

DOI: 10.13382/j.jemi.B2003092

随机森林算法在温度分布重建中的应用^{*}

陈敏鑫¹ 刘 石² 孙单勋² 刘兆宇¹

(1. 华北电力大学 能源动力与机械工程学院 北京 102206; 2. 华北电力大学 控制与计算机工程学院 北京 102206)

摘 要:为提高温度分布重建精度,提出了使用随机森林算法对温度测点进行优化布置的新方法。将测点位置作为样本特征,以不同的测点布置方式及其对应的重建误差作为样本数据集。使用样本数据集构建随机森林模型,评估样本特征重要性,根据特征重要性排序实现温度测点的优化布置。设定仿真实验与燃烧实验验证优化布置算法的可行性与有效性。分析实验数据,所提出的优化布置算法相对于原有算法,重建精度提升了20%以上。研究结果表明,随机森林算法在温度分布重建中具有良好的应用价值,并为解决工业实际问题提供了新思路。

关键词:随机森林;温度分布重建;特征重要性;测点优化布置

中图分类号: TN609; TH811 **文献标识码:** A **国家标准学科分类代码:** 510.4030

Application of random forest algorithm in temperature distribution reconstruction

Chen Minxin¹ Liu Shi² Sun Shanxun² Liu Zhaoyu¹

(1. School of Energy Power and Mechanical Engineering, North China Electric Power University, Beijing 102206, China;
2. School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)

Abstract:In order to improve reconstruction accuracy to resolve temperature distribution reconstruction problems, a method for optimal sensor placement based on random forest algorithm is proposed. Denoting different measurement sites as different sample features, a series of different sensor placements and the reconstruction errors which are calculated by these placements constitute a sample dataset. A random forest model is setting up by the sample dataset, feature importance is also evaluated, then the optimal sensor placement is determined by feature importance. Simulation test and combustion test are set up to verify the feasibility and practicability of the proposed method. Testing data shows that comparing the original method, the proposed method can improve the reconstruction accuracy by at least 20%. Research results indicate that the proposed method has a good practical value, it also provides a new probe of using random forest algorithm to solve industrial problems.

Keywords:random forest; temperature field reconstruction; feature importance; optimal sensor placement

0 引 言

在热力系统特别是工业燃烧装置中,温度作为测量控制的关键参数,表征了热能传输、转化和利用过程。快速准确的获得温度分布信息,对于提升设备运行稳定性、安全性以及运行效率具有深远的意义。

常用的温度信息获取方式大致分为测量法和数值计算法。但在工业应用中,两种方法都有各自的局限性。

受到测量环境、设备安装以及系统造价等因素的影响,测量法难以获得整体的温度分布信息;数值计算方法能够对计算域的整体温度分布进行求解,但数值计算方法往往伴随着建模困难、计算量巨大、计算时间冗长等问题,无法应用于工业实时监测^[1]。

利用机器学习中的主成分分析算法(principal component analysis, PCA)进行温度分布分析及重建^[2-8],获得了良好的应用效果,也成为了近些年来的研究热点。基于主成分分析的温度分布重建算法将测量法和数值计

算法结合,利用特征向量和测量数据能够实时获得温度分布数据。

尽管基于主成分分析的温度分布重建算法能够通过少量测量数据实时获得温度分布信息,但由于测量数据量较小,不同的测点布置方式对重建精度影响较大。为解决这一问题,Willcon 等将求取最小条件数作为目标函数,应用贪心算法获得测点优化布置,并对这一算法进行改进,提升了重建结果的准确性^[9-10]。贪心算法只能获得局部最优解,如何进一步通过优化测点布置方式降低重建误差成为了新的研究方向。

随机森林^[11]是一种集成学习算法,通过构建不同的决策树,每一棵决策树对样本数据集进行学习 and 分类,汇总各决策树结果最终实现回归或分类计算目标。由于具有实现简单、鲁棒性强等优点,随机森林算法被广泛应用于各个领域^[12-17]。

综上所述,提出基于随机森林的测点优化布置算法。从数据分析角度解决测点布置问题,挖掘测点位置与温度分布重建精度的关系。构建基于随机森林算法的分类模型,使用样本数据集对分类模型参数进行训练,分析测点位置在温度分布重建算法中的重要性,实现测点优化布置。与原有算法对比,基于随机森林的测点优化布置算法从原理上避免了局部最优解的出现并提升了重建算法的准确性。

1 温度分布重建算法

1.1 主成分分析算法

作为机器学习中的重要降维工具,主成分分析算法被广泛应用于数据降噪和数据压缩中。主成分分析算法能够从高维数据中提取少量线性独立的综合特性表征原有高维空间模型的特性。

利用主成分分析算法对数据降维的实质就是将原有高维度数据转化为低维数据,并最大化的保留原有数据的特性。从空间角度分析就是将全部高维数据投影至特定的超平面,并保证投影后数据散度最大。

