0254-6124/2023/43(4)-0703-08 *Chin. J. Space Sci.* 空间科学学报

REN Junzhu, XIAO Zhigang, YU Qiang. Modeling of Temperature Control System of Space Experiment High-temperature Furnace Based on XGBoost (in Chinese). *Chinese Journal of Space Science*, 2023, **43**(4): 703–710. DOI:10.11728/cjss2023.04.2022–0061

基于 XGBoost 的空间高温材料实验炉控制系统建模*

任俊竹[1,2] 肖志刚[1] 于 强[1]

1(中国科学院国家空间科学中心 北京 100190)

2(中国科学院大学 北京 100049)

摘 要 为确保高温材料科学实验柜科学实验系统能够成功地进行空间材料科学实验，在空间进行高温材料科学实验的时候要求其温度稳定在±0.25℃ 范围内。面对如此之高的温度稳定度要求，提出一个新的解决方案：在实验输入和输出数据的基础上，确定一个与高温炉控制系统内部等价的模型，为获得满足实验要求的控制参数提供依据。本文将高温炉控制系统内部看作黑箱模型，基于 XGBoost 方法分别对四类样品实验的高温炉内部温区 2 和温区 3 进行建模，模型精确度全部可达到 99.98% 以上。与传统建模方式传递函数相对比，在传统方法表现最好的情况下，模型精度仍提高了 3.8%，为获得控制参数以确保空间实验温度实现高稳定度提供了重要支持。

关键词 XGBoost，传递函数，系统辨识，高温炉控制系统

中图分类号 TP3

Modeling of Temperature Control System of Space Experiment High-temperature Furnace Based on XGBoost

REN Junzhu[1,2] XIAO Zhigang[1] YU Qiang[1]

1(*National Space Science Center*, *Chinese Academy of Sciences*, *Beijing* 100190)

2(*University of Chinese Academy of Sciences*, *Beijing* 100049)

Abstract With the development of China's space industry, the construction of China's space station has been completed in 2022. In the future, China will carry out a series of space material science experiments in space. The high-temperature furnace in the high-temperature material science experimental rack, as the main equipment of the space material science experiment, requires the high-temperature fur-

* 中国科学院空间科学先导专项科学卫星任务运控技术项目资助（XDA15040100）

2022-10-24 收到原稿，2023-02-07 收到修定稿

E-mail：renjunzhu20@mails.ucas.ac.cn. 通信作者 肖志刚，E-mail：xiaozhigang@nssc.ac.cn

©The Author(s) 2023. This is an open access article under the CC-BY 4.0 License (https://creativecommons.org/licenses/by/4.0/)

nace's temperature to be stable within ± 0.25℃ when conducting the high temperature material science experiment in space. In the face of such high temperature stability requirements, in order to ensure that the scientific experimental system of the high-temperature material science rack can successfully conduct the space material science experiment, it is necessary to first establish the mathematical model of the high-temperature furnace control system. Because the object of high-temperature furnace is a kind of nonlinear and time-delay complex control object, it is difficult to model based on mechanism. To solve this problem, this paper proposes a new solution: based on the experimental input and output data, an intelligent modeling method is adopted to determine an internal equivalent model of the high-temperature furnace control system, which provides a basis for obtaining control parameters that meet the experimental requirements. In this paper, the control system of high-temperature furnace is regarded as a black box model, and four representative sample experimental data are selected. Based on XGBoost method, the mathematical models of temperature zone 2 and temperature zone 3 control system of high-temperature furnace are established respectively. The accuracy of the models can all reach more than 99.98%. Compared with the traditional modeling method, the transfer function is used as the basic model for parameter estimation, and the modeling effect varies according to different samples. In addition, under the best performance of traditional methods, the accuracy of the model based on XGBoost is still improved by 3.8%. The experimental results show that the modeling effect of high-temperature furnace control system based on XGBoost method is good, and the model provides important support for obtaining control parameters to ensure high stability of space experimental temperature.

