一种有效的方法是建立有偏估计模型,使模型输出的预测值略小于实际值并维持在可以接受的精度范围内;这样一来,优化控制器便可将成品油的属性向上调节并保证在可接受的范围内,从而最大限度的保证一次调合成功并减少质量过剩.

查阅文献可知,基于有偏估计的辛烷值在线检测方法在理论和应用上的研究尚未引起足够的重视,也缺乏有效的解决方案.为解决这一问题,本文提出一种基于即时学习(just in time learning, JITL)和有偏最小最大概率学习机(biased minimax probability machine, BMPM)的有偏估计方法.该方法将有偏估计的理念引入到汽油属性的在线预测中;把最小最大概率学习机拓展到回归领域,使所建立的近红外模型能够较好地平衡预测精度和预测范围.此外, JITL策略的引入有效提高了模型对实时工况的跟踪能力,有利于减少模型维护的工作量.为了验证算法的性能,本文以国内某炼厂的实际工业数据以及公开的基准数据为例,与现有方法进行对比研究,结果验证了所提方法的有效性及其在调合优化控制中的重要现实意义.

2 有偏最小最大概率回归

2.1 回归模型概述

假设 x 服从任一有界分布 Λ , x 和 y 满足 $y = f(x) + \varepsilon$, 且 $E(\varepsilon) = 0$, $\text{var}(\varepsilon) = \sigma^2$, 则函数 $f(\cdot)$ 为 x 和 y 之间的回归模型^[15].给定同分布的数据集 $\{X, Y\} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 其中: $x_i = (x_{i,1} \ x_{i,2} \ \dots \ x_{i,m}) \in \mathbb{R}^{1 \times m}$, $Y = (y_1 \ y_2 \ \dots \ y_n) \in \mathbb{R}^{n \times 1}$, $i = 1, 2, \dots, n$. 则可通过 $\{X, Y\}$ 计算出函数 $f(\cdot)$ 的估计 $\hat{f}(\cdot)$, 进而对于任一 $x \in \Lambda$, 都可以得到因变量的估计值 $\hat{y} = \hat{f}(x)$. 假设估计值的绝对误差为 $\varepsilon = |\hat{y} - y|$, 考虑到实际问题对估计精度的要求, 对任一观测样本 x 的估计误差, 希望其满足

$$\Omega_f = \inf_{(x,y)} P\{|\hat{y} - y| \leq \varepsilon\}, \quad (1)$$

即要求对于所有观测样本, 估计值的绝对误差小于等于 ε 的概率达到最大.从这一点出发, 可将回归问题归结为如下二分类问题^[16-17]:

$$\max_{\alpha, a \neq 0, b} \alpha, \quad (2)$$

$$\text{s.t.} \quad \inf_{u \sim (\bar{u}, \sum u)} P\{a^T u - b \geq 0\} \geq \alpha, \quad (3)$$

$$\inf_{v \sim (\bar{v}, \sum v)} P\{a^T v - b \leq 0\} \geq \alpha, \quad (4)$$

其中: α 为分类正确概率的下确界, $u_i = (y_i + \varepsilon \ x_{i,1} \ x_{i,2} \ \dots \ x_{i,j} \ \dots \ x_{i,m})$, $U = (u_1 \ u_2 \ \dots \ u_n)^T$, $v_i = (y_i - \varepsilon \ x_{i,1} \ x_{i,2} \ \dots \ x_{i,j} \ \dots \ x_{i,m})$, $V = (v_1 \ v_2 \ \dots \ v_n)^T$.

对于线性回归问题, 对上述二分类问题进行求解, 解出分类超平面 $a^T z = b$ ($z = (\hat{y} \ x_1 \ x_2 \ \dots \ x_m)$), 即可按下式计算出回归系数:

$$b = a_1 \hat{y} + a_2 x_1 + \dots + a_{m+1} x_m \Rightarrow \\ \hat{y} = -\frac{a_2}{a_1} x_1 - \frac{a_3}{a_1} x_2 - \dots - \frac{a_{m+1}}{a_1} x_m + \frac{b}{a_1}$$

$$\beta^T x + \beta_0, \quad (5)$$

采用任一种二分类算法均可对该回归问题求解。

2.2 有偏最小最大概率分类机

式(2)–(4)中定义的二分类问题, 类别 U 和 V 的分类正确概率下确界都是 α , 属于无偏分类。为了引出本文所提的方法, 本节将介绍有偏二分类问题以及BMPM^[18–19]。

定义如下有偏二分类问题:

$$\max_{\alpha, a \neq 0, b} \alpha, \quad (6)$$

$$\text{s.t.} \quad \inf_{u \sim (\bar{u}, \sum u)} P\{a^T u - b \geq 0\} \geq \alpha, \quad (7)$$

$$\inf_{v \sim (\bar{v}, \sum v)} P\{a^T v - b \leq 0\} \geq \eta_0, \quad (8)$$

其中 $0 \leq \eta_0 < 1$ 为常数, 代表次要类别的可接受分类精度, 一般由先验知识确定。据文献[20], 式(6)–(8)可以归结为一个带约束的优化问题:

$$\max_{\alpha, a \neq 0, b} \alpha, \quad (9)$$

$$\text{s.t.} \quad 1 \geq \kappa(\alpha) \sqrt{a^T \Sigma_U a} + \kappa(\eta_0) \sqrt{a^T \Sigma_V a}, \quad (10)$$

$$a^T (\bar{u} - \bar{v}) = 1, \quad (11)$$

其中: $\kappa(\alpha) = \sqrt{\frac{\alpha}{1-\alpha}}$, $\kappa(\eta_0) = \sqrt{\frac{\eta_0}{1-\eta_0}}$ 。假设 Σ_U 为正定矩阵, 由式(10)可推出

$$\kappa(\alpha) \leq \frac{1 - \kappa(\beta_0) \sqrt{a^T \Sigma_V a}}{\sqrt{a^T \Sigma_U a}}, \quad (12)$$

