

# 矿井通风参数缺失数据插补方法

倪景峰<sup>1,2</sup>, 刘雪峰<sup>1,2</sup>, 邓立军<sup>1,2</sup>

(1. 辽宁工程技术大学 安全科学与工程学院, 辽宁 葫芦岛 125105; 2. 辽宁工程技术大学 矿山热动力灾害与防治教育部重点实验室, 辽宁 葫芦岛 125105)

**摘 要:** 矿井智能通风系统对矿山智能化建设至关重要。为解决矿井通风参数在实际测量时, 因为巷道不具备测试条件、仪器信号受到干扰、巷道断面风速不均一、人工操作不当等制约性因素, 造成的矿井通风参数数据缺失问题, 提出了 1 种基于随机森林-链式方程多重插补法的矿井通风参数缺失数据插补方法。采用链式方程多重插补法, 通过迭代对每个缺失的属性值产生  $n$  个插补值, 从而产生  $n$  个完整数据集, 对  $n$  个完整数据集进行分析优化得到 1 个最终的完整数据集。为了提高缺失值插补精度, 合理考虑了矿井通风参数缺失数据的不确定性对分析过程的影响, 在随机森林的预测任务中, 结合预测均值匹配模型对缺失数据进行插补。以潞新二矿为实验对象, 利用智能矿井通风仿真系统 IMVS 对潞新二矿矿井通风参数原始数据集进行数据预处理, 得到完整、准确的矿井通风参数完整数据集, 对完整数据集分别进行了不同缺失属性、不同数据缺失率、不同迭代次数的对比试验。以多种模型评价指标对模型有效性进行评估。结果表明: 基于随机森林的链式方程多重插补模型插补形成的完整数据集与原始数据集具有很好的相似性; 对不同缺失列进行插补实验的结果显示插补模型可以轻松处理混合类型的数据, 自主学习参数之间的相关性从而降低了插补复杂性; 迭代后形成的  $n$  个数据集通过分析合并成一个最终数据集, 提高了插补准确率; 对初始插补后的完整数据集进行不同迭代次数的试验, 发现迭代超过一定次数后, 数据相关性一定会收敛。

**关键词:** 矿井通风; 随机森林; 链式方程多重插补; 缺失数据; 数据插补

**中图分类号:** TD72 **文献标志码:** A **文章编号:** 0253-9993(2024)05-2315-09

## Method for filling missing data of mine ventilation parameters

NI Jingfeng<sup>1,2</sup>, LIU Xuefeng<sup>1,2</sup>, DENG Lijun<sup>1,2</sup>

(1. College of Safety Science & Engineering, Liaoning Technical University, Huludao 125105, China; 2. Key Laboratory of Mine Thermo-motive Disaster and Prevention Ministry of Education, Liaoning Technical University, Huludao 125105, China)

**Abstract:** The intelligent mine ventilation system is very important for the intelligent construction of coal mines. In order to solve the problem of missing mine ventilation parameter data caused by the lack of measurement conditions, instrument signal interference, uneven wind speed of roadway section, improper manual operation and other restrictive factors during actual measurement of mine ventilation parameters, a method for filling the missing data of mine ventilation parameters based on the multiple imputation method of random forest-chained equation was proposed. Multiple imputation with chained equations is used to generate  $n$  filled values for each missing attribute value by iterations, resulting in  $n$  complete

收稿日期: 2023-04-11 修回日期: 2023-05-28 责任编辑: 王晓珍 DOI: 10.13225/j.cnki.jccs.2023.0481

基金项目: 国家自然科学基金青年科学基金资助项目 (51904143)

作者简介: 倪景峰 (1975—), 男, 河北承德人, 教授, 博士。E-mail: njf\_3829@163.com

引用格式: 倪景峰, 刘雪峰, 邓立军. 矿井通风参数缺失数据插补方法[J]. 煤炭学报, 2024, 49(5): 2315-2323.

NI Jingfeng, LIU Xuefeng, DENG Lijun. Method for filling missing data of mine ventilation parameters[J]. Journal of China Coal Society, 2024, 49(5): 2315-2323.