将原有高维数据记做 $\mathbf{T}_{n \times m} \{t_1, t_2, t_3, \dots, t_m\}$, 降维后低维数据记做 $\mathbf{X}_{l \times m} \{x_1, x_2, x_3, \dots, x_m\}$ ($l < n$), 转换矩阵以 $\mathbf{W}_{n \times l} \{w_1, w_2, w_3, \dots, w_l\}$ 表示, 降维计算过程表述为:

$$\mathbf{X} = \mathbf{W}^T \mathbf{T} \quad (1)$$

要保证投影后的数据散度最大,需求取 $\mathbf{X}$ 的方差最大值。由于方差的大小由 $\mathbf{W}$ 的方向决定,与 $\mathbf{W}$ 的大小无关,此外,为了避免 $\mathbf{W}$ 出现向量重叠的问题, $\mathbf{W}$ 内各向量线性无关,设定约束条件 $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ 。总结可得目标函数为:

$$\begin{cases} \operatorname{argmax} \operatorname{tr}(\mathbf{W}^T \mathbf{S} \mathbf{W}) \\ \text{s. t. } \mathbf{W}^T \mathbf{W} = \mathbf{I} \end{cases} \quad (2)$$

式中: $\mathbf{S}$ 为 $\mathbf{T}$ 的协方差矩阵。

由拉格朗日乘数法可得:

$$J(\mathbf{W}) = \mathbf{W}^T \mathbf{S} \mathbf{W} - \lambda (\mathbf{W}^T \mathbf{W} - \mathbf{I}) \quad (3)$$

对 $\mathbf{W}$ 求导可得:

$$\mathbf{S} \mathbf{W} = \lambda \mathbf{W} \quad (4)$$

由特征值和特征向量的定义可知, λ 和 $\mathbf{W}$ 分别由原始高维数据协方差矩阵 $\mathbf{S}$ 的特征值和特征向量构成。将 λ 按照降序排列,前 1 个特征值对应的特征向量构成 $\mathbf{W}$, 通过式(1)将原始数据从 n 维降至 1 维,即实现对高维数据的降维计算。

1.2 温度分布重建算法

降维后的低维数据集记为 $\mathbf{B}_{l \times m} \{b_1, b_2, b_3, \dots, b_l, \dots, b_m\}$ ($l < n$), 式(1)改写为:

$$(b_1, b_2, b_3, \dots, b_l, \dots, b_m) = \mathbf{W}^T (t_1, t_2, t_3, \dots, t_l, \dots, t_m) \quad (5)$$

结合约束条件 $\mathbf{W}^T \mathbf{W} = \mathbf{I}$, 任意温度分布可以表示为:

$$t_i = \mathbf{W} \mathbf{b}_i \quad (6)$$

在重建计算中, t_i 为重建温度分布, $\mathbf{W}$ 可以通过仿真计算结果或者历史数据构成的经验数据集提取获得。当系数向量 $\mathbf{b}_i$ 确定时,即可计算出温度分布重建结果,因此,温度分布重建问题转化成为系数向量 $\mathbf{b}_i$ 求解的问题。

通过测量获得少量温度数据,这一过程的数学表达记做:

$$t_M = \mathbf{M} t \quad (7)$$

式中: t_M 为温度测量值; t 为待重建温度分布; $\mathbf{M}$ 为测量矩阵,表征了温度测点的位置信息。重建目标为使重建值 t_i 尽可能的接近 t , 即 $t_i \approx t$, 结合式(6)和(7)可得:

$$t_M = \mathbf{M} \mathbf{W} \mathbf{b}_i \quad (8)$$

式(8)建立起了温度测量数据与未知系数向量之间的关系,对式(8)使用最小二乘法求解,整理后可得:

$$(\mathbf{M} \mathbf{W})^T \mathbf{M} \mathbf{W} \mathbf{b}_i = (\mathbf{M} \mathbf{W})^T t_M \quad (9)$$

未知系数向量 $\mathbf{b}_i$ 求解后,温度分布重建结果即可通过式(6)计算得出。

为了表征重建效果,定义重建误差如下:

$$e = \frac{\|t - t_i\|_2}{\|t\|_2} \times 100\% \quad (10)$$

1.3 测点优化布置算法

由式(9)可知,测量矩阵 $\mathbf{M}$ 直接影响温度分布重建结果,优化测点布置方式对提升温度分布重建精度有着深远的意义。

现阶段,常用的测点优化布置算法为最小条件数法。从线性方程组角度分析,最小化式(9)中系数矩阵

$(\boldsymbol{M}\boldsymbol{W})^T\boldsymbol{M}\boldsymbol{W}$ 的条件数对提升系数向量 $\boldsymbol{b}_i$ 求解稳定性以及准确性都有着重要作用,最小条件数法以最小化系数矩阵 $(\boldsymbol{M}\boldsymbol{W})^T\boldsymbol{M}\boldsymbol{W}$ 的条件数为目标,使用贪心算法遍历测点位置,最终获得最小化条件数的测点优化布置方式。