又因 $\kappa(\alpha)$ 是 α 的增函数, 故式(9)–(11)定义的优化问题等价于

$$\max_{a \neq 0} \frac{1 - \kappa(\eta_0) \sqrt{a^T \Sigma_V a}}{\sqrt{a^T \Sigma_U a}}, \quad (13)$$

$$\text{s.t.} \quad a^T (\bar{u} - \bar{v}) = 1, \quad (14)$$

这是一个分数规划问题, 可采用梯度映射解出使 α 最大的参数 a , b 。

3 融合即时学习的有偏最小最大概率回归

在汽油管道调合过程中, 调合后的油品首先泵入成品油罐, 静置一段时间再采样检测, 待检验合格后方可出厂^[21]。若检验不合格, 该罐成品油需作为一种特殊的组分油进行重新调合, 直至油品属性满足出厂标准^[22]。重调过程不仅过多占用储罐资源、延误订单交付, 而且重调循环还会带来油品挥发损失; 因此, 在实际调合过程中要尽可能的保证一次调合成功。现有的调合优化控制方案以油品属性的在线检测为基础: 优化控制器根据调合头后成品油属性的实时检测值, 参调组分油的属性, 罐底油属性等计算成品油罐内属性的累积平均值并确定下一个优化周期的配方, 从而达到优化成品油罐内油品属性的目的。从这一角度来看, 如果调合头后油品属性的检测值偏高, 则调

合配方经优化调整后会使下个周期的属性降低; 如果检测值持续偏高, 调合配方向下的调整幅度会变大, 这极有可能造成该批次油品属性不合格。因此, 希望调合头后的油品属性检测值偏低实际值, 并保持在允许的误差范围内; 这样配方经优化调整后不但能使罐内油品属性达标、保证一次调合成功, 而且能够尽可能减少质量过剩。为了解决这一问题, 本文采用有偏最小最大概率机回归(biased minimax probability machine regression, BMPMR)算法建立模型, 即用有偏最小最大概率分类机(式(6)–(8))代替式(2)–(4)中定义的二分类回归问题, 并用梯度下降法求解分类超平面, 再根据式(5)求出回归方程系数。

考虑到有偏最小最大概率回归算法是一种全局建模方法: 即训练样本结构、模型参数均为离线设定, 在线应用过程中保持不变。对于汽油调合过程, 组分油种类的改变, 上游物料属性的波动, 配方的改变, 订单切换以及生产条件的变化等都会使过程样本偏离初始训练数据覆盖的范围, 从而造成模型衰减, 预测偏差变大。因全局模型的更新只能离线进行, 在实际工业现场往往需要配备专门人员负责模型的维护和更新, 需要耗费大量的人力成本。为了解决这一问题, 本文将JITL策略^[23–24]融入有偏最小最大概率回归算法中, 提出自适应有偏最小最大概率机回归(adjust BMPMR, ad-BMPMR)算法以扩展BMPM, 实现模型的在线自适应更新。

3.1 即时学习建模方法

JITL是一种惰性学习策略, 它不预先建立模型, 而是当每个待测样本到来时再建立局部模型, 并且给出相应的预测输出^[25–27]。通常来说, 对于 t 时刻的待测样本 x_t , 用JITL策略建立局部模型需要以下3个步骤: 1) 计算 x_t 与历史数据 $\{X, y\}$ 中所有 $x_i \in X$, $i = 1, 2, \dots, n$ 的相似度, 根据预先设定的相似度阈值选出 k 个样本组成局部训练集; 2) 基于该训练集, 用预先选定的建模算法建立预测模型 $f(x)$; 3) 计算预测值 $\hat{y}_t = f(x_t)$ 并释放局部模型。上述步骤中, 相似度的定义决定了每个局部训练集的规模以及训练数据对当前工况的代表性, 一般采用基于欧式距离和马氏距离的方法。本文中, 采用下式定义相似度^[28–30]:

$$\omega_i = \exp\left(-\frac{d_i^2}{\sigma_d^2}\right), \quad (15)$$

$$d_i = \sqrt{(x_t - x_i)^T (x_t - x_i)}, \quad (16)$$

式中: σ_d 是控制参数, 用于控制相似度 ω_i 随距离 d_i 的衰减率; σ_d 越大, ω_i 衰减的越快, 反之亦然。图1给出了 $\omega = \exp\left(-\frac{d^2}{\sigma_d^2}\right)$ 在不同参数 σ_d 下的变化曲线。根据式(15), 给定阈值 $\omega_{\lim}$ 后, 相似度满足 $\omega \geq \omega_{\lim}$ 的样本将被选中组成局部训练集。