移动阅读

datasets, and a final complete dataset is obtained by analyzing and optimizing the  $n$  complete datasets. In order to improve the filling accuracy of missing values, the influence of the uncertainty of missing data of mine ventilation parameters on the analysis process is reasonably considered, and the missing data is filled in the prediction task of random forest in combination with the prediction mean matching model. Taking the Luxin No.2 Mine as an experimental example, the intelligent mine ventilation simulation system IMVS was used to preprocess the original data set of ventilation parameters of the Luxin No.2 Mine to obtain a complete and accurate complete dataset of mine ventilation parameters. Comparative experiments with different missing attributes, different data missing rates, and different number of iterations were conducted separately for the complete data set. The effectiveness of the model was evaluated by a variety of model evaluation indicators. The results show that the complete data set formed by the multiple imputation method of random forest-chained equation has good similarity with the original data set. Results of filling experiments with different missing columns show that the filling model can easily handle mixed data types, autonomously learning the correlations between parameters and thus reducing filling complexity. The  $n$  datasets formed after iterations are combined into a final dataset by analysis, which improves the filling accuracy. Experiments with different iterations on the complete data set after initial filling show that the data correlation will converge after a certain number of iterations.

**Key words:** mine ventilation; random forest; multiple interpolation of chain equations; missing data; data completion

矿山智能化建设离不开矿井智能通风,通风系统的智能化、信息化是保障我国煤炭工业转型升级和高质量发展的核心技术之一<sup>[1]</sup>。完整、准确的矿井通风参数是对矿井智能通风系统进行通风网络解算、通风状态分析、异常或灾变智能决策的重要数据支撑<sup>[2]</sup>。矿井通风参数包括风速、风量、风压、风阻、巷道断面积、干湿温度等,然而,在井下实际测量时,经常因为仪器受到干扰、巷道断面风速不均一、人工操作不当、巷道环境恶劣等制约性因素,导致矿井通风参数存在数据缺失或数据质量下降的现象<sup>[3]</sup>。经实际测量发现,常见易缺失的参数包括巷道断面积、风量、风阻等。存在缺失的参数一方面无法准确反映矿井通风系统的工作状态、无法对通风系统进行状态分析,另一方面会导致煤矿通风工作人员无法对通风设施进行修改、维护,影响其对通风系统的运维决策。为解决上述问题,对矿井通风参数的缺失数据进行插补至关重要。

大多数统计分析方法都需要完整的数据集,但在几乎所有的研究中都存在数据缺失现象<sup>[4]</sup>。常规的补全方法有均值填充法<sup>[5]</sup>、众数填充法<sup>[6]</sup>、函数逼近法<sup>[7]</sup>、期望最大化算法<sup>[8]</sup>、最近邻插补法<sup>[9]</sup>、多重插补法<sup>[10]</sup>、机器学习模型预测插补法<sup>[11-13]</sup>等方法。关于矿井通风参数缺失数据插补方面,目前有对矿井通风摩擦阻力系数预测的一些研究。刘泽功<sup>[14]</sup>利用通风阻力测定数据和分支风量结合起来求算复杂通风网络分支风阻,并给出了具体的测算步骤和测算数据。司俊鸿等<sup>[15]</sup>针对复杂通风网络易构造病态测风求阻模型的问题,提出了基于贪婪进化算法的 Tikhonov 正则化参数求解方法,对模型进行修正处理。邓立军<sup>[16]</sup>基于最

小二乘原理,综合考虑压力、风量以及通风阻力系数范围约束,将通风阻力系数反演问题转化非线性优化问题,最终提出了一种基于反映通风系统阻力系数变化的巷道风量测点和节点压力测点布置方法。刘剑等<sup>[17]</sup>理论分析了矿井定常湍流对通风阻力测试的影响。李雨成等<sup>[18]</sup>研究了基于节点压能的风量反演风阻算法,提出使用 2 组风量数据及部分节点压能数据反演风阻。机器学习算法在摩擦阻力系数的预测中已有一定的研究,通过机器学习与缺失值插补算法结合的方法直接对矿井通风参数进行插补的研究还较为少见。

结合矿井通风参数缺失数据产生的原因与矿井通风参数缺失数据类型进行分析,可以得到矿井通风参数缺失模式是多变量缺失模式与一般缺失模式相结合,同时数据缺失类型属于随机缺失类型。鉴于此,在传统多重插补法的基础上,笔者构建了一种基于随机森林的链式方程多重插补法用于矿井通风参数缺失数据插补。该方法为每个缺失数据找到多个插补值,经过有限次迭代形成多个完整的数据集,从多个数据集中选择一个最好的数据集或者综合多个数据集最终形成一个最优数据集来用于后续工作。该方法不需要太多的超参数调整就可以完成缺失的分类和连续数据的缺失值单一或多重插补,具有快速、高效、内存利用率高的特征。为验证插补方法的可行性与准确性,首先利用智能矿井通风仿真系统 IMVS<sup>[19]</sup>构建潞新二矿矿井通风系统,在初步形成矿井通风系统的情况下,经过反复调试,得到完整、准确、与现场一致的二矿矿井通风参数完整数据集。再对完整数据集进行随机缺失处理,利用笔者提出的缺失数据插