为了进一步提升温度重建算法的准确性,本文提出了基于随机森林的测点优化布置算法,从数据分析的角度,探索测点布置方式与重建误差之间的关系,进而实现测点的优化布置。

定义误差变化率,计算重建误差的变化,用以量化分析不同布置算法对重建精度的影响:

$$r = \frac{|e - e_o|}{e} \times 100\%$$

(11)

式中: e 为原有算法重建误差; e_o 为优化算法重建误差。

2 随机森林算法

随机森林是一种建立在决策树上的集成学习算法,广泛应用于分类和回归问题。

2.1 决策树

作为经典的机器学习算法,决策树采用树形结构通过对数据进行逻辑分析并分组,寻找总结数据中的深层次规律,以此对新数据进行处理。

决策树模型由根节点、中间节点以及叶子节点三部分组成。决策树常用的计算方法有 ID3、C4.5、C5.0、CART 等。决策树根据选定的计算规则对数据进行分类,确定根节点,并继续划分不同的中间节点,直至节点中只含有一类特征,即叶子节点。决策树结构如图 1 所示。

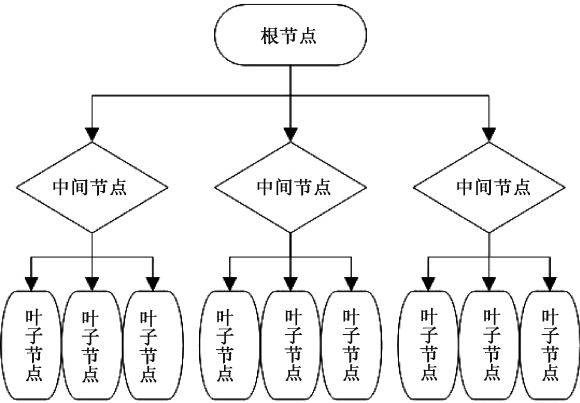


图 1 决策树结构
Fig. 1 Structure of decision tree

决策树作为一种单模型算法,在处理大量复杂样本数据时,计算精度较低。尽管通过剪枝等技术手段能够提升算法精度,但应用中依然存在计算复杂以及过拟合等问题。因此,决策树算法难以适用于工业过程中大量数据的分析和处理。

2.2 随机森林算法

为了解决决策树存在的问题,随机森林算法在决策树的基础上使用样本数据集建立起包含多个决策树的集成模型,综合考量各个决策树的输出。构建随机森林模型的步骤如下。

- 1) 在样本数据集中,采用有放回取样的方式随机抽取数据,形成多个样本数据子集。
- 2) 使用样本数据子集构建多棵决策树,每棵决策树按照分裂规则进行分生长,直至节点不纯度最小,即叶子节点时,决策树停止生长。
- 3) 将所有决策树整合成随机森林,并以投票的方式输出最终结果。

随机森林采用数据子集的方式构建决策树,并未使用所有特征,且训练数据子集以及特征选择具有随机性,所以随机森林模型中的各个决策树相互独立。通过对各个决策树模型计算结果综合分析得出随机森林模型的输出结果,既提升了算法的准确性,也避免了过拟合问题。此外,随机森林算法是一种并行算法,具有较快的处理速度和较强的数据处理能力,因此随机森林能够处理高维数据。