Key words XGBoost, Transfer function, System identification, High-temperature furnace

0 引言

随着中国载人航天工程的快速发展,空间材料科学逐步发展成为中国材料科学中最活跃的学科之一[1]。2022 年中国载人空间站完成建造,将支持开展大规模多学科的空间科学研究、技术验证和空间应用。空间站的重要应用之一是支持空间材料科学研究。高温材料科学实验柜(以下简称高温柜)作为空间材料科学实验的承载装置,其性能的优劣决定着空间材料科学实验的成败。

温度是晶体生长过程的重要控制参数之一,只有温度始终维持在实验设定的范围内,才能够确保样品完全熔化和晶体生长状态良好。例如,根据空间材料实验的综合要求,高温炉在保温时的温度稳定度要求控制在±0.25℃ 以内。为了获得实验过程中的最优控制参数,需要对高温炉的控制系统进行系统建模,在输入和输出数据的基础上,构建一个与所测系统等价的模型[2],从而确保高温炉温度的高稳定度。

理想化的简单系统可以使用白箱建模对系统进行直接辨识,通过物理公式对系统进行推导;复杂系统可以使用黑箱法进行辨识,只关注系统输入输出找到等价的模型;将白箱法和黑箱法结合得到灰箱法,给定系统模型,进行模型参数辨识[3]。文献 [4] 根据热平衡原理对空间生长炉进行了机理建模,并利用相关系数法和最小二乘法结合,进行了生长炉模型辨识。文献 [5] 根据传热学基本原理对三温区的空间晶体生长炉进行了建模,并利用数值方法求解模型参数。文献 [6] 为了满足先进控制系统设计需求,针对板翅式换热器动态模型,通过列写简化的机理方程,提出采用传递函数动态模型进行系统辨识。文献 [7] 通过推导电子节气门体的转矩传递过程,建立系统动力学方程,创建了状态空间模型,并对模型进行参数辨识。

高温柜中的高温炉控制系统作为控温的主要设备,由于其结构模块化、复杂化、多元化,采用传统方法基于机理直接对被控对象温度建立准确的数学模型难度较大,效果较差。本文在进行系统建模时,将高温炉控制系统内部传热过程看作黑箱模型,基于数据驱动采用极致梯度提升(eXtreme Gradient Boost-

ing, XGBoost)算法对高温炉控制系统进行系统建模分析, 根据输出的预测值与实际值的差异来评估系统辨识模型的准确度。对比使用传递函数方法进行系统参数辨识的结果, 该模型精确度更高。依据此模型进行地面匹配实验, 可提高材料生长的温度稳定度, 为空间材料科学实验的成功奠定技术基础。

1 研究方法

机器学习中通过采用某种策略将多个弱学习器组合成强学习器, 通常会获得比单一弱学习器显著优越的性能。梯度提升决策树(Gradient Boosting Decision Tree, GBDT)是一种基于集成思想的加法模型, 训练时采用贪婪策略进行基于前向分布算法的学习, 每次迭代都学习一棵分类回归树(Classification and Regression Tree, CART)来拟合之前 $t-1$棵树的预测结果与训练样本真实值的残差。XGBoost[8]是基于 GBDT 的一种算法, 在 GBDT 的基础上进行了改进, 采用二级泰勒公式展开损失函数, 加入代表模型复杂度的正则化项[9], 并利用启发式搜索的方法确定每一棵回归树的最佳切分点, 算法中每棵回归拟合树拟合的都是上一棵回归树的残差值, 最后把所有的回归树相加便得到最终的模型[10]。

采用 XGBoost 方法则将系统完全视为黑箱模型, 只关注系统输入和输出, 输出为系统某温区炉温, 输入为经过特征处理后得到的与炉温相关的 m 个特征参数, 据此定义温度输出的数据模型如下:

$$\widehat{y}_i=\sum_{k=1}^{K}f_k\left(x_i\right),\quad f_k\in\mathcal{F}. \tag{1}$$

其中, i 表示第i个样本, K表示迭代训练过程中的 CART 树的总棵数, f_k表示要训练的模型, 是在一个函数空间$\mathcal{F}$里面的函数; $\widehat{y}_i$表示最终预测的第 i 个样本的炉温, x_i表示第 i 个样本 m 个特征的值。

为了获取最佳温度模型, 设计的目标函数包含损失函数以及正则项两部分, 即

$$O^{(t)}=\sum_{i=1}^{n}l\left[y_i,\widehat{y}_i^{(t-1)}+f_t\left(x_i\right)\right]+\sum_{k=1}^{t-1}\Omega\left(f_k\right)+\Omega\left(f_t\right). \tag{2}$$

式中, n为样本数量, i为样本序号, l表示损失函数, y_i为温度实际值, $\widehat{y}_i^{(t-1)}$为第$t-1$次迭代输出的预测值, f_t为第t次迭代的树结构和叶子结点的权重函数, x_i为输入, $\sum_{k=1}^{t-1}\Omega\left(f_k\right)$为前$t-1$次迭代的正则化项, $\Omega\left(f_t\right)$为第t次迭代的正则化项。

假设第t棵树有T个叶子结点, 叶子结点的输出用$[w_1,w_2,\ldots,w_T]$表示, 定义正则项如下:

$$\Omega\left(f_t\right)=\gamma T+\frac{1}{2}\lambda\sum_{j=1}^{T}w_j^2. \tag{3}$$

