

基于核的冶金煤气流量在线区间预测

赵 珺[†], 杜雅楠, 盛春阳, 王 伟

(大连理工大学 信息与控制研究中心, 辽宁 大连 116023)

摘要: 针对冶金企业煤气系统的流量区间预测问题, 本文提出一种基于核的在线区间预测构造方法, 该方法将传统预测区间构造过程中对雅克比矩阵的复杂计算转化为对核的计算, 大大降低了计算成本. 为确定所提方法的超参数, 采用共轭梯度下降算法来优化模型预测误差, 使其逼近样本数据中有效噪声的方差. 为验证本文所提方法的有效性, 对现场实时数据库中的煤气流量数据进行了仿真实验, 其结果表明本文方法在预测精度、可靠性和实时性三方面都表现出明显的优势.

关键词: 煤气流量; 核方法; 区间预测; 参数优化; 共轭梯度

中图分类号: TP273 **文献标识码:** A

Kernel-based method for predicting online gas flow interval in metallurgical enterprises

ZHAO Jun[†], DU Ya-nan, SHENG Chun-yang, WANG Wei

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

Abstract: In the kernel-based method, the calculation of the Jacobian matrix for determining the predicted interval in routine methods is converted into the calculation of kernels, thus greatly reducing the calculation costs. The hyper-parameters of the proposed model are determined by employing the conjugate gradient algorithm to minimize the model prediction error, making it to approach the variance of the effective noises in the sample data. To verify the effectiveness of the proposed method, we apply this method to construct prediction intervals of real gas flow data collection from the energy center of a steel plant. Results indicate that the proposed method is highly accurate and reliable with low computational costs.

Key words: gas flow; kernel-based method; interval prediction; parameter optimization; conjugate gradient

1 引言(Introduction)

冶金企业副产煤气是重要的二次能源, 其有效合理的利用直接影响到钢铁企业的经济效益和节能减排水平. 为了实现副产煤气的有效利用, 须对煤气系统进行有效的实时调度. 然而由于煤气系统具备非线性、复杂性和耦合性等特点, 很难通过基于机理和工艺的方法对煤气系统进行调度. 现场调度人员多结合基于数据的煤气流量预测结果对系统进行定性平衡调度, 这使得副产煤气流量预测至关重要^[1-3].

目前已经存在大量的关于副产煤气流量预测的研究, 主要分为基于神经网络的预测方法和基于支持向量回归的预测方法. 文献[4-5]采用了BP方法对高炉煤气的发生量进行了预测, 不同的是前者是利用影响高炉煤气发生量的因素进行预测, 而后者是基于时间序列的预测. 文献[6]采用了回声状态网络对高炉煤气流量进行了预测, 并针对工业数据含噪声的特点, 提

出采用对偶卡尔曼滤波方法对网络的权值参数进行估计, 有效的改善了预测效果. 文献[1]采用改进BP神经网络方法对焦炉煤气流量进行了预测, 提高了模型的收敛速度和预测精度. 文献[2]采用了基于并行的粒子群优化算法的支持向量回归算法对转炉煤气流量进行了预测, 其预测效果优于普通的支持向量回归和BP方法. 文献[3]针对煤气数据的特点, 提出一种基于数据驱动的子空间方法来预测煤气流量, 并采用粒子群算法来优化参数, 得到了较好的预测结果. 以上的预测研究都是基于点的预测的模式, 这种模式只能给出预测的结果, 而不能给出任何关于预测可靠性的说明, 这增加了现场调度人员调度操作的风险. 区间预测不仅能够给出预测结果, 同时能够给出与预测结果相应的可信度, 这是调度人员所期望的. 常用的构造预测区间的方法是Bootstrap^[7]和Delta^[8]等, 但由于其构造过程需要计算雅克比矩阵, 时间冗长, 而不利

于在工业实际这种实时问题中进行应用. 文献[9]提出了一种基于支持向量回归的区间预测方法, 由于支持向量回归模型的参数中包含偏置 b 而使得模型的参数估计往往是有偏的, 尽管文献[9]中做了相应的处理, 但大大增加了计算的复杂性.

考虑到冶金企业煤气系统的流量数据复杂多变, 性质各异的特点, 本文提出一种基于双高斯核的在线区间预测构造方法, 通过两个核的相互调节同样可以很好的描述非线性系统的动态, 同时在构造预测区间时, 可以将对雅克比矩阵的计算转化为对核的计算, 降低了计算复杂度. 为确定所提方法的超参数, 本文利用共轭梯度下降算法来优化模型预测误差, 使其尽可能逼近样本数据中有效噪声的方差, 为验证本文所提方法的有效性, 对现场实时数据库中的煤气流量数据进行了仿真实验, 其结果表明本文方法在预测精度、可靠性和实时性3个方面都表现出明显的优势.