随机森林模型的结构如图 2 所示。

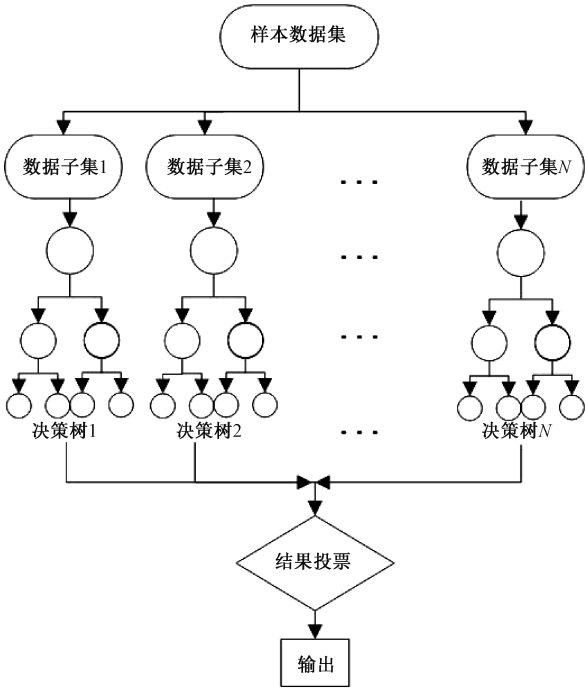


图 2 随机森林模型
Fig. 2 Model of random forest

3 基于随机森林的测点优化布置算法

分类与回归树 (classification and regression tree,

CART)算法中使用基尼指数(Gini index)作为节点分割的标准。实际应用中基尼指数更为准确的定义为基尼不纯系数(Gini impurity)^[18],基尼不纯度越小表示基于所选特征对应分类错误概率越小,即所选特征在决定分类结果过程中的重要性越大。随机森林算法对不同决策树中的基尼不纯系数进行综合累加,分析特征在分类计算过程中的重要性。基于随机森林的测点优化布置算法正是运用这一思想,通过样本数据构建随机森林模型,对不同特征进行重要性排序,提取重要性排序靠前的特征,将其所对应的位置作为测点位置,实现测点位置的优化布置。

基于随机森林的测点优化布置算法计算流程如图 3 所示。

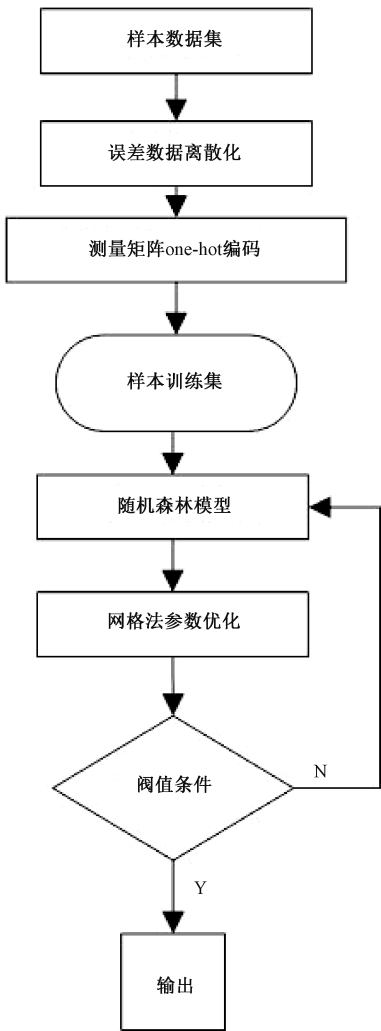


图 3 基于随机森林的测点优化布置算法计算流程

Fig. 3 Calculation process of optimal sensor placement algorithm based on random forest

总结计算过程主要包括以下步骤:

1) 利用经验数据构建样本数据集,样本数据集包括

测点的布置方式以及对应的重建误差。对重建误差离散化,形成一定数量的不同样本分类。

2) 将所有测点位置作为样本特征,测量矩阵进行 one-hot 编码^[19],与所对应的样本分类共同作为样本训练集。

3) 使用样本训练集构建随机森林模型。

4) 采用网格法调整随机森林模型参数,同时调整样本分类阈值,保证随机森林模型计算结果的准确性。

5) 由于随机森林算法在构建模型时采用特征子集进行计算,为降低随机性的干扰并提升计算所得测点布置方式的抗噪能力,设定条件数阈值评价计算结果。当阈值条件满足时,输出运算结果,阈值条件不满足时,重复步骤 3)~5) 直至获得测点优化布置。

4 实验分析

为了验证基于随机森林的测点优化布置算法的有效性,使用仿真实验以及燃烧实验对算法实际应用效果进行分析。

4.1 仿真实验设置

使用 Fluent 软件构建燃烧仿真模型,燃料气体(民用液化石油气,丙烷占比约为 99%,为简化计算过程,仿真计算模型中燃料成分设定为丙烷)和空气在燃烧器内引燃并扩散至炉膛内继续燃烧,高温烟气通过引风机引至尾部烟道排出。模型结构如图 4 所示。

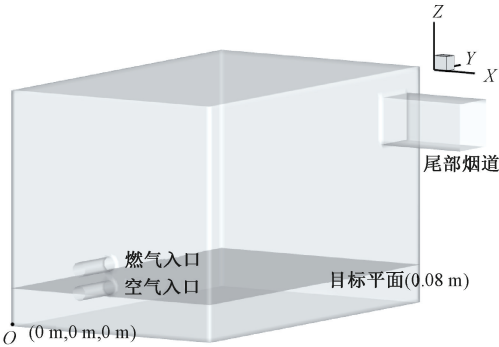


图 4 燃烧仿真模型结构示意图

Fig. 4 Schematic diagram of combustion simulation model

仿真模型炉膛尺寸为 0.5 m×0.8 m×0.5 m,空气入口管径为 DN32,管径中心点坐标为(0.25 m, 0 m, 0.1 m),燃料入口管径为 DN32,管径中心点坐标为(0.25 m, 0 m, 0.15 m),烟道出口截面尺寸为 0.15 m×0.15 m。以高度 0.08 m 平面作为目标平面,对其温度分布进行重建。