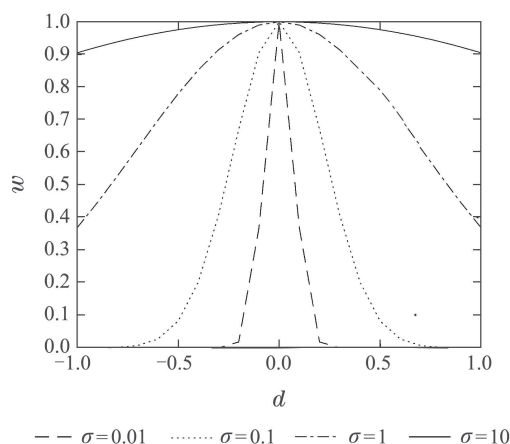


图1 相似度函数特性曲线

Fig. 1 Characteristic curves of similarity function

3.2 融合JITL的有偏最小最大概率回归

在汽油调合优化控制过程中,希望汽油属性的预测值尽量偏低,即 $-\varepsilon \leq \hat{y} - y \leq 0$.因此希望 $P\{a^T u - b \geq 0\}$ 的概率下确界 α 达到最大,而 $P\{a^T v - b \geq 0\}$ 的概率下确界 η_0 由先验知识给定.此外,为了增强局部模型对当前工况的代表性,递归地将最新获取的样本 $\{x_{\text{new}}, y_{\text{new}}\}$ 更新入历史数据中.对于每个待测样本,当用JITL方法确定局部训练集后,即可运用BMPM算法建立模型,用梯度优化求解分类超平面,并按式(5)解出回归系数,进而给出对当前样本 x_t 的预测输出. ad-BMPMR算法的具体流程如图2所示.

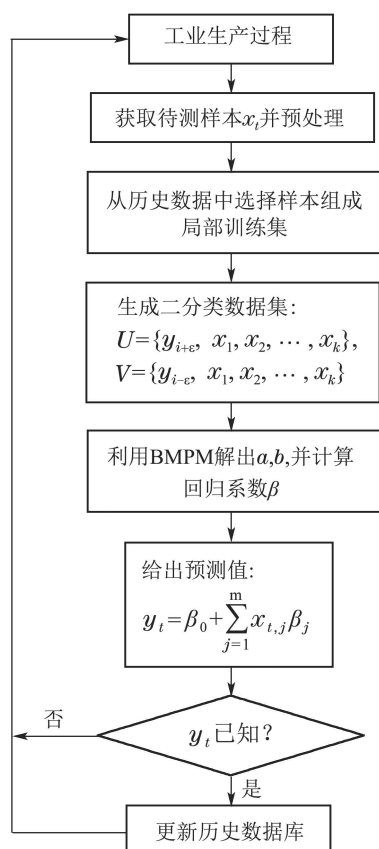


图2 ad-BMPMR算法流程图

Fig. 2 Model structure of ad-BMPMR

4 实验验证与工业应用

在汽油管道调合过程中,调合后成品汽油的RON,芳烃含量(aromatic, ARO),苯含量等属性的检测主要依赖在线近红外光谱分析仪.该仪表是二次仪器,不能直接检测属性,而是需要把在线采集的汽油NIR光谱输入到预先建立的回归模型中,从模型输出相应属性的预测值.本文采用所提出的自适应有偏最小最大概率回归(ad-BMPMR)算法建立模型,在柴油芳烃-近红外光谱(aromatic-NIR, ARO-NIR)基准数据上对算法进行了比较分析,验证算法性能,之后基于国内某炼厂汽油管道调合过程中采集的汽油数据集对所提方法进行应用验证.

4.1 工业数据

1) 柴油ARO-NIR光谱数据:该数据为一组公开的基准数据,数据来源于<http://www.eigenvector.com/data/index.htm>.本文选用该数据包中柴油ARO-NIR光谱数据集进行实验.将该数据集按2:1的比例随机划分为训练集和测试集,其中训练集包含182个样本,测试集包含85个样本.该数据集每条柴油NIR光谱包含401个波长点,采用一阶微分法对光谱进行预处理,以消除基线漂移和背景噪声,处理后的光谱变量作为自变量,柴油芳烃含量作为因变量.

2) 汽油RON-NIR光谱数据:该数据采集于国内某炼化企业半成品车间汽油调合工业过程,NIR光谱采用Guidewave公司的扫描光栅型在线近红外光谱仪获取,波长范围为1100 nm~1300 nm,波长间隔1 nm,共包含201个波长点,所有光谱数据都采用Unscrambler商用建模软件中自带的背景扣除法进行预处理,以消除基线漂移.处理后的汽油NIR光谱作为自变量,RON含量作为因变量.所用数据集中共包含324组样本,按2:1的比例随机选择216个样本组成训练集,其余108个样本组成测试集.

4.2 实验方法与性能参数

为了比较分析本文所提出方法的性能和特点,该部分用不同算法建立模型,并对比分析不同算法在同一数据上的均方根误差(root mean square error, RMSE)和决定系数(R^2);RMSE越小, R^2 越大,则模型的精度越高、对数据的解释能力越强. RMSE, R^2 的具体形式如下:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}, \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (18)$$

此外, 考虑到对油品属性在线预测的目的是对当前批次的油品属性进行优化, 以保证该批次油品的总体质量合格、并减少质量过剩. 而且国内炼油企业的成品汽油、柴油等都是经调合后泵入储罐, 然后对储罐内的油品进行抽样检查, 合格后准予出厂; 因此在线过程中单个样本的预测偏差难以表征储罐内由累积效应产生的最终成品油的属性偏差. 为了客观反映成品油罐内油品的属性偏差情况, 并定量评价分析模型在一个连续批次内的精度, 本文定义了平均累积偏差, 具体如下:

$$\overline{\text{AcErr}} = \frac{\sum_{i=1}^n (\hat{y}_i - y_i) \times f v_i}{\sum_{i=1}^n f v_i}, \quad (19)$$

式中 $f v_i$ 为 $t = 0$ 至 $t = 1$ 时刻的总管流量, 其余以此类推. 当采样均匀、流量稳定时, 式(19)等价于

$$\overline{\text{AcErr}} = \sum_{i=1}^n (\hat{y}_i - y_i) / n, \quad (20)$$

那么, 一个调合批次的累积偏差为

$$\text{AcErr} = \sum_{i=1}^n (\hat{y}_i - y_i). \quad (21)$$

由上式可知, 在一个调合批次内, 单个样本的正负偏差累积后会有一部分抵消, 抵消后余下的偏差为成品油罐内油品属性的真实预测偏差; 这一点符合油品调合实际工况.