其中, T 为叶子结点个数, w_j为叶子结点j 的输出, γ为每增加 1 个叶子结点的复杂度, λ为惩罚系数, 可有效防止过拟合。

优化目标是降低损失函数与模型复杂度, 即寻求目标函数的最小值。使用泰勒公式对目标函数中的损失函数进行二阶展开, 由于前$t-1$棵树已知, 所以前$t-1$棵树的损失函数$L^{(t-1)}$和复杂度$\Omega^{(t-1)}$也已知, 可将其看为常数。假设$q\left(x\right):R^d\to\{1,2,3,\ldots,T\}$表示样本到叶子结点的映射, 样本全部映射到叶子结点上, 定义$I_j=\{i|q\left(x_i\right)=j\}$为叶子结点的实例集, 令$G_j=\sum_{i\in I_j}g_i$, $H_j=\sum_{i\in I_j}h_i$, 则最终目标函数可化简为

$$O^{(t)}=\sum_{j=1}^{T}\left[G_jw_j+\frac{1}{2}\left(H_j+\lambda\right)w_j^2\right]+\gamma T. \tag{4}$$

该函数是一个关于w_j的累加独立二次函数, 对w_j 求偏导, 并令偏导为 0, 即由

$$\frac{\partial O^{(t)}}{\partial w_j}=G_j+\left(H_j+\lambda\right)w_j=0. \tag{5}$$

可以得到最优参数$w_j^*=-\dfrac{G_j}{H_j+\lambda}$, 目标函数最小值$O_{\min}^{(t)}=-\dfrac{1}{2}\sum_{j=1}^{T}\dfrac{G_j^2}{H_j+\lambda}+\gamma T$。

将$O_{\min}^{(t)}$用作衡量结点是否需要进行分裂的评分函数, 该评分函数类似于评估决策树划分选择的信息增益评分, 不过其是针对更广泛的目标函数推导的。通常情况下, 不可能枚举所有树的结构, 取而代之的是, 使用单个叶子结点从树深为 0 开始并迭代地向树添加分支的贪婪算法。假设I_{L}和I_{R}是分割后左右结点的实例集, $I=I_{\mathrm{L}}\cup I_{\mathrm{R}}$, 那么分割后的损失减少$G_{\mathrm{ain}}$可以表示为分裂前的评分减去分裂后的评分, 即

$$G_{ain}=\frac{1}{2}\left[\frac{\left(\sum_{i\in I_{\mathrm{L}}}g_i\right)^2}{\sum_{i\in I_{\mathrm{L}}}h_i+\lambda}+\frac{\left(\sum_{i\in I_{\mathrm{R}}}g_i\right)^2}{\sum_{i\in I_{\mathrm{R}}}h_i+\lambda}-\frac{\left(\sum_{i\in I}g_i\right)^2}{\sum_{i\in I}h_i+\lambda}\right]-\gamma=\frac{1}{2}\left[\frac{G_{\mathrm{L}}^2}{H_{\mathrm{L}}+\lambda}+\frac{G_{\mathrm{R}}^2}{H_{\mathrm{R}}+\lambda}-\frac{(G_{\mathrm{L}}+G_{\mathrm{R}})^2}{H_{\mathrm{L}}+H_{\mathrm{R}}+\lambda}\right]-\gamma. \tag{6}$$

其中, $G_{\mathrm{L}}=\left(\sum_{i\in I_{\mathrm{L}}}g_i\right)$, $G_{\mathrm{R}}=\left(\sum_{i\in I_{\mathrm{R}}}g_i\right)$, $G_{\mathrm{L}}+G_{\mathrm{R}}=\sum_{i\in I}g_i$, $H_{\mathrm{L}}=\left(\sum_{i\in I_{\mathrm{L}}}h_i\right)$, $H_{\mathrm{R}}=\left(\sum_{i\in I_{\mathrm{R}}}h_i\right)$, $H_{\mathrm{L}}+H_{\mathrm{R}}=\sum_{i\in I}h_i$; γ为加入新叶子结点引入的复杂度代价。G_{ain}越大说明分裂该结点后目标函数减少越多, 故每次选取G_{ain}值最大的结点进行分裂。当G_{ain}值大于 0 或者未达到系统设置的树的最大深度时, 叶子结点持续分裂; 当G_{ain}值小于 0 或者达到系统设置的树的最大深度时, 叶子结点停止分裂, 获得权重最优解和预测温度最优值, 此时的模型是与系统最符合的模型。