采用四面体和六面体对模型进行网格划分,在流体结构复杂位置对网格进行加密,综合考虑计算精度、计算

收敛速度等因素,采用 GCI(grid convergence index)^[20] 对网格独立性进行验证,网格节点数为 784 370 时,计算效果较好,后续研究中采用节点数为 784 370 的网格进行仿真计算。

计算模型选用标准 $k-\varepsilon$ 湍流模型作为流动模型。启动化学组分传输和反应。反应物为丙烷-空气。空气入口与燃料入口边界条件设定为速度入口,尾部烟道出口设定为压力出口,出口压力设为 -20 Pa 。残差收敛条件设为 10×10^{-5} ,综合残差、观测点温度以及流场质量守恒情况综合判断计算是否收敛。

为了构建重建过程所需的经验数据集,并考虑燃烧过程中的不同工况,选择不同的过量空气系数 α ,对完全燃烧、不完全燃烧等情况进行模拟,仿真工况设定如表 1 所示。

表 1 样本工况
Table 1 Sample conditions

燃料流量/ ($\text{m}\cdot\text{s}^{-1}$)	空气流量/($\text{m}\cdot\text{s}^{-1}$)		
	$\alpha=0.75$	$\alpha=1.25$	$\alpha=1.40$
0.41	1.57	2.62	2.94
0.52	1.97	3.28	3.67
0.62	2.36	3.93	4.40
0.72	2.75	4.59	5.14
0.83	3.15	5.24	5.87
0.92	3.54	5.90	6.60
1.03	3.93	6.55	7.34
1.13	3.33	7.21	8.07

4.2 燃烧实验设置

燃烧实验台结构如图 5 所示。

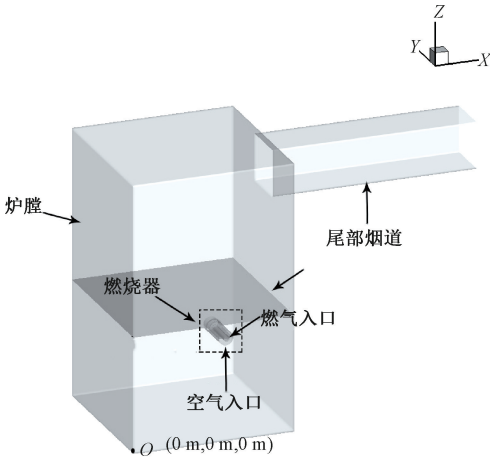


图 5 燃烧实验台结构示意图

Fig. 5 Schematic diagram of combustion test bench

炉膛尺寸为 $0.9\text{ m}\times0.9\text{ m}\times1.5\text{ m}$,空气入口管径为 DN80,内部燃料入口管径为 DN32,烟道出口截面尺寸为 $0.20\text{ m}\times0.23\text{ m}$ 。实验台主要装置如图 6 所示。

炉墙由保温砖堆砌而成,为了避免烟气温度过高并



图 6 燃烧实验装置

Fig. 6 The basic structure of combustion test equipment

且便于调整工况,设有喷水减温装置以及循环水冷装置。

实验过程中,以高度 0.65 m 平面的温度分布作为研究对象。温度信息由热电偶测量获得,目标平面设有 85 个测点位置。实验所用热电偶信息如表 2 所示。

表 2 热电偶参数
Table 2 Parameters of thermocouples

型号	热电偶参数	数量
HT-116	K 型,外径 12 mm,总长 700 mm,量程 $0\text{ }^{\circ}\text{C}\sim1\text{ }250\text{ }^{\circ}\text{C}$	7
HT-129	S 型,外径 12 mm,总长 700 mm,量程 $0\text{ }^{\circ}\text{C}\sim1\text{ }500\text{ }^{\circ}\text{C}$	1
WRR-130	B 型,外径 12 mm,总长 1 000 mm,量程 $0\text{ }^{\circ}\text{C}\sim1\text{ }800\text{ }^{\circ}\text{C}$	2

温度测点纵坐标分别为 0.15 、 0.30 、 0.45 、 0.61 、 0.75 m ,相邻测点横坐标间隔为 0.05 m ,测点位置如图 7 所示。

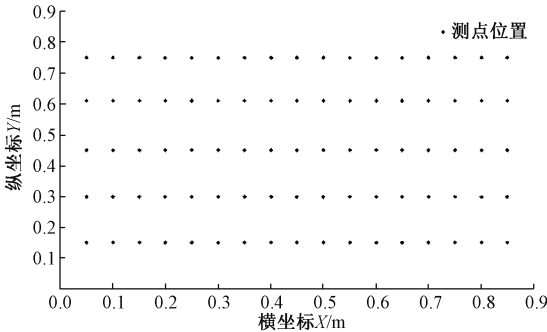


图 7 测点位置图

Fig. 7 Sensor locations

为获得重建算法所需经验数据集,使用 Fluent 模拟实验台燃烧工况,仿真模型几何结构与燃烧实验台保持一致,计算模型按照实际情况设定。模拟过程中采用 GCI 对网格独立性进行验证,并与燃烧实验数据对比,最终选择节点数为 2641471 的网格进行计算。经验数据集仿真工况设定如表 3 所示。