2 问题描述(Description of the problem)

图1所示为某一钢铁企业的副产煤气系统结构简

图. 从工艺上看副产煤气系统主要由发生源、消耗源、存储设备和输送设备组成. 副产煤气系统的发生源主要是高炉炼铁过程产生高炉煤气, 焦炉炼焦过程产生焦炉煤气, 转炉炼钢过程回收转炉煤气; 煤气柜是副产煤气系统主要的存储设备, 在系统中起缓冲作用, 管网是主要的输送设备; 副产煤气系统的消耗用户主要是制造过程的工序或重要设备, 如热风炉、焦炉、冷轧热处理、加热炉、电炉、钢管厂等, 这些用户有的燃烧多种煤气的混合气体, 有的仅燃烧某一种煤气介质, 如连铸过程仅消耗焦炉煤气. 理想状态下副产煤气系统应保持着供需平衡, 而在实际生产过程中, 由于受钢铁生产工艺和生产计划的影响, 煤气系统经常趋于不平衡, 这就需要调度人员及时地对煤气系统进行调整, 使煤气系统重新达到平衡, 以保证生产的安全和高效. 鉴于副产煤气系统的特点, 调度人员很难建立一套定量调度方案, 实际生产中往往通过预测, 提前了解系统中各煤气单元对煤气的产量或者用量, 并结合预测结果定性对系统进行调度.

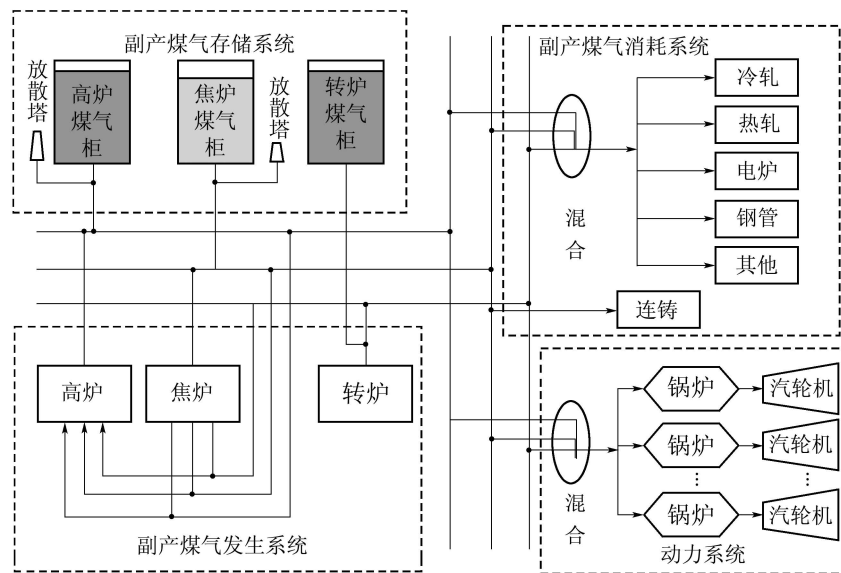


图 1 某钢铁企业副产煤气系统管网图

Fig. 1 Structure of the gas system in an iron and steel enterprise

煤气系统各产消单元机理和工艺非常复杂, 如高炉炼铁和焦炉炼焦, 很难通过机理建模的方法预测煤气系统的产消流量. 由于现场实时数据库中已保存有大量的可用数据, 通过基于数据的预测方法对煤气系统的产消流量进行预测. 但是已有的关于这方面的研究缺乏对预测结果的可靠性的说明^[1-3,6], 由于数学模型不可能完全描述数据的非线性特性, 且工业数据普遍存在噪声, 因此预测结果的精度难以得到保证. 而区间预测综合考虑了模型的不确定性和数据的不确定性, 不仅能够给出更准

确的预测结果, 同时能够给出预测结果可能的变化区间^[10].

3 煤气流量区间预测方法(Interval prediction method of gas flow)

3.1 核回归方法(Kernel regression)

由于工业数据普遍存在噪声, 因此在回归学习过程中观测值可用系统真实输出值和附加数据噪声来表示:

$$t_n = f(\mathbf{x}_n, \mathbf{w}) + \gamma_n, \quad (1)$$

其中: \mathbf{w} 是线性回归模型的权值参数, 可以定义 $f(\mathbf{x}_n, \mathbf{w}) = \mathbf{w}^T \phi(\mathbf{x}_n)$, γ_n 是零均值高斯白噪声, 考虑上述的线性回归模型, 其参数可以通过最小化一个正则化的均方和误差来求得^[11], 具体如式(2)所示:

$$J(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{\mathbf{w}^T \phi(\mathbf{x}_n) - t_n\}^2 + \frac{\lambda}{2} \mathbf{w}^T \mathbf{w}, \quad (2)$$

其中 $\lambda \geq 0$. 如果将式(2)对权值参数 \mathbf{w} 求导数, 因函数 $J(\mathbf{w})$ 是凸的, 所以将其导数等于零的点视为最优点, 可得到式(4)所示的结果, 权值 \mathbf{w} 可以写成向量 $\phi(\mathbf{x}_n)$ 的线性组合形式:

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \sum_{n=1}^N \{\mathbf{w}^T \phi(\mathbf{x}_n) - t_n\} \phi(\mathbf{x}_n) + \lambda \mathbf{w} = 0, \quad (3)$$

$$\mathbf{w} = -\frac{1}{\lambda} \sum_{n=1}^N \{\mathbf{w}^T \phi(\mathbf{x}_n) - t_n\} \phi(\mathbf{x}_n) =$$

$$\sum_{n=1}^N a_n \phi(\mathbf{x}_n) = \Phi^T \mathbf{a}, \quad (4)$$

其中 $a_n = -\{\mathbf{w}^T \phi(\mathbf{x}_n) - t_n\} / \lambda$, 将式(4)的结果代入式(2)可以重新构造以 \mathbf{a} 为自变量的新的优化目标, 也就是原函数的对偶表示形式.