本文所采用的对比算法, 具体如下:

1) 支持向量机回归 (support vector machine regression, SVR): SVR 采用台湾大学林智仁教授等开发的 libsvm 工具箱 (<http://www.csie.ntn.edu.tw/~cjlin/libsvm>) 实现, 以历史数据作为训练集, 并且保持训练集结构不变;

2) 最小最大概率机回归 (minimax probability machine regression, MPMR): 用 MPM 求解本文定义的二分类问题, 并给出回归系数和预测结果, 同样以历史数据作为训练集, 并且保持训练集结构不变;

3) 有偏最小最大概率机回归 (biased minimax probability machine regression, BMPMR): 用 BMPM 求解本文定义的二分类问题, 并给出回归系数和预测结果, 同样以历史数据作为训练集, 并且保持训练集结构不变;

4) 递归偏最小二乘 (recursive partial least squares, RPLS): 以历史数据作为初始训练集建立 PLS 模型, 在线过程中递归的利用新采样的标记样本更新 PLS 的协方差矩阵, 该算法的具体细节可参考文献 [31].

4.3 结果分析

4.3.1 柴油芳烃含量预测结果分析

考虑到柴油近红外光谱包含的变量较多, 而且波

长点变量之间的耦合严重, 本案例中采用主成分分析法对 NIR 光谱降维, 并按累积贡献率 $\geq 98\%$ 提取主成分, 共提取 3 个主成分; 降维后的数据分别送入 ad-BMPMR, BMPMR, MPMR 和 SVR 建立回归模型. RPLS 采用原始数据建模, 公平起见潜变量个数取 3, 遗忘因子采用试差法在训练集中确定, 该案例中整定值为 0.3. 其他参数设置如下: $\eta_0 = 0.9$, $\varepsilon = 3.65$, $\omega_{\text{lim}} = 0.95$ 均采用试差法确定; SVR 算法选用 RBF 核, 其余采用 libsvm 里的默认参数.

实验结果详见表 1, 结果保留 2 位小数. 从表 1 中可见, ad-BMPMR 在测试集上性能最佳. 从 RMSE 和 R^2 的角度来看, ad-BMPMR, BMPMR, RPLS 和 MPMR 差别不大, 4 种算法对数据的解释能力以及预测精度相当. 但是从累积误差来看, 本文方法的累积误差的绝对值较小而且是负偏差. 也就是说, 在一个批次内所有样本的正负预测误差抵消后的剩余误差较小, 且预测值比实际值偏低. 这一点非常有利于对整个批次内属性进行优化控制, 非常符合调合优化控制的需求. BMPMR 的累积误差也是负的, 但是偏差绝对值较大, 整体性能不如 ad-BMPMR, 这主要是由其静态模型的特性决定的. SVR 算法在该案例中的性能指标较差, 一是因为 SVR 也是静态模型, 难以及时跟踪过程数据的变化; 二是由于 SVR 算法内部的参数较多, 难以确定最优模型参数; 此外 SVR 算法难以控制有偏估计, 导致累积误差偏大. 作为常用的在线算法 RPLS, 其 RMSE 比本文方法高出很多, 没有获得满意的精度. 所有对比算法的时间损耗如表 1 所示, ad-BMPMR 在测试集上总耗时 1.06 s, 比其他算法略高, 但是平均到每个待测样本的平均计算时间小于 0.1 s, 完全能够满足在线预测的需求.

表 1 柴油数据不同建模算法的实验结果对比

Table 1 Comparison of different modeling strategies for diesel dataset

方法	R^2	RMSE	AcErr	耗时/s
ad-BMPMR	0.85	2.46	-9.04	1.06
BMPMR	0.82	2.66	-38.37	0.83
MPMR	0.83	2.60	9.42	0.41
SVR	0.79	2.86	10.30	0.06
RPLS	0.77	3.04	3.45	0.07

图 3 为 ad-BMPMR, BMPMR, MPMR 以及 SVR 4 种算法对柴油 ARO 预测值的散点图, 从中可见本文方法的预测值更为均匀的分布在误差限之内, 而其他 3 种方法的拟合曲线均有较大的非期望偏置. 这一点也验证了表 1 中的实验结果. 这 4 种对比算法的误差曲线见图 4, 从中可见 ad-BMPMR 的误差变化范围要小于其他对比算法, 同样验证了上述分析结果.