表 3 经验数据集仿真工况

Table 3 Sample conditions for empirical dataset

燃料流量/ ($\text{m}\cdot\text{s}^{-1}$)	空气流量/($\text{m}\cdot\text{s}^{-1}$)				
	$\alpha=0.80$	$\alpha=0.90$	$\alpha=1.00$	$\alpha=1.10$	$\alpha=1.25$
0.43	1.22	1.37	1.52	1.68	1.90
0.54	1.52	1.71	1.90	2.09	2.38
0.65	1.83	2.06	2.28	2.51	2.86
0.75	2.13	2.40	2.67	2.93	3.33

仿真计算后,由不同工况下的目标平面温度值构建经验数据集,提取重建算法所需特征向量。

由于实验台测量条件限制,每次最多只能同时测量 10 个温度值,考虑火焰波动等情况对温度测量带来的影响,每个测点位置在进行温度测量时,测量过程持续 1 min,所得数据平均值作为测量值。

4.3 数据分析

1) 仿真实验数据分析

为了验证基于随机森林的测点优化布置算法在仿真实验中的应用效果,设置了 6 组测试工况,测试工况边界条件如表 4 所示。

表 4 测试工况

Table 4 Test conditions.

工况编号	燃料流量/($\text{m}\cdot\text{s}^{-1}$)	空气流量/($\text{m}\cdot\text{s}^{-1}$)
1	0.58	2.20
2	0.58	3.67
3	0.58	4.11
4	1.23	4.72
5	1.23	7.86
6	1.23	8.81

将测试工况下目标平面的温度仿真值作为待重建温度分布。从测试工况温度仿真值中提取少量的温度数据作为测量数据,结合温度分布重建算法以及随机森林的测点优化布置算法实现温度分布的重建。

为了使算法的应用场景更接近实际情况,并测试算法在存在测量误差情况下的效果,提取温度测量数据时加入信噪比(signal noise ratio, SNR)为 40 dB 的高斯噪声。

经过对比多次重建结果可知,仿真实验中,温度分布重建算法在使用 7 个特征向量,12 个测点时,重建效果较好。因此,在测试优化布置算法时,采用 7 个特征向

量,12 个测点进行分析计算。
通过原有最小条件数法和基于随机森林的测点优化布置算法所得的 12 个测点布置如图 8 所示。

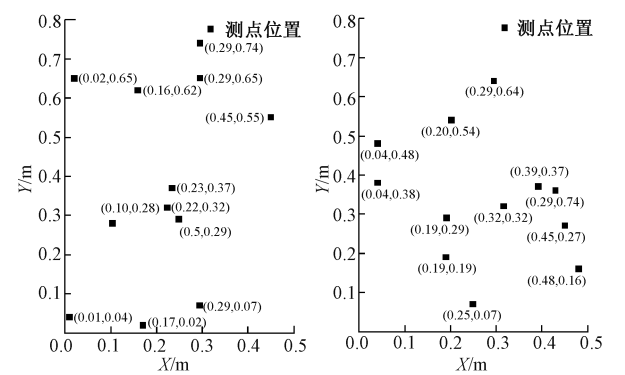


图 8 测点位置
Fig. 8 Sensor placements calculated by different methods

为了减少噪声随机性对重建结果带来的影响,获得更为普遍的结论。分别使用随机布置、原有算法(最小条件数法)以及基于随机森林的测点优化布置算法计算所得的测点布置方式进行 1 000 次重建,求取重建误差的平均值作为最终算法性能评价指标,测试工况 1~6 的比较结果如图 9 所示。

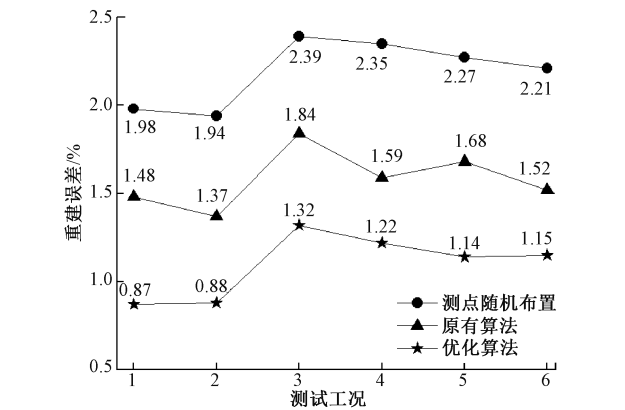


图 9 不同测点布置方式下的重建误差
Fig. 9 Reconstruction errors in different sensor placements

由图 9 可知,使用原有算法和基于随机森林的测点优化布置算法计算所得的测点布置方式均能在一定程度上降低温度分布重建的误差。在 6 组测试工况中,使用优化算法对比原有算法,重建精度分别提升 41%、36%、28%、23%、32%、24%。因此,在仿真实验中,使用基于随机森林的测点优化布置算法所得测点布置能够提升重建算法的精度。