$$J(\mathbf{a}) = \frac{1}{2} \mathbf{a}^T \mathbf{K} \mathbf{K} \mathbf{a} - \mathbf{a}^T \mathbf{K} \mathbf{t} + \frac{1}{2} \mathbf{t}^T \mathbf{t} + \frac{\lambda}{2} \mathbf{a}^T \mathbf{K} \mathbf{a}, \quad (5)$$

其中: $\mathbf{t} = (t_1, t_2, \dots, t_N)^T$, 定义矩阵 $\mathbf{K} = \Phi \Phi^T$, \mathbf{K} 是一个 $N \times N$ 的矩阵, 包含元素 K_{nm} , K_{nm} 可以用核函数 $k(\mathbf{x}_n, \mathbf{x}_m) = \phi^T(\mathbf{x}_n) \phi(\mathbf{x}_m)$ 来表示, 将式(5)中函数对 \mathbf{a} 进行求导数并置零, 可以得到参数 \mathbf{a} 的表达式

$$\mathbf{a} = (\mathbf{K} + \lambda \mathbf{I}_N)^{-1} \mathbf{t}. \quad (6)$$

如此, 将式(4)(6)的结果代入原回归函数, 可以得到新的线性回归模型:

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) = \sum_{n=1}^N a_n \phi(\mathbf{x}_n) \phi(\mathbf{x}) =$$

$$\mathbf{k}(\mathbf{x})^T (\mathbf{K} + \lambda \mathbf{I}_N)^{-1} \mathbf{t}, \quad (7)$$

其中 $\mathbf{k}(\mathbf{x})$ 是由 N 个核函数组成的向量, 每个元素可以表示为 $k_n(\mathbf{x}) = k(\mathbf{x}_n, \mathbf{x})$. 基于核的方法避开了对基函数 $\phi(\mathbf{x})$ 的研究, 而转换为对核函数的选择, 通常情况下核函数有多种形式, 其中最常用的是高斯核函数, 其应用效果已有大量验证^[12-14]. 本文所提方法的核函数是两个不同高斯核的加权形式:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \beta k_1(\mathbf{x}_i, \mathbf{x}_j) + (1 - \beta) k_2(\mathbf{x}_i, \mathbf{x}_j), \quad (8)$$

$$k_l(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_l^2}\right), \quad l=1, 2. \quad (9)$$

3.2 基于核的区间预测方法(Kernel-based interval prediction methods)

对于工业预测问题, 很大程度调度人员还关心预测结果的可靠性, 因此本文为预测结果构造预测区间. 预测区间可以描述出预测结果带来的风险, 主要包含两部分的内容, 一是参数估计对预测结果的影响, 二是数据含有的固有噪声对预测结果的影响^[10]. 其中参数估计不可能完全的描述非线性系统的动态, 如式(10), $\hat{\mathbf{a}}$ 是估计的参数, \mathbf{a}^* 是非线性系统的实际动态, 是一个隐变量, 零均值高斯白噪声 ε_n 描述的是系统输出值与模型输出值的偏差.

$$y(\mathbf{x}_n, \mathbf{a}^*) = y(\mathbf{x}_n, \hat{\mathbf{a}}) + \varepsilon_n. \quad (10)$$

实际系统通常都是伴随噪声和干扰的, 也就是说真实的观测值与系统的真实输出值同样存在着偏差, 这个偏差可以用 γ_n 来表示, 于是

$$t_n = y(\mathbf{x}_n, \hat{\mathbf{a}}) + \varepsilon_n + \gamma_n. \quad (11)$$

虽然模型不可能完全描述非线性系统, 但是可以将模型输出值 $y(\mathbf{x}_0, \hat{\mathbf{a}})$ 在真实系统参数 \mathbf{a}^* 附近进行一阶的泰勒公式展开, 那么 $y(\mathbf{x}_0, \hat{\mathbf{a}})$ 可以近似地表示为

$$y(\mathbf{x}_0, \hat{\mathbf{a}}) \approx y(\mathbf{x}_0, \mathbf{a}^*) + \mathbf{g}_0^T (\hat{\mathbf{a}} - \mathbf{a}^*), \quad (12)$$

其中

$$\mathbf{g}_0^T = \left[\frac{\partial y(\mathbf{x}_0, \mathbf{a}^*)}{\partial a_1^*} \quad \frac{\partial y(\mathbf{x}_0, \mathbf{a}^*)}{\partial a_2^*} \quad \dots \quad \frac{\partial y(\mathbf{x}_0, \mathbf{a}^*)}{\partial a_N^*} \right].$$

对于基于核的方法, 特殊的是

$$\frac{\partial y(\mathbf{x}_0, \mathbf{a}^*)}{\partial a_1^*} = \mathbf{k}(\mathbf{x}_0, \mathbf{x}_1),$$

令 $\mathbf{k}(\mathbf{x}_0) = [k(\mathbf{x}_0, \mathbf{x}_1) \quad k(\mathbf{x}_0, \mathbf{x}_2) \quad \dots \quad k(\mathbf{x}_0, \mathbf{x}_N)]$, 则 $\mathbf{g}_0^T = \mathbf{k}(\mathbf{x}_0)$.