该算法的核心思想就是通过不断进行特征分裂来添加新的树, 每次添加后都需要学习新的函数来拟合上次预测的残差, 最终训练完成后得到 t 棵树。要预测一个样本的分数, 就是将这个样本的特征落到每一棵树的叶子结点上, 得到每个叶子结点对应的分数, 最后将每棵树对应的分数加起来就是该样本的预测值。相比较传统的 GBDT, XGBoost 采用泰勒二级展开, 有利于梯度更快更准地下降; 加入正则化项控制模型的复杂度, 更好地防止过拟合, 提高了模型的泛化能力。

2 系统建模

2.1 实验设备与数据

2.1.1 实验设备

整个高温柜分为实验柜支撑系统和科学实验系统两部分, 其中科学实验系统包括 4 个模块, 分别是实验电控模块、批量样品管理模块、高温炉模块, 以及 X 射线实时观察模块, 各个模块的位置如图 1 所示[11]。

实验电控模块用于控制实验系统的实验过程, 包括控制高温炉模块的温度, 批量样品模块的样品更换和样品提拉过程, 并且控制 X 射线实时观察模块中的加热过程; 同时该模块还接收实验柜的控制指令, 将实验数据传输给实验柜。高温炉加热体(炉膛)由 4 个可控制加热的加热区(温区)组成, 每个温区由一组炉丝控制加热, 一支热电偶进行测温, 其测出的温度还用于反馈控制加热过程。加热炉膛、加热温区和测控温热电偶的简化结构如图 2 所示。

2.1.2 实验数据

空间材料科学实验开展之前需要在地面进行大

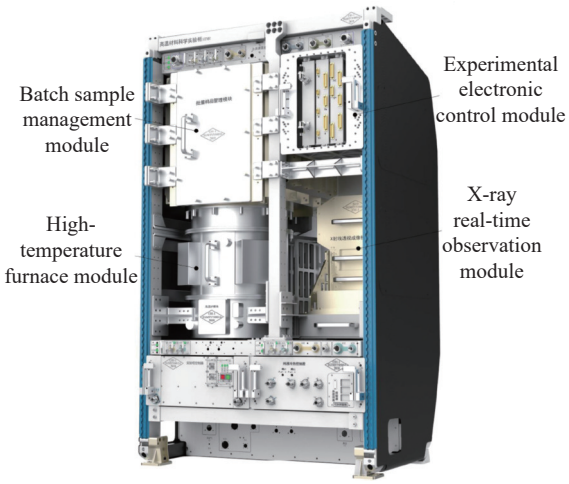

图 1 高温材料科学实验柜内部结构

Fig. 1 Internal structure of high-temperature material science experiment rack

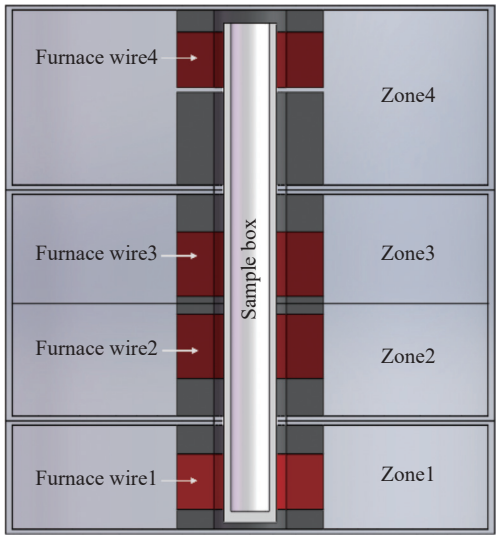

图 2 高温炉加热炉体简化结构

Fig. 2 Simplified structure of high-temperature furnace

量的匹配实验, 以保证实验材料能够在高温柜中按照设定的实验工艺进行加工, 并得到期望的实验结果, 为空间材料科学实验的理论研究和实验方法提供数据支撑, 从而提高空间材料科学实验的成功率[12]。

选用高温柜多次地面匹配实验中具有代表性的两种非金属材料实验、两种金属材料实验进行建模。平均每次实验的时长超过 14 h, 实验数据的采集频率为每秒钟采集两条数据, 每次采集到的一条数据包括高温炉模块、实验电控模块、样品盒模块、X 射线模块等多个设备模块的工程采集参数, 共 462 个参数, 产生的数据量巨大。对于高温炉的温控系统来说, 其实验过程要求的热特性决定了其温度、电流电压等连续数值型数据, 理论上变化比较平稳, 但仍不能排除噪音干扰的存在, 采用均值滤波可帮助减少数据噪音干扰, 尽量逼近真实数据值。本文首先根据通信协议对源码数据文件进行解析, 之后对解码数据进行均值数据滤波处理, 将数据处理为一秒钟一条(见表 1)。