补方法进行插补。通过对比试验,证明了方法的可行性,缺失数据插补的准确率得到进一步的提升,可为矿井通风系统智能化建设提供数据支撑。

## 1 基于随机森林的链式方程多重插补方法

### 1.1 多重插补法

多重插补法(MI)是指利用不完整数据集中的现有数据对缺失值进行多次插补,由此生成多个完整数据集的插补方法<sup>[20]</sup>。由于缺失值是数据集中的未知

因素,在一定程度上导致了分析结果不确定。针对缺失数据的不确定问题,多重插补方法将不完整数据集插补过程分为插补、分析和合并3个阶段。通过执行多次插补得到若干组插补值,在此基础上求解多个分析结果并对这些结果实行有效合并。相比单一插补,多重插补合理考虑缺失值不确定性对分析结果造成的影响,从而获得更为合理的推断。

基于多重插补法的不完整数据集插补过程如图1所示。

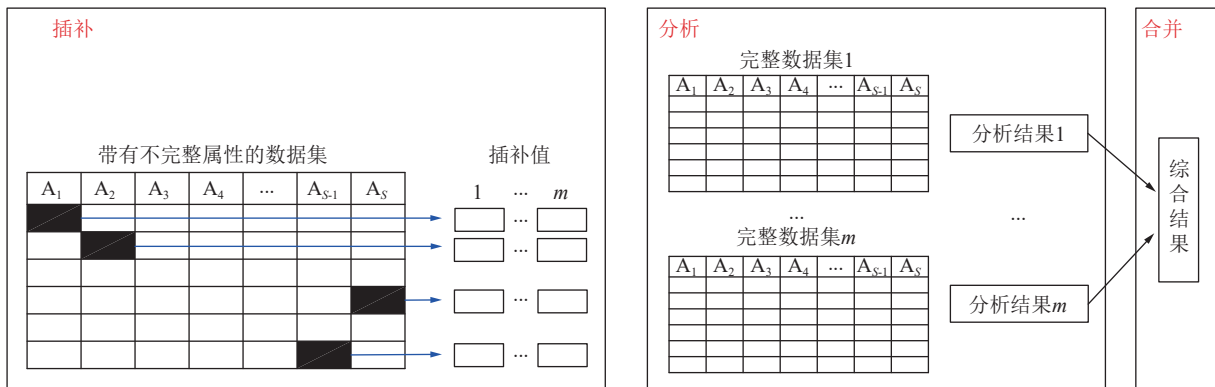


图1 基于多重插补法的不完整数据集插补过程

Fig.1 Completion process of incomplete data sets based on multiple imputation method

图1中, $S$ 表示数据集的属性数量; $A_j(j=1,2,\dots,S)$ 表示第 $j$ 个属性的名称;黑色方框表示缺失值;白色方框表示现有值; $m$ 表示插补次数。

(1) 插补。对不完整数据集展开 $m$ 次插补,得到 $m$ 组插补值,这些值反映了模型的不确定性,每组值都被用来插补数据集中的缺失值,由此产生 $m$ 个完整数据集。

(2) 分析。采用同样的分析算法对所有完整数据集进行独立的分析,并求解出 $m$ 个分析结果。

(3) 合并。综合上一阶段的所有结果获得最终分析结果。

相比于拟合插补和均值替换等求解单一插补值

的缺失值处理方法,多重插补更关注于不完整数据的分析结果。多重插补并没有试图通过模拟值去估计每个缺失值,而是采用多组插补值表征缺失值的不确定性,对不同插补值下的完整数据集进行多次分析,并综合若干个分析结果以获得最终的推断。该方法合理考虑了缺失值的不确定性对分析过程的影响,使得统计推断更有效,有助于提高分析质量。