如此, 根据式(1)(11)–(12), 实际观测值与模型输出值的偏差可以近似为式(13), 假设参数 $\hat{\mathbf{a}}$ 是真实系统 \mathbf{a}^* 的无偏估计, 那么, 式(13)中的偏差的期望值将满足式(14).

$$t_0 - \hat{y}_0 \approx \gamma_0 - \mathbf{g}_0^T (\hat{\mathbf{a}} - \mathbf{a}^*), \quad (13)$$

$$E[t_0 - \hat{y}_0] \approx E[\gamma_0] - \mathbf{g}_0^T E[(\hat{\mathbf{a}} - \mathbf{a}^*)] \approx 0. \quad (14)$$

因为统计的独立性, 所以 $t_0 - \hat{y}_0$ 的方差可以写成

$$\text{var}[t_0 - \hat{y}_0] \approx \text{var}[\gamma_0] + \text{var}[\mathbf{g}_0^T (\hat{\mathbf{a}} - \mathbf{a}^*)]. \quad (15)$$

假设高斯白噪声 γ_0 的方差为 σ_γ^2 , 即 γ_0 服从分布 $\mathcal{N}(0, \sigma_\gamma^2)$, 那么分布 $(\hat{\mathbf{a}} - \mathbf{a}^*)$ 可以近似为分布 $\mathcal{N}(0, \sigma_\gamma^2 [F^T(\hat{\mathbf{a}}) F(\hat{\mathbf{a}})]^{-1})$. $F(\hat{\mathbf{a}})$ 为函数 $y(\mathbf{x}_0, \hat{\mathbf{a}})$ 关于 $\hat{\mathbf{a}}$ 的

雅克比矩阵, 如式(16):

$$F(\hat{\mathbf{a}}) = \frac{\partial y(\mathbf{x}, \hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}} = \begin{bmatrix} \frac{\partial f_1(\mathbf{x}_1, \hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}_1} & \frac{\partial f_1(\mathbf{x}_1, \hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}_2} & \dots & \frac{\partial f_1(\mathbf{x}_1, \hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}_N} \\ \frac{\partial f_2(\mathbf{x}_2, \hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}_1} & \frac{\partial f_2(\mathbf{x}_2, \hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}_2} & \dots & \frac{\partial f_2(\mathbf{x}_2, \hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_N(\mathbf{x}_N, \hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}_1} & \frac{\partial f_N(\mathbf{x}_N, \hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}_2} & \dots & \frac{\partial f_N(\mathbf{x}_N, \hat{\mathbf{a}})}{\partial \hat{\mathbf{a}}_N} \end{bmatrix} \quad (16)$$

因为 $\frac{\partial y(\mathbf{x}_i, \mathbf{a}^*)}{\partial a_j^*} = k(\mathbf{x}_i, \mathbf{x}_j)$, 所以 $F(\hat{\mathbf{a}}) = K$ 是一个核矩阵, 由此体现了基于核的方法的优势, 大大的降低了计算的复杂性, 对于工业问题要求实时性具有重要的意义. 这样, 式(15)的方差可表示为

$$\text{var}[t_0 - \hat{y}_0] \approx \sigma_\gamma^2 + \sigma_\gamma^2 \mathbf{g}_0^T (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{g}_0, \quad (17)$$

其中: 第1项 σ_γ^2 是指因数据的固有噪声而导致预测结果的偏差, 而第2项则是指由于模型参数估计不可能完全逼近真实系统而导致预测结果的偏差. 基于上述结论, 对于预测值 \hat{y}_0 其 $(1-\alpha)\%$ 的预测区间可以描述为

$$\hat{y}_0 \pm t_{n-p}^{\alpha/2} (\sigma_\gamma^2 + \sigma_\gamma^2 \mathbf{g}_0^T (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{g}_0)^{1/2}, \quad (18)$$

其中: α 为 t 分布上的分位点, $n-p$ 为 t 分布的自由度.

4 基于噪声估计的超参数优化方法(Parameter optimization method based on noise estimates)

基于核计算的方法将回归模型的参数转化为对核的运算, 如式(6)所示. 如此, 基于核计算的方法仅仅需要估计超参数, 即惩罚因子 λ , 核的加权因子 β 和核函数的宽度 σ_1 和 σ_2 . 对于式(18)所示的预测区间, 数据所含固有噪声的方差 σ_γ^2 也需估计. 本节采用 Gamma 实验技术对数据中的有效噪声进行估计^[15].

为了确定较优的模型参数, 定义优化目标如式(19)所示. 因为预测误差一般认为由数据噪声和模型的欠精确共同导致, 若将预测误差的方差逼近噪声方差, 那么相当于将模型的误差逼近于零, 这样求得模型参数将会有最优的性能.

$$e = \text{var}(\text{train}) = |\text{var}(\text{pred}) - \text{var}(\text{noise})|, \quad (19)$$

$$\text{var}(\text{pred}) = \text{var}(t - f(\mathbf{x}, \hat{\mathbf{a}})) =$$

$$\text{var}(f(\mathbf{x}, \hat{\mathbf{a}}) + \varepsilon + \gamma - f(\mathbf{x}, \hat{\mathbf{a}})) =$$

$$\text{var}(\varepsilon) + \text{var}(\gamma), \quad (20)$$

其中预测误差可以用式(20)表示, 而噪声的方差

$\text{var}(\text{noise}) = \text{var}(\gamma)$. 可见, 模型训练误差 e 实际上就是 $|\text{var}(\varepsilon)|$, e 是一个假设存在的隐变量.