由于目前只对两个温区进行了实验, 因此建模时也只考虑两个温区。模型的输出向量为单个温区的温度, 输入向量由特征选择决定。

2.2 特征选择

原数据集中含有大量特征, 这些特征与本文的建模输出对象——高温炉炉温相关程度可大致分为两类, 一类是相关程度几乎为 0 的特征参数, 诸如通信类的特征参数等; 另一类则是相关程度相对较高的参数。两类特征区分度较大, 故本文使用 Pearson 相关系数[13] 度量特征之间的相关程度。Pearson 相关系数是两个变量之间的协方差和标准差的商, 公式如下:

$$\rho_{x,y}=\frac{\operatorname{cov}(x,y)}{\sigma_x\sigma_y}=\frac{E\left[(x-\mu_x)(y-\mu_y)\right]}{\sigma_x\sigma_y}. \tag{7}$$

式中, σ_x为x的标准差, μ_x为x的均值, σ_y为y的标准差, μ_y为y的均值。

Pearson 相关系数的值均在 [–1, 1] 之间, 值接近于 0 表示两者相关性不强, 值接近于 1 表示两者呈正相关, 值接近于–1 表示两者呈负相关, 绝对值大于 0.8 为强相关。挑选出对于单个温区温度而言相关性高于 0.8 的特征, 从而将 461 维特征减少至 33 维特征作为模型输入特征向量, 大大减少模型的计算量。部分输入特征列于表 2。

2.3 模型设置

模型的参数设置会影响模型的训练速度和模型的准确度, 参数选取不合适也会出现欠拟合或过拟合的情况。模型的目标函数由损失函数和正则化项构成, 其中损失函数设置为均方差(Mean Square Error, MSE), 正则化项 Lambda 设置为固定值 1。除了设置正则化项减少过拟合之外, 还需要设置一些边界条件, 例如树的最大深度、树的棵数(迭代次数)、每棵树的权重(学习率), 防止模型进行无限次训练[14]。

表 1 数据滤波后实验数据基本信息

Table 1 Basic information of experimental data after data filtering

实验名称	实验时长/h	数据条数	数据维度	温区2工况/℃	温区3工况/℃
非金属样品1号实验	30.25	108900	462	750～800	860～900
非金属样品2号实验	15.57	56040	462	600～800	600～750
金属样品1号实验	14.05	50580	462	900	900
金属样品2号实验	16.95	61020	462	1050～900	1050～900

表 2 部分输入特征

Table 2 A portion of input feature vector

序号	特征名称	相关系数	序号	特征名称	相关系数
1	目标温度	0.9971	6	炉丝电源温度	0.9298
2	计算上限值	0.9942	7	高温炉冷端	0.9050
3	输出占空比	0.9537	8	炉丝霍尔电流	0.8995
4	加热电压	0.9479	9	主控板28 V温度	0.8791
5	100 V霍尔电流	0.9449	10	高温炉壳温	0.8575

并且这三个参数的设置很大程度上决定了模型的训练速度、训练效果，通常需要进行优化。

目前优化超参数的常用策略是网格搜索和随机优化。其中随机优化方法由于搜索过程是随机的，容易错失最优解。本文选用网格搜索方法进行参数优化，该算法是在所有候选的参数选择中，通过循环遍历，尝试每一种可能性，通过 5 折交叉验证找到在验证集上精度最高的参数，表现最好的参数就是最终的结果。模型最优参数列于表 3。

2.4 评价指标

模型采用E_{mse}进行误差测量，E_{mse}指标计算的是预测值与真实值对应样本点的误差平方和的均值，其值越小说明模型效果越准确，具体计算公式如下：

$$E_{\mathrm{mse}}=\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i-y_i)^2. \tag{8}$$

式中，y_i为实际测量所得的值，$\hat{y}_i$为预测的值，n为样本总数。

E_{mse}只考虑了预测数值的正确性，未考虑模型本身是否能够捕捉到数据的规律。如果前半部分数据拟合较好，但后半部分数据拟合较差，误差经过平均之后，只用以上提到的指标可能难以发现模型的缺陷。

为能够更好地反映系统建模的整体效果，这里还选用决定系数R^2[15]这一指标评估模型的准确度。R^2也称判定系数或者拟合优度。其表征回归方程在多大程度上解释了因变量的变化，或者说方程对观测值的拟合程度。