### 1.2 基于随机森林的链式方程多重插补方法

在多重插补方法的基础上,使用链式方程进行迭代优化,使用随机森林进行缺失值预测,基于随机森林的链式方程多重插补的方法过程如图2所示。

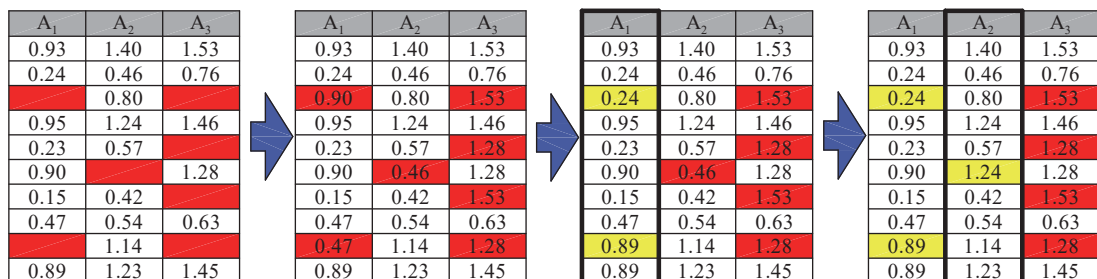


图2 基于随机森林的链式方程多重插补法

Fig.2 Multiple imputation method of chain equation based on random forest

Step1: 获得初始的随机插补的完整数据集。对含有缺失值的数据集,在当前属性值中随机选择一些值

对当前属性的缺失值进行插补,得到初始的随机插补的完整数据集 Rnd\_Dataset。



Step2: 对于初始的完整数据集开始迭代 (iterations =  $n$ )。对于 Rnd\_Dataset, 对所有不完整属性  $A_1$ 、 $A_2$ 、 $A_3$  进行缺失值的随机森林预测。在随机森林的预测任务中, 不是使用随机森林直接预测的值, 而是使用随机森林预测值中相近的值所对应的原始数据集中的值, 使用随机森林预测均值匹配的过程如图3所示。

图3中, 如  $A_1$  属性存在缺失数据, 则利用 Step1 中插补后的  $A_2$ 、 $A_3$  属性作为特征矩阵,  $A_1$  属性作为标签列, 构建随机森林模型对  $A_1$  属性预测出一组值  $E[A_1|A_2, A_3]$ , 利用预测值中与缺失属性对应的预测值相近的值所对应的原始数据集中的值对缺失属性进行插补。图3(b)中,  $A_1$  属性缺失值对应的预测值为 0.60 和 1.20, 在整组预测值中找到与之相近的 1 个所

对应的原始列中的值对缺失属性进行插补。设置参数 mean\_match\_candidates 来限制所利用值的数量, 若 mean\_match\_candidates = 5, 则选择与缺失属性对应的预测值中相近的 5 个值所对应的原始数据集中的值的均值进行插补。

对所有待插补属性均执行上述操作, 直到所有属性不含有缺失值, 得到一个完整数据集后, 就完成了完整数据集的一次迭代。

Step3: 对于一次迭代后的完整数据集, 重复 Step2, 直到迭代次数为  $n$  时停止, 得到最终的 1 个完整数据集。

Step4: 利用 Step1~3 得到  $m$  个完整数据集, 对  $m$  个完整数据集进行分析, 得到最终插补结果。

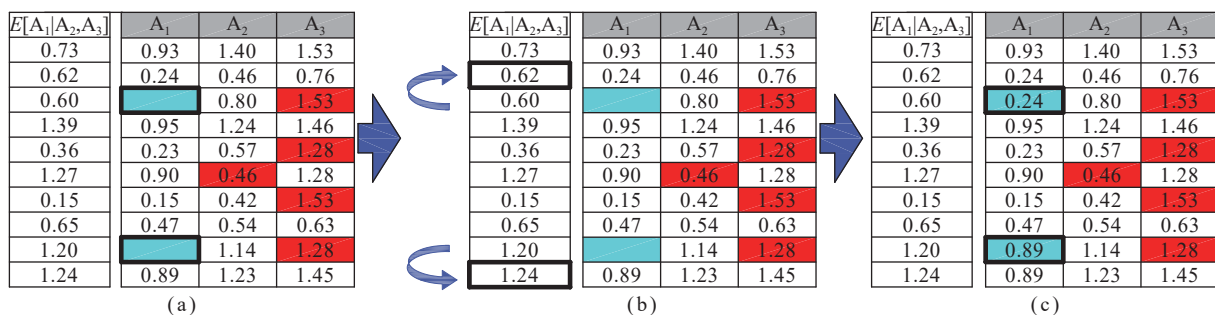


图3 基于随机森林的预测均值匹配过程

Fig.3 Prediction mean matching process based on random forest

## 2 矿井通风参数缺失数据插补流程与评价指标

### 2.1 矿井通风参数缺失数据插补流程

基于随机森林的链式方程多重插补法的矿井通风参数缺失数据插补流程共分为 3 个阶段:

Step1: MICE-Forest() 函数利用非缺失的矿井通风参数对原始数据集中的缺失值进行插补得到  $m$  个完整数据集。通过设置程序中 datasets 的值控制插补后数据集的数量。