根据式(19)–(20), 训练误差 e 可以用下述的公式描述, 预测模型的最优参数, 就是将式(21)的结果逼近于最小值 0 时的结果. 公式中的噪声方差由 Gamma 实验估计获得, 而 $\{\mathbf{x}_n, t_n\}_{n=1,2,\dots,N}$ 为训练样本.

$$e = |\text{var}(\varepsilon)| = \left| \frac{1}{2} \sum_{n=1}^N \{t_n - \mathbf{k}^T(\mathbf{x}_n) \hat{\mathbf{a}}\}^2 - \sigma_\gamma^2 \right|. \quad (21)$$

本文这里采用共轭梯度法来解决上述优化问题, 以求得回归模型的超参数, 首先对将式(21)中的函数推导关于自变量的偏导数, 也就是关于惩罚因子 λ , 核的加权因子 β 和核函数的宽度 σ_1 和 σ_2 的偏导数, 具体如式(22)所示:

$$\begin{cases} \frac{\partial e}{\partial \sigma_1} = \frac{e}{\text{var}(\varepsilon)} \sum_{n=1}^N \{t_n - \mathbf{k}^T(\mathbf{x}_n) \hat{\mathbf{a}}\} \frac{\partial (\mathbf{k}^T(\mathbf{x}_n) \hat{\mathbf{a}})}{\partial \sigma_1}, \\ \frac{\partial e}{\partial \beta} = \frac{e}{\text{var}(\varepsilon)} \sum_{n=1}^N \{t_n - \mathbf{k}^T(\mathbf{x}_n) \hat{\mathbf{a}}\} \frac{\partial (\mathbf{k}^T(\mathbf{x}_n) \hat{\mathbf{a}})}{\partial \beta}, \\ \frac{\partial e}{\partial \lambda} = \frac{e}{\text{var}(\varepsilon)} \sum_{n=1}^N \{t_n - \mathbf{k}^T(\mathbf{x}_n) \hat{\mathbf{a}}\} \mathbf{k}^T(\mathbf{x}_n) \frac{\partial \hat{\mathbf{a}}}{\partial \lambda}, \\ \frac{\partial (\mathbf{k}^T(\mathbf{x}_n) \hat{\mathbf{a}})}{\partial \sigma_1} = -\frac{\partial \mathbf{k}^T(\mathbf{x}_n)}{\partial \sigma_1} \hat{\mathbf{a}} - \mathbf{k}^T(\mathbf{x}_n) \frac{\partial \hat{\mathbf{a}}}{\partial \sigma_1}, \\ \frac{\partial (\mathbf{k}^T(\mathbf{x}_n) \hat{\mathbf{a}})}{\partial \beta} = \frac{\partial \mathbf{k}^T(\mathbf{x}_n)}{\partial \beta} \hat{\mathbf{a}} + \mathbf{k}^T(\mathbf{x}_n) \frac{\partial \hat{\mathbf{a}}}{\partial \beta}, \end{cases} \quad (22)$$

其中

$$\mathbf{k}^T(\mathbf{x}_n) = [k(\mathbf{x}_n, \mathbf{x}_1) \ \dots \ k(\mathbf{x}_n, \mathbf{x}_N)],$$

由于式(22)中存在多个中间变量, 如核向量 $\mathbf{k}(\mathbf{x}_n)$ 和模型的参数 $\hat{\mathbf{a}}$ 是超参数 β, σ_1 和 σ_2 的函数, 因此需要对中间变量求解关于自变量的偏导数, 具体如式(23)所示:

$$\begin{cases} \frac{\partial \mathbf{k}^T(\mathbf{x}_n)}{\partial \sigma_1} = \left[\frac{\partial k(\mathbf{x}_n, \mathbf{x}_1)}{\partial \sigma_1} \ \frac{\partial k(\mathbf{x}_n, \mathbf{x}_2)}{\partial \sigma_1} \ \dots \right. \\ \left. \frac{\partial k(\mathbf{x}_n, \mathbf{x}_N)}{\partial \sigma_1} \right], \\ \frac{\partial \mathbf{k}^T(\mathbf{x}_n)}{\partial \beta} = \left[\frac{\partial k(\mathbf{x}_n, \mathbf{x}_1)}{\partial \beta} \ \frac{\partial k(\mathbf{x}_n, \mathbf{x}_2)}{\partial \beta} \ \dots \right. \\ \left. \frac{\partial k(\mathbf{x}_n, \mathbf{x}_N)}{\partial \beta} \right], \\ \frac{\partial \hat{\mathbf{a}}}{\partial \sigma_1} = -t_n (\mathbf{K} + \lambda \mathbf{I}_N)^{-1} \frac{\partial \mathbf{K}}{\partial \sigma_1} (\mathbf{K} + \lambda \mathbf{I}_N)^{-1}, \\ \frac{\partial \hat{\mathbf{a}}}{\partial \beta} = -t_n (\mathbf{K} + \lambda \mathbf{I}_N)^{-1} \frac{\partial \mathbf{K}}{\partial \beta} (\mathbf{K} + \lambda \mathbf{I}_N)^{-1}, \\ \frac{\partial \hat{\mathbf{a}}}{\partial \lambda} = -t_n (\mathbf{K} + \lambda \mathbf{I}_N)^{-1} (\mathbf{K} + \lambda \mathbf{I}_N)^{-1}, \end{cases} \quad (23)$$