设y_i为实际测量值，其均值为$\bar{y}$；$\hat{y}_i$为预测值，记总平方和$E_{\mathrm{sst}}=\sum_{i=1}^{n}(y_i-\bar{y})^2$，残差平方和$E_{\mathrm{sse}}=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$，决定系数

$$R^2=1-\frac{E_{\mathrm{sse}}}{E_{\mathrm{sst}}}=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}. \tag{9}$$

决定系数R^2既考虑了预测值与真实值的差异，也兼顾了真实值的离散程度。通常R^2为负表示模型不可信，故一般R^2的取值范围是 [0, 1]，值越大表示模型辨识的准确度越高。

表 3 网格搜索参数优化

Table 3 Optimization of grid search parameters

参数名	寻优范围	取值
迭代次数	[50, 65, 75, 80, 100, 150, 200, 350]	150
学习率	[0.01, 0.03, 0.05, 0.08, 0.1, 0.2, 0.25, 0.3]	0.2
树的最大深度	[1, 3, 5, 8]	5

3 试验结果分析

3.1 传统方法试验

选择具备时间顺序的炉丝功率作为建模时的输入变量 x，选择炉温作为建模时的输出变量y，采样间隔为 1 s，系统阶数为 2，利用 MATLAB 系统辨识工具箱进行传统方法系统辨识。采用传递函数方法辨识得到的结果列于表 4。

由传递函数的系统辨识结果可以看出，采用传递函数作为基础模型进行参数估计，结果是十分不稳定的，模型准确性在 0.11～0.96 之间波动，传递函数建模效果较差。对比温区 2 和温区 3 的结果可知，2 阶传递函数模型在温区 3 上表现更好，4 种样品的R^2均大于 0.45，而温区 2 的R^2则全部小于 0.45。在两个温区上金属样品 1 号的结果最接近，非金属样品 2 号差异最大，这可能是由于样品伸入温区内的体积大小以及保温时长不同造成的。

3.2 XGBoost 方法试验

表 5 给出了采用 XGBoost 方法系统建模得到的结果。根据 XGBoost 方法的系统辨识结果可以看出，采用 XGBoost 进行黑箱建模相比传统方法进行参数估计，效果提高十分明显。误差E_{mse}较小，模型预测准确，模型精确度R^2可以达到 0.99999，说明模型整体对数据规律的捕捉也较好。对比金属样品和非金属样品模型预测准确度，可以看出采用 XGBoost 方法预测的非金属材料模型准确度略高于金属材料模型准确度，这与实际实验中金属样品温度控制难度略

表 4 传递函数参数估计结果

Table 4 Estimation results of transfer function parameters

传递函数	温区2		温区3	
	R^2	E_{mse}	R^2	E_{mse}
非金属样品1号	0.1105	5.45×10^4	0.5386	4.00×10^4
非金属样品2号	—	3.28×10^4	0.9613	952.5176
金属样品1号	0.4114	3.56×10^3	0.4659	3.15×10^3
金属样品2号	0.4369	2.94×10^3	0.4982	2.51×10^3

大于非金属样品控制难度相符。并且模型稳定性良好, 对于不同种样品的预测效果都很好。

3.3 结果比对分析

表 6 列出了 XGBoost 方法与传统建模方法获得的模型对比情况, 由于R^2评价指标越大越好, 所以定义d_{R^2}表示的是 XGBoost 模型R^2与传递函数模型R^2的差, 公式如下:

$$d_{R^2} = R^2_{\text{XGBoost}} - R^2_{\text{tf}}. \tag{10}$$

其中, R^2_{XGBoost}表示 XGBoost 模型的R^2, R^2_{tf}表示传递函数模型的R^2。而E_{mse}评价指标越小越好, 所以定义$d_{E_{\text{mse}}}$表示传递函数模型E_{mse}与 XGBoost 模型E_{mse}的差, 公式如下:

$$d_{E_{\text{mse}}} = E_{\text{mse_tf}} - E_{\text{mse_XGBoost}}. \tag{11}$$

其中, $E_{\text{mse_XGBoost}}$表示 XGBoost 模型的E_{mse}, $E_{\text{mse_tf}}$表示传递函数方法的E_{mse}。

根据表 6 可以得知, 在R^2评价指标上, XGBoost 与传递函数相比, 最小提升了 0.039, 最大提升了 0.889。这主要是因为 R^2评价指标不仅评估预测数据与真实数据的误差, 还评估了建模的整体效果, 因此传统方法在R^2这项指标上的表现波动较大。对比金属材料与非金属材料模型准确度, 非金属材料的准确度更高, 这与控温系统中金属材料控温比非金属材料控温难度大的实际情况相符。传统传递函数方法对系统辨识的结果较差, 这是由于样品的特性对控制系统阶数识别造成了很大的影响, 并且在不同温区、不同样品上表现出极不稳定的效果。XGBoost 方法表现最好, 模型稳定且准确。所建模型可用于寻找系统 PID 控制参数时对实验温度进行仿真, 根据预测值与目标温度的误差, 通过 PID 控制器, 修正部分输入向量的值, 循环多次获得预测实验温度曲线。仿真如图 3 所示。