2) 燃烧实验数据分析

为了验证基于随机森林的测点优化布置算法在实际

中的应用效果,在实验台组织燃烧实验。工况设为燃料入口速度为 1.5 m/s,空气入口速度为 7.46 m/s。

燃烧状态平稳后,采集目标平面温度数据,作为待重建温度分布。经过分析对比,针对燃烧实验数据,重建过程中选择 6 个特征向量,6 个测点时重建效果较好。由原有算法及优化算法分别计算出测点布置方式,并随机选取六组测点布置作为对比。不同布置方式下的测点坐标如表 5 所示。

表 5 测点坐标

Table 5 Coordinates of sensor locations

布置方式	测点坐标/m
原有算法	(0.40,0.15),(0.70,0.15),(0.20,0.45)
	(0.35,0.61),(0.05,0.61),(0.45,0.15)
优化算法	(0.15,0.15),(0.45,0.30),(0.75,0.45)
	(0.55,0.45),(0.50,0.45),(0.45,0.15)
随机布置 1	(0.30,0.15),(0.15,0.15),(0.10,0.15)
	(0.20,0.45),(0.35,0.61),(0.50,0.75)
随机布置 2	(0.55,0.30),(0.45,0.30),(0.15,0.30)
	(0.45,0.45),(0.35,0.61),(0.25,0.61)
随机布置 3	(0.25,0.15),(0.45,0.30),(0.60,0.45)
	(0.25,0.45),(0.10,0.45),(0.60,0.75)
随机布置 4	(0.15,0.15),(0.70,0.15),(0.55,0.45)
	(0.45,0.45),(0.75,0.61),(0.60,0.15)
随机布置 5	(0.80,0.30),(0.85,0.45),(0.70,0.15)
	(0.05,0.45),(0.85,0.75),(0.70,0.75)
随机布置 6	(0.85,0.15),(0.35,0.45),(0.70,0.61)
	(0.15,0.61),(0.40,0.75),(0.10,0.75)

采用不同的测点布置方式对目标温度分布进行重建,与实验测量值作比较,将重建误差作为评价指标,8 组测点布置方式对应重建误差如表 6 所示。

表 6 测试工况

Table 6 Test conditions

布置方式	重建误差/%
原有算法	1.94
优化算法	1.45
随机布置 1	3.44
随机布置 2	14.15
随机布置 3	8.91
随机布置 4	2.76
随机布置 5	4.70
随机布置 6	2.63

由表 6 可知,在燃烧实验中,重建误差受到测点布置方式的影响。不同的测点布置方式重建误差相差甚至大于 10%,这为温度分布重建带来极大的不确定性。采用原有算法和基于随机森林的测点优化布置算法计算所得的测点布置方式降低了重建误差。而优化算法在原有算法基础上,将重建精度进一步提升了 25%。因此,在燃烧

实验中,使用基于随机森林的测点优化布置算法所得测点布置同样能够提升重建算法的精度。

5 结 论

在应用主成分分析算法进行温度重建过程中,测点位置对重建精度有着深远的影响。本文将随机森林算法引入温度分布重建中,提出了基于随机森林的测点优化布置算法,在仿真实验和燃烧实验中进行了测试,对比原有最小条件数算法,确定了基于随机森林的测点优化布置算法能够提升温度分布重建算法精度,验证了基于随机森林的测点优化布置算法的有效性,并为随机森林解决工业问题提供了新思路。

参考文献

[1] MU H, LI Z, WANG X, et al. Temperature distribution measurement using the Gaussian process regression method [J]. Mathematical Problems in Engineering, 2017, 2017(1): 1-12.

[2] LUH G C, LIN C Y. PCA based immune networks for human face recognition [J]. Applied Soft Computing, 2011, 11(2): 1743-1752.

[3] PLATON R, DEHKORDI V R, MARTEL J. Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis [J]. Energy & Buildings, 2015, 92(1): 10-18.

[4] 胡帅,顾艳,姜华,等. 基于 PCA-BPNN 的学生写作成绩预测模型研究 [J]. 国外电子测量技术, 2015, 34(12): 41-44.

HU SH, GU Y, JIANG H, et al. Study of student score prediction model based on PCA-BPNN [J]. Foreign Electronic Measurement Technology, 2015, 34(12): 41-44.

[5] 许桢英,罗来齐,王匀,等. 基于 PCA 的管道缺陷导波信号特征优化方法 [J]. 电子测量技术, 2016, 39(4): 160-163.