其中: \mathbf{K} 是核矩阵, $\frac{\partial \mathbf{K}}{\partial \sigma_1}$ 是以 $\frac{\partial k(\mathbf{x}_i, \mathbf{x}_j)}{\partial \sigma_1}$ 为单元的维数为 $N \times N$ 的矩阵, 式中关于核函数的定义可见式

(6)-(7), $\frac{\partial K}{\partial \beta}$ 是以 $\frac{\partial k(\mathbf{x}_i, \mathbf{x}_j)}{\partial \beta}$ 为单元的维数为 $N \times N$ 的矩阵, 因此还要求解核函数 $k(\mathbf{x}_i, \mathbf{x}_j)$ 关于超参数 β, σ_1 和 σ_2 的偏导数, 如式(24)所示:

$$\begin{cases} \frac{\partial k(\mathbf{x}_i, \mathbf{x}_j)}{\partial \sigma_1} = \beta k_1(\mathbf{x}_i, \mathbf{x}_j) \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma_1^3}, \\ \frac{\partial k(\mathbf{x}_i, \mathbf{x}_j)}{\partial \sigma_2} = (1 - \beta) k_2(\mathbf{x}_i, \mathbf{x}_j) \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma_2^3}, \\ \frac{\partial k(\mathbf{x}_n, \mathbf{x}_1)}{\partial \beta} = k_1(\mathbf{x}_n, \mathbf{x}_1) - k_2(\mathbf{x}_n, \mathbf{x}_1). \end{cases} \quad (24)$$

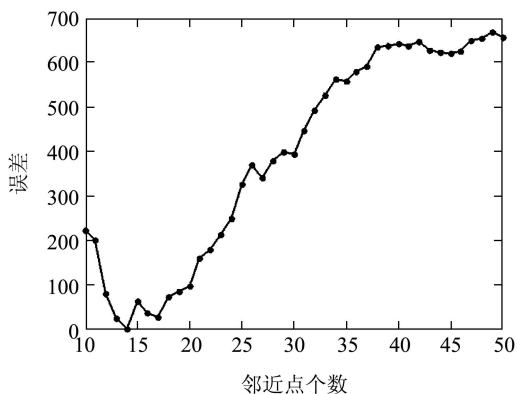
本文以共轭梯度法优化超参数, 步长搜索采用 Wolfe-Powell 非精确一维搜索方法, 虽然梯度下降算法易陷入局部最优, 但是本文实验发现, 如果选取初始点合适, 算法会快速收敛到一个较优参数组合, 因此初始值的选定对算法至关重要, 可以先基于网格搜索的快速交叉验证方法离线的确定初始值, 对一个序列来讲初始值一旦确定, 一般可以在应用时保持不变. 如此可以为预测模型获得较优的超参数.

5 数据试验及分析(Data experiment and analysis)

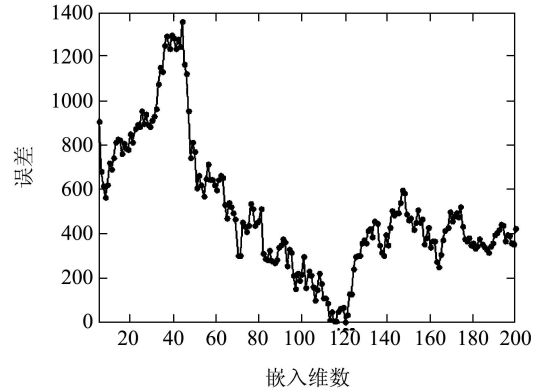
为验证所提出的基于核的冶金煤气流量区间预测方法的有效性, 选取宝钢能源中心的煤气系统 2012 年 11 月份的高炉煤气用户数据进行验证, 数据采样频率为 1 分钟. 实验部分具体包括参数优化实验和区间预测验证实验两部分.

5.1 模型参数优化(Model parameter optimization)

影响 Gamma 实验噪声估计结果的参数主要有邻近点个数 p 和嵌入维数 m 两个参数. 根据经验, 本文选取 700 个训练样本, 采用交叉验证的方法来确定邻近点个数和嵌入维数, 通过试验得到当邻近点个数为 14, 嵌入维数为 120 时, 可以估计出较为准确的数据噪声. 图 2 为采用交叉验证方法优化模型参数的结果图.



(a) 邻近点个数对噪声方差的影响



(b) 嵌入维数对噪声方差的影响

图 2 模型参数优化结果

Fig. 2 Model parameter optimization results

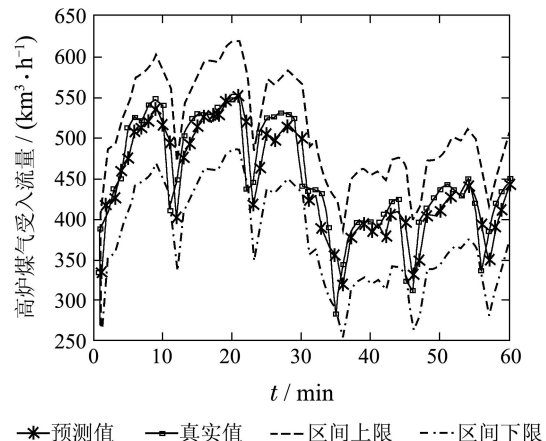
图 2(a) 中, 当模型嵌入维数为 120 时, 可以看出当临近点等于 14 时, Gamma 实验估计噪声的误差最小, 图 2(b) 中, 当临近点个数设置为 14 时, 模型嵌入维数等于 120 时, Gamma 实验估计噪声的误差最小.