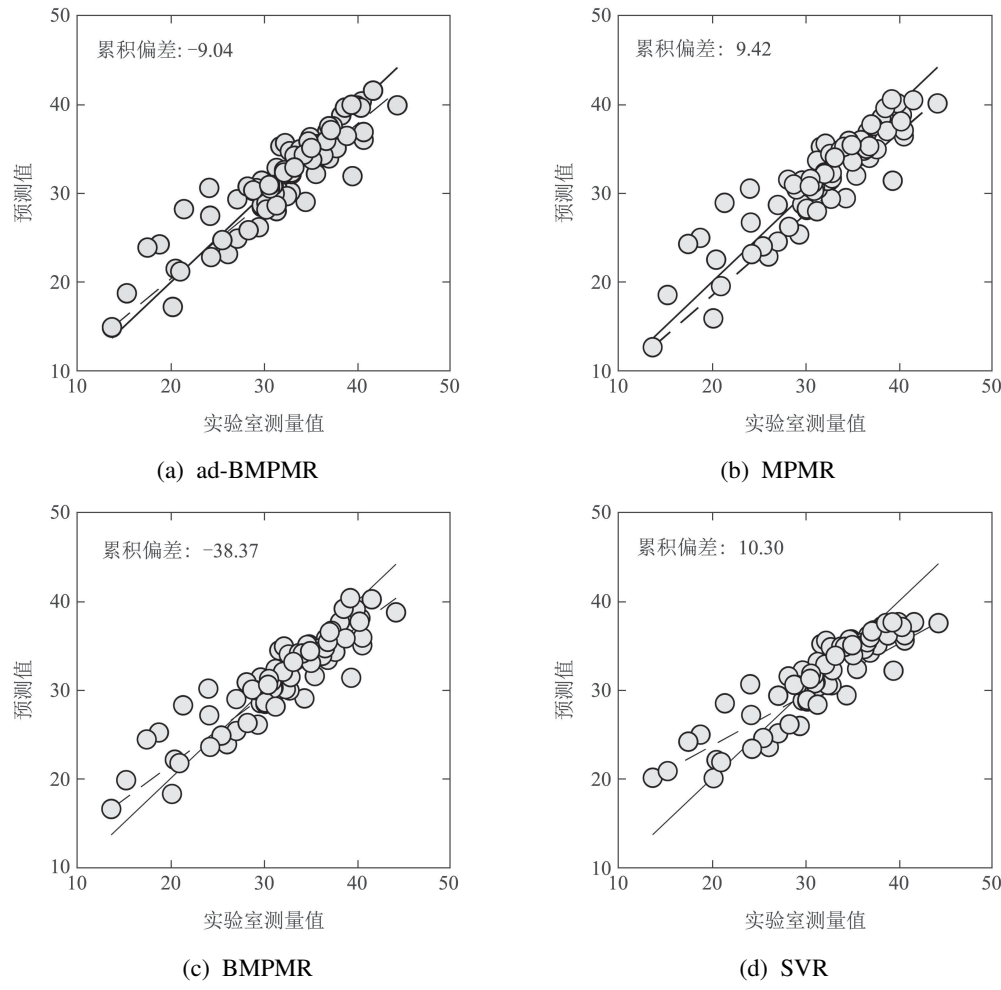


图 3 柴油ARO预测值散点图

Fig. 3 Scatter plots of model predictions: the ARO property of diesel

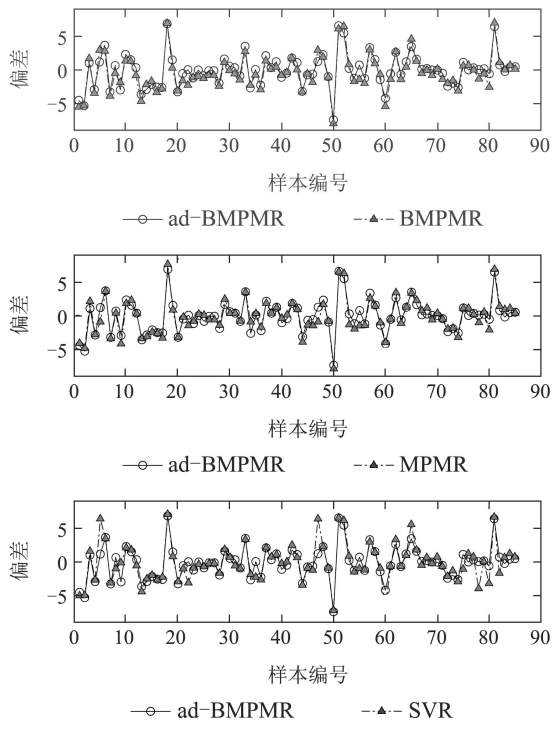


图 4 柴油ARO的预测偏差比较

Fig. 4 Bias curves for diesel ARO estimation

4.3.2 汽油RON含量预测分析

由于本案例收集的汽油NIR光谱的波长变量相对较少,为了不损失有用信息,这里采用全光谱(201个变量)作为输入变量,不对原始NIR光谱进行降维.参数设置如下: $\eta_0 = 0.93$, $\varepsilon = 0.5$, $\omega_{lim} = 0.98$ 均采用试差法确定. SVR算法选用RBF核,其余采用libsvm里的默认参数, RPLS的主元个数取6,遗忘因子取0.1,采用试差交叉验证法在训练集中整定,实验结果详见表2.