XU ZH Y, LUO L Q, WANG Y, et al. Feature optimization method of pipe defect guided signals based on PCA [J]. Electronic Measurement Technology, 2016, 39(4): 160-163.

[6] TERASHIMA O, SAKAI Y, ITO Y. Measurement of fluctuating temperature and POD analysis of eigenmodes in a heated planar jet [J]. Experimental Thermal and Fluid Science, 2018, 92: 113-124.

[7] YIN H, ZHANG C, JI Y. Distributed clustering using distributed mixture of probabilistic PCA [C]. IEEE International Conference on Fuzzy Systems and

- Knowledge Discovery, 2014.
- [8] 王帅, 黄海鸿, 韩刚, 等. 基于 PCA 与 GA-BP 神经网络的磁记忆信号定量评价[J]. 电子测量与仪器学报, 2018, 32 (10): 190-196.
- WANG SH, HUANG H, HAN G, et al. Quantitative evaluation of magnetic memory signal based on PCA&GA-BP neural network [J]. Journal of Electronic measurement and Instrumentation, 2018, 32 (10): 190-196.
- [9] JIANG C, SOH Y C, LI H. Sensor and CFD data fusion for airflow field estimation [J]. Applied Thermal Engineering, 2016, 92(5): 149-161.
- [10] LASAGNA D, TUTTY O R. Wall-based reduced-order modelling [J]. Numerical Method in Fluids, 2016, 80(9): 511-535.
- [11] SEMAAN R. Optimal sensor placement using machine learning[J]. Computers & Fluids, 2017, 159: 167-176.
- [12] GENUER R, POGGI J M, TULEAU-MALOT C. Variable selection using random forests [J]. Pattern Recognition Letters, 2010, 31(14): 2225-2236.
- [13] HAPFELMEIER A, HOTHORN T, ULM K, et al. A new variable importance measure for random forests with missing data[J]. Statistics & Computing, 2014, 24(1): 21-34.
- [14] HAPFELMEIER A, ULM K. Variable selection by random forests using data with missing values [J]. Computational Statistics & Data Analysis, 2014, 80: 129-139.
- [15] 刘金海, 付明芮, 唐建华. 基于漏磁内检测的缺陷识别方法 [J]. 仪器仪表学报, 2016, 37 (11): 2572-2581.
- LIU J H, FU M R, TANG J H. MFL inner detection based defect recognition method[J]. Chinese Journal of Scientific Instrument, 2016, 37(11): 2572-2581.
- [16] 刘仁明, 毛建旭, 历艳琨. 基于随机森林回归的人脸特征点定位 [J]. 电子测量与仪器学报, 2016, 30(5): 684-693.
- LIU R M, MAO J X, LI Y K. Facial landmarks location based on random forest regression [J]. Journal of Electronic Measurement and Instrumentation, 2016, 30(5): 684-693.
- [17] 黄南天, 王达, 刘座铭, 等. 复杂噪声环境下电能质量复合扰动特征选择 [J]. 仪器仪表学报, 2018, 39(4): 82-90.
- HUANG N T, WANG D, LIU Z M, et al. Feature selection of composite power quality disturbances under complex noise environment [J]. Chinese Journal of Scientific Instrument, 2018, 39(4): 82-90.
- [18] BOULESTEIX A, BENDER A, BERMEJO J L, et al. Random forest Gini importance favours SNPs with large minor allele frequency: Impact sources and recommendations[J]. Briefings in Bioinformatics, 2011, 13(3): 292-304.
- [19] ALAYA M, BUSSY S, GAIFFAS S, et al. Binarisity: A penalization for one-hot encoded features[J]. Journal of Machine Learning Research, 2019, 20: 1-34.
- [20] MANNA P, DHARAVATH M, SINHA P K, et al. Optimization of a flight-worthy scramjet combustor through CFD [J]. Aerospace Science & Technology, 2013, 27(1): 138-146.

作者简介



陈敏鑫, 分别在 2009 年和 2012 年于华北电力大学获得学士学位和硕士学位, 现为华北电力大学博士研究生, 主要研究方向为现代测量技术、数据分析、机器学习。

E-mail: chenminxin@ncepu.cn

Chen Minxin received B. Sc. and M. Sc. from North China Electric Power University in 2009 and 2012. Currently he is pursuing the Ph. D. degree in thermal engineering in North China Electric Power University. His main research interests include measurement technology, numerical analysis and machine learning.



刘石, 1992 年于英国剑桥大学获得博士学位, 现为华北电力大学教授, 主要研究方向为多相流测量、数值模拟和计算流体力学。

E-mail: liushi@ncepu.edu.cn

Liu Shi received Ph. D. from the University of Cambridge, UK, in 1992. He is currently a professor at North China Electric Power University. His main research interests include multiphase flow measurement, numerical simulation and computational fluid dynamic.