通过 Gamma 实验得到训练样本个数为 $n = 700$, 嵌入维数为 $m = 120$, 临近点个数 $p = 14$, 选取网格搜索 γ 范围为 (1, 800), σ_1 范围为 (10, 500), σ_2 范围为 (1, 10) 网格数为 1000, 取 50 段数据进行参数优化, 取得到的 50 个超参数点的几何中心点作为最终的最优超参数, 本实验的结果为

$$(\gamma, \sigma_1, \sigma_2)_{\text{best}} = (11.8, 255.9, 1.7).$$

5.2 高炉煤气受入流量区间预测 (Interval prediction of the generation amount of the blast furnace gas)

在确定好模型的参数之后, 开始为高炉煤气受入流量构造预测区间, 如图 3 所示为采用本文所提方法构造的高炉煤气受入流量预测区间, 区间置信度为 95%.



(a) 预测效果

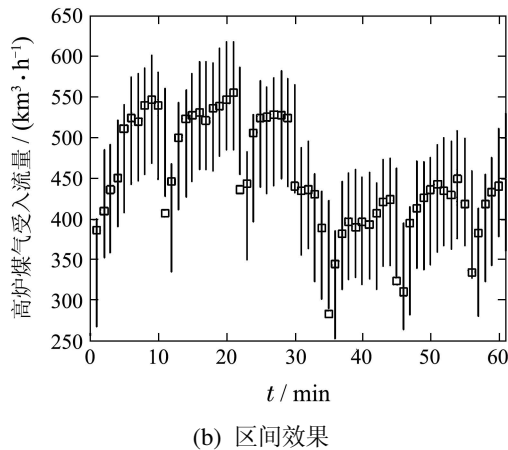


图 3 高炉煤气受入流量预测区间

Fig. 3 Prediction interval of the generation amount of the blast furnace gas

从图3(a)中可以看出本文方法能够准确对高炉煤气受入流量进行预测,从图3(b)中可以看出高炉煤气受入流量的真实值基本都落在所构造的预测区间之内,图3整体上说明了本文所提方法在为高炉煤气受入流量构造预测区间时的有效性。

为进一步说明本文算法在构造高炉煤气受入流量预测区间时的优势,文中还与基于Bootstrap的神经网络方法、基于Delta的神经网络方法和最小二乘支持向量机方法进行了比较,比较结果如表1所示,表1给出了50次实验的统计结果,表中RMSE为均方

表 1 各算法区间预测效果比较

Table 1 Comparison of the interval prediction results of each algorithm

算法	PICP	RMSE/($\text{km}^3 \cdot \text{h}^{-1}$)	NMPIW	CWC	预测耗时/s
Bootstrap NN	92.7227	31.1626	0.3912	1.6634	40.7670
Delta NN	89.1945	31.1920	0.4127	1.6956	36.9772
LSSVM	92.2163	30.0976	0.4140	1.6542	11.4048
本文方法	93.1179	29.7217	0.3599	1.4019	12.3595

表1中, PICP表示区间覆盖率, RMSE表示均方根误差, NMPIW表示区间平均宽度, CWC是式(26)中的综合评价指标。

6 结论(Conclusions)

冶金企业副产煤气是重要的二次能源,其利用是否合理直接影响企业的经济效益和节能减排水平,完成副产煤气的产消流量预测是合理利用煤气的前提.针对冶金企业煤气系统的流量预测问题,本文提出一种基于核的在线区间预测构造方法,该方法将传统预测区间构造过程中对雅克比矩阵的复杂计算转化为对核的计算,大大降低了计算成本.为确定所提方法的超参数,本文利用共轭梯度下降

根误差用于衡量预测的精度,如式(25)所示:

$$\text{RMSE} = \frac{1}{n} \sqrt{\sum_{i=1}^n (Y_i - F_i)^2}, \quad (25)$$

其中: n 是预测数据点的个数, Y_i 为实际的观测值, F_i 为模型的预测值.评价预测区间的指标,通常有两个即区间覆盖率PICP和区间平均宽度NMPIW^[8],但是这两个参数是矛盾的, CWC兼顾了区间覆盖率和区间平均宽度两个参数,是一个更广的评价指标:

$$\text{CWC} = \text{NMPIW} \cdot f_{\text{PICP}}, \quad (26)$$

$$f_{\text{PICP}} = (1 + \gamma(\text{PICP}) \exp(-\eta(\text{PICP} - \mu))), \quad (27)$$

其中:

$$\gamma = \begin{cases} 0, & \text{PICP} \geq \mu, \\ 1, & \text{PICP} < \mu, \end{cases}$$