表 2 汽油数据不同建模算法实验结果对比
Table 2 Comparison of different modeling strategies for gasoline dataset

方法	R^2	RMSE	AcErr	耗时/s
ad-BMPMR	0.92	0.32	-5.41	1.92
BMPMR	0.90	0.36	-16.20	0.09
MPMR	0.76	0.56	2.22	0.28
SVR	0.84	0.47	5.64	0.11
RPLS	0.89	0.39	7.56	0.05

从 R^2 和 RMSE 的角度来看, ad-BMPMR 算法取得了最好的结果, BMPMR 的性能稍差于 ad-

BMPMR, 这两种算法的预测精度以及对数据的解释能力均能够满足汽油调合工况的需求. RPLS与BMPMR获得了相当的精度, 但是RPLS的累积误差绝对值较小, 这是由算法的动态特性决定的. 在实际调合应用中, 若使用RPLS仍然难以在建模过程中预判其预测输出与实际属性的偏差正负; 若调合体积较大, 即使累积正偏差较小也仍可能造成调合产品不合格. SVR的误差较大, 模型的解释能力也较低, 其主要原因同第4.3.1节中的分析: 主要由于SVR算法内部参数较多, 参数设置以及核函数的选择很难达到最优. 此外由于SVR的模型维护较为繁琐, 在实际工业现场很少使用. MPMR的性能表现最差, 已经超出0.5, 难以满足调合优化的需求精度. 从累积误差AcErr来看, MPMR算法拥有最低的累积误差, 但是因其RMSE太高, 在调合优化过程中会导致调合配方前后调整幅度较大; 大幅度的配方调整会导致调合不均匀, 进而会使成品油罐内的油品发生分层, 难以保证调合成功. 所以MPMR难以兼顾预测精度和累积误差. 相比之下, ad-BMPMR的累积误差为-5.41, 偏低于是实际值, 这一点与所期望的一致; 累积误差绝对值比BMPMR要小很多. 综合RMSE, R^2 和本文采用的累积误差指标, ad-BMPMR具有最佳的性能, 最适合用于汽油调合过程中油品属性的在线预测. 所有对比算法的时间损耗如表2所示, ad-BMPMR在测试集上总耗时1.92 s, 虽然比其他算法略高, 但是平均到每个待测样本的平均计算时间小于0.1 s, 完全能够满足在线预测的需求.

图5为其中4种算法在测试集上的拟合散点图. 从中可见, ad-BMPMR的预测值比其他几种算法更为均匀的分布在真实值稍偏下的位置, 这一点也符合前文的分析. MPMR和SVR算法中, 有较多预测点严重偏离理想拟合线, 这在调合优化过程中会造成局部短时间内调合配方的大幅度调整. BMPMR算法的拟合曲线也较为理想, 但是由于算法的静态特性造成整体上的累积误差较大.

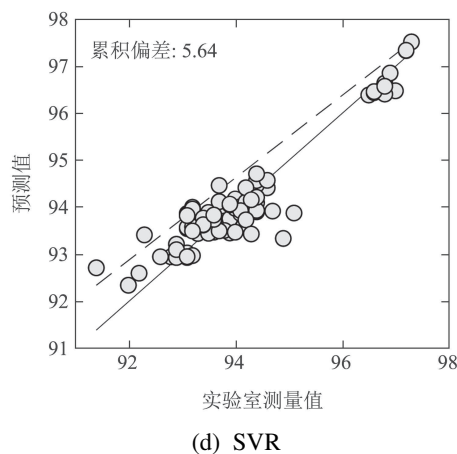
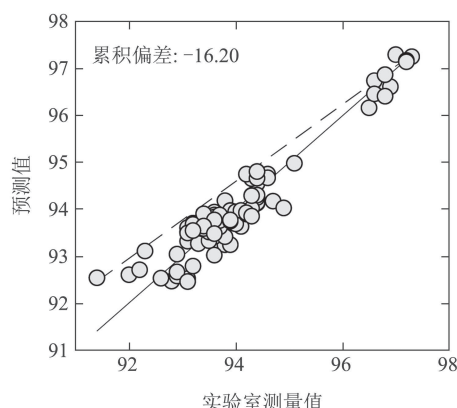
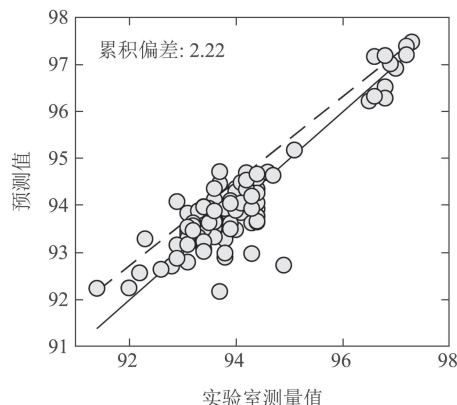
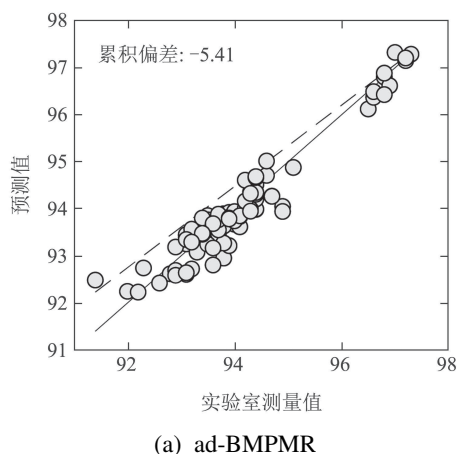


图5 汽油RON预测值散点图

Fig. 5 Scatter plots of predictions: the RON property of gasoline

图6为ad-BMPMR与BMPMR, MPMR以及SVR三种对比算法在测试集上的误差曲线, 从中可见, 本文方法正负误差的变化幅度明显小于MPMR和SVR. 与BMPMR相比, 误差曲线相当, 但是其正负误差的分布更为对称, 这也是ad-BMPMR拥有更小的累积偏差的原因. 综上, 从对数据的解释能力、预测精度以及累积偏差来看, ad-BMPMR非常适合用于油品调合优化控制过程中油品属性的在线预测.