表 5 XGBoost 建模结果

Table 5 XGBoost modeling results

XGBoost	温区2		温区3	
	R^2	E_{mse}	R^2	E_{mse}
非金属样品1号	0.999999	0.064715	0.999999	0.096201
非金属样品2号	0.999999	0.077839	0.999999	0.077566
金属样品1号	0.999998	1.56×10^{-1}	0.999998	1.57×10^{-1}
金属样品2号	0.999998	2.48×10^{-1}	0.999998	2.33×10^{-1}

4 结论

基于机理和传统方法对高温柜高温炉控制系统进行系统建模需要掌握部分系统机理, 由于高温炉内部系统复杂, 难以精确地用机理进行描述, 据此建立的模型精确度也较低, 因此提出基于 XGBoost 的高温柜高温炉控制系统系统辨识方法。本文选取地面匹配实验中的两次非金属材料实验、两次金属材料实

表 6 XGBoost 模型较传统模型的性能比较

Table 6 XGBoost model performance improvemen compared with traditional models

实验名称		d_{R^2}	$d_{E_{\text{mse}}}$
非金属样品1号	温区2	0.889	5.4500×10^{4}
	温区3	0.461	4.0000×10^{4}
非金属样品2号	温区2	—	3.2799×10^{4}
	温区3	0.039	9.5244×10^{2}
金属样品1号	温区2	0.588	3.5598×10^{3}
	温区3	0.534	3.1500×10^{3}
金属样品2号	温区2	0.563	2.9398×10^{3}
	温区3	0.502	2.5098×10^{3}

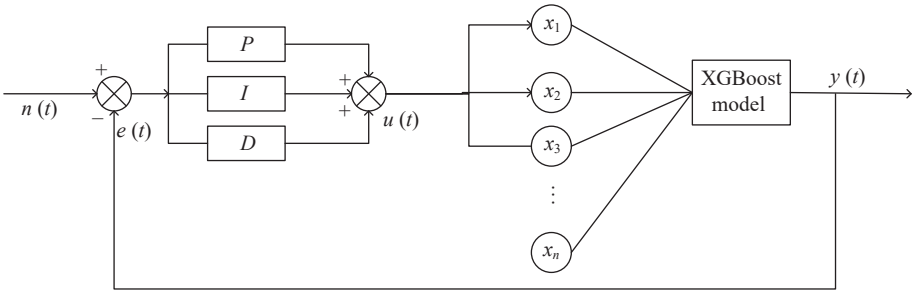

图 3 模型应用

Fig. 3 Model application

验，分别利用 XGBoost 方法对高温柜高温炉的温区 2 和温区 3 进行单输出系统建模，实验的系统识别精度达到 99.98%。相比传递函数这种传统建模方式，即便在传统方法表现最好的情况下，模型评价指标 R^2仍提高了 0.039 左右，并且模型对于不同种样品预测差异较小，表现稳定。实验证明该方法在对高温柜实验数据的预测中具有实际应用价值，未来对不同样品进行实验时，可依据该方法建立的模型建立不同种样品的系统模型，为获得控制参数以确保空间实验温度稳定度达到±0.25℃ 提供依据。

致谢 中国科学院上海硅酸盐研究所和高温材料科学实验柜研制团队以及参加空间材料科学实验的科学家给予了支持。

参考文献

[1] RUAN Ying, HU Liang, XIE Wenjun, *et al.* Recent advances and future perspectives of space materials science[J]. *Scientia Sinica Technologica*, 2020, **50**(6): 603-649 (阮莹, 胡亮, 闫娜, 等. 空间材料科学研究进展与未来趋势[J]. 中国科学: 技术科学, 2020, **50**(6): 603-649)

[2] Wang Qihong, Shi Guodong, Xue Guoxin, *et al.* A Study on the Modeling of Chemical Processes with BP Neural Networks[J]. *Journal of Changzhou University(Natural Science Edition)*, 2000(02): 45-47 (王其红, 史国栋, 薛国新, 张奕. 基于BP神经网络的化工过程建模研究[J]. 江苏石油化工学院学报, 2000(02): 45-47)

[3] LI Yu, YUAN Lei. Overview on system identification for linear systems[J]. *Journal of Detection & Control*, 2021, **43**(3): 22-29 (李昱, 袁磊. 线性系统的系统辨识综述[J]. 探测与控制学报, 2021, **43**(3): 22-29)