η 和 μ 是控制CWC跳变点以及跳变幅度的两个参数,本文取 $\eta = 50$, $\mu = 0.95$ 。

表1中的结果说明了本文所提方法的有效性,无论是从预测精度还是区间效果,本文方法都要优于其他3种方法的结果,在计算时间方面,本文方法相比于LSSVM略微耗时,但要明显优于神经网络方法.从预测精度,区间效果和计算时效性3个方面考虑,无疑本文方法是最好的,适合为冶金煤气流量构造预测区间。

算法来优化模型预测误差,使其尽可能逼近样本数据中有效噪声的方差.本文所提的核方法是建立在两个不同核函数的基础上,不同的是核函数的宽度,可以控制函数的径向函数范围,核函数的参数不同会隐式的改变输入空间到高维空间的映射关系,进而改善回归模型的性能.文中最后通过对现场实际数据的仿真,说明了本文方法的有效性。

参考文献(References):

- [1] 王伟, 吴敏, 雷琪, 等. 炼焦生产过程综合生产指标的改进神经网络预测方法 [J]. 控制理论与应用, 2009, 26(12): 1419 - 1424.
(WANG Wei, WU Min, LEI Qi, et al. An improved neural network method for the prediction of comprehensive production indices in

- coking process [J]. *Control Theory & Applications*, 2009, 26(12): 1419 – 1424.)
- [2] 张晓平, 赵珺, 王伟, 等. 转炉煤气柜位的多输出最小二乘支持向量机预测 [J]. *控制理论与应用*, 2010, 27(11): 1463 – 1470. (ZHANG Xiaoping, ZHAO Jun, WANG Wei. Multi-output least squares support-vector-machine for level prediction in Linz Donaniz gas holder [J]. *Control Theory & Applications*, 2010, 27(11): 1463 – 1470.)
- [3] 张颜颜, 唐立新. 改进的数据驱动子空间算法求解钢铁企业能源预测问题 [J]. *控制理论与应用*, 2012, 29(12): 1616 – 1622. (ZHANG Yanyan, TANG Lixin. Improved data-driven subspace algorithm for energy prediction in iron and steel industry [J]. *Control Theory & Applications*, 2012, 29(12): 1616 – 1622.)
- [4] 肖冬峰, 杨春节, 宋执环. 基于改进BP网络的高炉煤气发生量预测模型 [J]. *浙江大学学报(工学版)*, 2012, 46(11): 2103 – 2108. (XIAO Dongfeng, YANG Chunjie, SONG Zhihuan. The forecasting model of blast furnace gas output based on improved BP network [J]. *Journal of Zhejiang University (Engineering Science)*, 2012, 46(11): 2103 – 2108.)
- [5] 张琦, 谷延良, 提威, 等. 钢铁企业高炉煤气供需预测模型及应用 [J]. *东北大学学报(自然科学版)*, 2012, 31(12): 1731 – 1740. (ZHANG Qi, GU Yanliang, TI Wei, et al. Supply-demand forecasting model of blast furnace gas in iron and steel works and its application [J]. *Journal of Northeastern University (Natural Science)*, 2012, 31(12): 1731 – 1740.)
- [6] SHENG C Y, ZHAO J, LIU Y. Prediction for noisy nonlinear time series by echo state network based on dual estimation [J]. *Neurocomputing*, 2012, 82: 186 – 195.
- [7] PAPADOPOULOS G, EDWARDS P J, MURRAY F. Confidence estimation methods for neural networks: a practical comparison [J]. *IEEE Transactions on Neural Networks*, 2001, 12(6): 1278 – 1287.
- [8] KHOSRAVI A, NAHAVANDI S, CREIGHTON D. Comprehensive review of neural network-based prediction intervals and new advances [J]. *IEEE Transactions on Neural Networks*, 2011, 22(9): 1341 – 1356.
- [9] BRABANTER K D, BRABANTER J D, SUYKENS A K. Approximate confidence and prediction intervals for least squares support vector regression [J]. *IEEE Transactions on Neural Networks*, 2011, 22(1): 110 – 121.
- [10] HEKKES T. Practical confidence and prediction intervals [M] // *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 1997, 9: 176 – 182.
- [11] BISHOP C M. *Pattern Recognition and Machine Learning* [M]. New York: Springer Press, 2006.
- [12] RASMUSSEN C E, WILLIAMS C K. *Gaussian Processes for Machine Learning* [M]. Cambridge: MIT Press, 2006.
- [13] GOLDBERG P W, WILLIAMS C K, BISHOP C M. Regression with input-dependent noise: a Gaussian process treatment [J]. *Advances in Neural Information Processing Systems*, 1998, 10: 493 – 499.
- [14] GIRARD A, MURRAY R. Gaussian process: prediction at a noisy input and application to iterative multiple-step ahead forecasting of time-series [R] // *Lecture Notes in Computer Science*. Heidelberg: Springer-Verlag, 2005, 3355: 158 – 184.
- [15] AN S, LIU W, VENKATESH S. Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression [J]. *Pattern Recognition*, 2007, 40(8): 2154 – 2162.

作者简介:

赵珺 (1981–), 男, 博士, 讲师, 主要研究方向为生产计划与调度、现代集成制造系统、工业生产一体化优化技术, E-mail: zhaoj@dlut.edu.cn;

杜雅楠 (1988–), 男, 硕士研究生, 主要研究方向为工业生产计划与优化调度, E-mail: duyuanan159@126.com;

盛春阳 (1986–), 男, 博士研究生, 主要研究方向为流程工业建模与优化调度;

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.