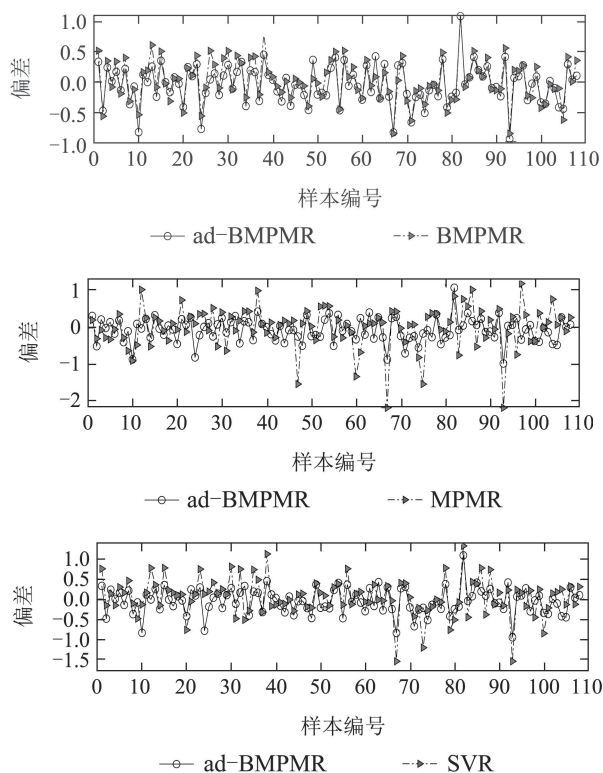


图6 汽油RON的预测偏差曲线

Fig. 6 Bias curves for gasoline RON estimation

5 结语

汽油调合是汽油产品出厂前的最后一道工序,调合过程的优化控制直接关系到产品的质量. 为了保证一次调合成功,避免重调带来的油品损失、储罐资源浪费,订单交付延迟等;希望优化过程中汽油属性的预测值等于或偏低实际值,并保持在可接受的误差范围内. 这样优化控制器会始终确保调合产品合格,同时又能有效控制属性质量过剩. 针对这一特性,本文提出采用有偏最小最大概率回归算法建立汽油属性的近红外分析模型;考虑到初始近红外模型会随着组分油属性,调合配方等的改变而逐渐衰减,使模型预测误差不断增大,本文将即时学习算法与最小最大概率回归相结合,采用局部建模策略提高模型的自适应能力,从而既可保证对油品属性的有偏估计,又能及时跟踪生产过程的最新动态. 在柴油和汽油近红外数据上的实验结果验证了方法的有效性,本文方法有望大幅度提高油品调合优化控制的投运率.

参考文献:

[1] FANG Yongchun, WANG Ning, WANG Shuqing. DMC-expert control of gasoline in-line blending processes. *Control Theory & Applications*, 2008, 25(3): 560 – 563.
(方勇纯, 王宁, 王树青. 汽油管道调合过程的动态矩阵预测—专家控制. 控制理论与应用, 2008, 25(3): 560 – 563.)

[2] LIU Zhao. *Design and application of gasoline blending system*. Shanghai: East China University of Science and Technology, 2010.
(刘朝. 汽油调合系统的设计及应用. 上海: 华东理工大学, 2010.)

[3] KELLY J J, BARLOW C H, JINGUJI T M, et al. Prediction of gasoline octane numbers from near-infrared spectral features in the range 660~1215 nm. *Analytical Chemistry*, 1989, 61(4): 313 – 320.

[4] WAN J H, HAN Z Z, LIU K W. RON predicted of gasoline by NIR based on ICA and SVM. *International Computer Conference on Wavelet Active Media Technology & Information Processing*. Chengdu, China: IEEE, 2016: 498 – 501.

[5] CHEN M, KHARE S, HUANG B, et al. Recursive wavelength-selection strategy to update near-infrared spectroscopy model with an industrial application. *Industrial & Engineering Chemistry Research*, 2013, 52(23): 7886 – 7895.

[6] HE Yanlin, WANG Xiao, ZHU Qunxiong. Modeling of acetic acid content in purified terephthalic acid solvent column using principal component analysis based improved extreme learning machine. *Control Theory & Applications*, 2015, 32(1): 80 – 85.
(贺彦林, 王晓, 朱群雄. 基于主成分分析-改进的极限学习机方法的精对苯二甲酸醋酸含量软测量. 控制理论与应用, 2015, 32(1): 80 – 85.)

[7] SOARES F, ANZANELLO M J. Support vector regression coupled with wavelength selection as a robust analytical method. *Chemometrics & Intelligent Laboratory Systems*, 2018, 172: 167 – 173.

[8] BALABIN R M, SAFIEVA R Z, Lomakina E I. Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. *Chemometrics & Intelligent Laboratory Systems*, 2007, 88(2): 183 – 188.

[9] RODRIGUEZ-GALIANO V, SANCHEZ-CASTILLO M, CHICA-OLMO M, et al. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 2015, 71: 804 – 818.

[10] HUANG Caifeng, LI Xin, LI Shaoyuan. Gasoline blend optimization based on Ethyl RT-70 models. *Control Engineering of China*, 2007, 14(3): 256 – 259.
(黄彩凤, 李信, 李少远. 基于Ethyl RT-70模型的汽油调合优化. 控制工程, 2007, 14(3): 256 – 259.)

[11] DU Mi. *Dynamic modeling of gasoline properties based on near infrared spectroscopy*. Nanjing: Southeast University, 2017.
(杜咪. 基于近红外光谱的汽油性质动态建模方法研究. 南京: 东南大学, 2017.)