[4] YU Qiang, LV Xutao, SHI Chun, *et al.* Modeling of temperature system in space crystal growth furnace[J]. *Journal of System Simulation*, 2008, **20**(13): 3596-3599 (于强, 吕旭涛, 石春, 等. 空间晶体生长炉温度系统建模[J]. 系统仿真学报, 2008, **20**(13): 3596-3599)

[5] WANG Gang. Mechanism Modeling and Control of the Temperature System of Three Zone Furnace for Space Crystal Growth[D]. Hefei: University of Science and Technology of China, 2009 (王刚. 三温区空间晶体生长炉温度系统机理建模与控制[D]. 合肥: 中国科学技术大学, 2009)

[6] LIANG Xingzhuang, HUANG Zhiyuan, AI Fengming, *et al.* Dynamic model by transfer function and parameter determination method of plate fin heat exchanger[J/OL]. *Journal of Beijing University of Aeronautics and Astronautics*, 1-14[2022-10-09]. https://doi.org/10.13700/j.bh.1001-5965.2022.0202 (梁兴壮, 黄志远, 艾凤明, 等. 板翅式换热器传递函数动态模型及参数确定方法[J/OL]. 北京航空航天大学学报, 1-14[2022-10-09]. https://doi.org/10.13700/j.bh.1001-5965.2022.0202)

[7] LU Yang, DENG Xiaoting. Analysis of state space model of electronic throttle system based on nonlinear factors[J]. *Automobile Applied Technology*, 2022, **47**(13): 63-69 (鲁杨, 邓晓亭. 基于非线性因素的电子节气门系统状态空间模型的分析[J]. 汽车实用技术, 2022, **47**(13): 63-69)

[8] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22 nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016: 785-794

[9] DUAN Meiling, PAN Julong. Research on elderly fall detection algorithms based on Lasso-LGB[J]. *Journal of China University of Metrology*, 2021, **32**(1): 67-73,117 (段美玲, 潘巨龙. 基于Lasso-LGB的老人跌倒检测算法研究[J]. 中国计量大学学报, 2021, **32**(1): 67-73,117)

[10] GUO Yinjing, SONG Xianqi, YANG Lei, *et al.* A coal mine underground positioning algorithm based on gradient boost regression tree[J]. *Science Technology and Engineering*, 2019, **19**(8): 138-144 (郭银景, 宋先奇, 杨蕾, 等. 基于梯度提升回归树的井下定位算法[J]. 科学技术与工程, 2019, **19**(8): 138-144)

[11] China Manned Space. China Space Station Science Experiment Resource Manual [EB/OL]. (2019-04-22)[2022-09-05]. http://www.cmse.gov.cn/gfgg/201904/P020200225328481496691.pdf (中国载人航天工程办公室. 中国空间站科学实验资源手册[EB/OL]. (2019-04-22)[2022-09-05]. http://www.cmse.gov.cn/gfgg/201904/P020200225328481496691.pdf)

[12] DAI Chunyuan. Telecontrol Client System for Matching Test of High Temperature Cabinet Client of Space Station[D]. Beijing: National Space Science Center, the Chinese Academy of Sciences, 2021. DOI: 10.27562/d.cnki.gkyyz.2021.000063 (戴春媛. 面向空间站高温柜匹配试验的遥操作客户端系统[D]. 北京: 中国科学院大学(中国科学院国家空间科学中心), 2021. DOI: 10.27562/d.cnki.gkyyz.2021.000063)

[13] TAO Yang, ZHU Xiaojun, YANG Liu. Multi-sensor data fusion based on Pearson coefficient and information entropy[J]. *Journal of Chinese Computer Systems*, 2023, **44**(5): 1075-1080 (陶洋, 祝小钧, 杨柳. 基于皮尔逊相关系数和信息熵的多传感器数据融合[J]. 小型微型计算机系统, 2023, **44**(5): 1075-1080)

[14] XIAO Yu, ZHAO Jianyou, CHIGAN Du, *et al.* A short-term prediction model for taxi speed based on XGBoost[J]. *Journal of Transport Information and Safety*, 2022, **40**(3): 163-170 (肖宇, 赵建有, 叱干都, 等. 基于XGBoost的短时出租车速度预测模型[J]. 交通信息与安全, 2022, **40**(3): 163-170)

[15] XIE Yu. Regression Analysis[M]. Beijing: Social Sciences Academic Press (China), 2010: 59-61 (谢宇. 回归分析[M]. 北京: 社会科学文献出版社, 2010: 59-61)