[12] HE Renchu, CHEN Haiquan, YANG Chaowen. Discussion and analysis on optimization technology of gasoline blending for Chinese national standard IV. *Chemical Industry and Engineering Progress*, 2018, 37(3): 962 – 969.
(何仁初, 陈海泉, 杨超文. 面向国IV标准的汽油调合优化技术探讨与分析. 化工进展, 2018, 37(3): 962 – 969.)

[13] WANG W, JIANG H, LIU G, et al. Quantitative analysis of yeast growth process based on FT-NIR spectroscopy integrated with Gaussian mixture regression. *RSC Advances*, 2017, 7(40): 24988 – 24994.

[14] CHEN Jie, WANG Ning, WANG Shuqing. Neuro-internal model based optimal control for gasoline blending process. *Control Theory & Applications*, 1999, 26(4): 532 – 535.
(陈捷, 王宁, 王树青. 汽油调合过程的神经内模优化控制. 控制理论与应用, 1999, 26(4): 532 – 535.)

[15] CHU Xiaoli, LU Wanzhen. Research and application progress of near infrared spectroscopy analytical technology in China in the past five years. *Spectroscopy and Spectral Analysis*, 2014, 34(10): 2595 – 2605.
(褚小立, 陆婉珍. 近五年我国近红外光谱分析技术研究与应用进展. 光谱学与光谱分析, 2014, 34(10): 2595 – 2605.)

- [16] STROHMANN T, GRUDIC G Z. A formulation for minimax probability machine regression. *Advances in Neural Information Processing Systems*. Massachusetts, USA: MIT Press, 2003: 769 – 776.
- [17] SAMUI P. Use of minimax probability machine regression for modelling of settlement of shallow foundations on cohesionless soil. *International Journal of Performability Engineering*, 2014, 10(3): 325 – 328.
- [18] SHAH V S, SHAH H R, SAMUI P, et al. Prediction of fracture parameters of high strength and ultra-high strength concrete beams using minimax probability machine regression and extreme learning machine. *Computers Materials & Continua*, 2014, 44(2): 73 – 84.
- [19] HUANG K, YANG H, KING I, et al. Learning classifiers from imbalanced data based on biased minimax probability machine. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, DC, USA: IEEE, 2004, 2: 558 – 563.
- [20] HUANG K, YANG H, KING I. The minimum error minimax probability machine. *Journal of Machine Learning Research*, 2004, 5(4): 1253 – 1286.
- [21] LI J, KARIMI I A, SRINIVASAN R. Recipe determination and scheduling of gasoline blending operations. *AIChE Journal*, 2010, 52(1): 441 – 465.
- [22] HE K X, QIAN F, DU W L. Improved integrated optimization method of gasoline blend planning and real-time blend recipes. *Industrial & Engineering Chemistry Research*, 2016, 55(16): 4632 – 4645.
- [23] MIN H, LUO X L. Calibration of soft sensor by using just-in-time modeling and AdaBoost learning method. *Chinese Journal of Chemical Engineering*, 2016, 24(8): 1038 – 1046.
- [24] SHAO W M, TIAN X M, WANG P. Supervised local and nonlocal structure preserving projections with application to just-in-time learning for adaptive soft sensor. *Chinese Journal of Chemical Engineering*, 2015, 23(12): 1925 – 1934.
- [25] GE Z Q, SONG Z H. A comparative study of just-in-time-learning based methods for online soft sensor modeling. *Chemometrics and Intelligent Laboratory Systems*, 2010, 104(2): 306 – 317.
- [26] MIN H, LUO X L. Enhanced just-in-time soft sensor calibration method using data density estimation. *Chemometrics & Intelligent Laboratory Systems*, 2017, 161: 79 – 87.
- [27] YUAN X F, HUANG B, GE Z Q. Double locally weighted principal component regression for soft sensor with sample selection under supervised latent structure. *Chemometrics & Intelligent Laboratory Systems*, 2016, 153: 116 – 125.
- [28] LIU Y, GAO Z, LI P. Just-in-time kernel learning with adaptive parameter selection for soft sensor modeling of Batch processes. *Industrial & Engineering Chemistry Research*, 2012, 51(11): 4313 – 4327.
- [29] YUAN X F, GE Z Q, SONG Z H. Locally weighted kernel principal component regression model for soft sensing of nonlinear time-variant processes. *Industrial & Engineering Chemistry Research*, 2014, 53(35): 13736 – 13749.
- [30] KIM S, KANO M, HASEBE S, et al. Long-term industrial applications of inferential control based on just-in-time soft-sensors: economical impact and challenges. *Industrial & Engineering Chemistry Research*, 2013, 52(35): 12346 – 12356.
- [31] QIN S J. Recursive PLS algorithms for adaptive data modeling. *Computers & Chemical Engineering*, 1998, 22(4/5): 503 – 514.

作者简介:

贺凯迅 讲师, 硕士生导师, 主要研究方向为数据驱动软测量、故障诊断及异常工况识别, E-mail: kaixunhe@sdust.edu.cn;

刘晶晶 硕士研究生, 主要研究方向为数据驱动软测量建模, E-mail: qsskliujingjing@163.com;

王小邦 硕士研究生, 主要研究方向为故障诊断、异常工况识别, E-mail: 759750603@qq.com;

苏照阳 硕士研究生, 主要研究方向为故障诊断、异常工况识别, E-mail: zhaoyangsu8@gmail.com.