Step2: with() 函数分别对每个插补后的完整数据集进行数据分析, 得到不同的分析结果<sup>[21]</sup>。

Step3: pool() 函数将 Step2 得到的结果按照最优原则整合成最终数据集。

矿井通风参数缺失数据插补流程如图4所示。

### 2.2 评价指标

对于线性回归及其他回归模型来说, 评价连续性可拟合的数据不能使用离散二分类器的评价指标对回归模型进行评价, 因此引入均方误差、平均绝对误差、可解释方差和决定系数等评价指标对模型进行度量。以上模型度量方式均可通过 python 调用 sklearn

库中函数由程序直接运行得出。

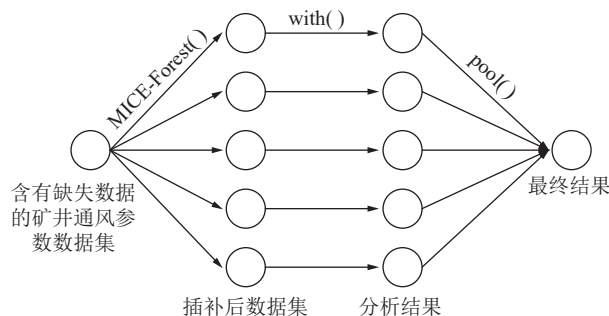


图4 矿井通风参数缺失数据插补流程

Fig.4 Completion process of missing data on mine ventilation parameters

(1) 均方误差 (Mean Square Error,  $E_{MS}$ ): 该指标计算的是拟合数据和原始数据对应样本点的误差的平方和的均值, 用于衡量拟合数据与原始真实值的匹配程度, 其值越小说明拟合效果越好, 均方误差计算公式为

$$E_{MS} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (1)$$

式中,  $N$ 为样本数;  $i$ 为每一个数据样本;  $y_i$ 为原始真实值;  $\hat{y}_i$ 为预测值。

(2) 解释方差 (explained\_variance,  $E_{\text{var}}$ ): 衡量所有预测值和样本之间的差的分散程度与样本本身的分散程度的相近程度<sup>[22]</sup>, 其取值范围是 $[0,1]$ , 值越大表示预测和样本值的分散分布程度越相近, 值越低越差, 其计算公式为

$$E_{\text{var}}(y, \hat{y}) = 1 - \frac{\frac{1}{N} \sum_{i=1}^N (z_i - \bar{z}_i)^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (2)$$

式中,  $\bar{y}_i$ 为原始值均值;  $z_i$ 为样本值与预测值之差;  $\bar{z}_i$ 为 $z_i$ 的均值。

(3) 决定系数  $R^2$ : 通过数据的变化来表征一个拟合的好坏,  $R^2$ 接近1, 说明模型效果越好。  $R^2$  计算公式为

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2} \quad (3)$$

(4) 平均绝对误差 (Mean Absolute Error,  $E_{\text{MA}}$ ): 表示预测值与真实值之间绝对误差的平均值, 用于评估预测结果和真实数据集的接近程度, 其值越小说明拟合效果越好, 计算式为

$$E_{\text{MA}} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (4)$$

### 3 潞新二矿通风参数插补试验

潞新二矿位于新疆维吾尔自治区哈密市三道岭矿区东部, 矿井主要可采煤层5层, 矿田面积约13.4 km<sup>2</sup>, 井田面积37.02 km<sup>2</sup>。潞新二矿矿井通风方式为中央并列式, 通风方法为机械抽出式, 主斜井、副斜井、进风立井及行人斜井进风, 回风立井回风。目前矿井总进风量为8 311 m<sup>3</sup>/min左右, 矿井通风阻力为1 380 Pa, 等积孔为4.87 m<sup>2</sup>。

#### 3.1 矿井通风参数数据集

由于井下环境复杂多变, 难以通过少次测量得到完整、准确的样本, 笔者借助自主研发的智能矿井通风仿真系统IMVS, 在初步形成潞新二矿通风系统的情况下, 固定部分风路风量, 进行风流分配, 经过反复调试后得到完整、与现场一致的矿井通风参数完整数据集。将完整数据集进行缺失处理后, 再进行缺失值插补, 最后与原始数据集对比以进行模型评价, 可以有效验证插补方法的可行性与准确性。构建潞新二矿矿井通风参数数据集, 具体步骤如下:

(1) 构建矿井通风仿真系统。基于2023年初对潞新二矿矿井通风系统进行普查和基础参数测定的结果, 利用智能矿井通风仿真系统IMVS构建潞新二矿矿井通风仿真系统。

(2) 获取矿井通风参数完整数据集。通过向潞新二矿矿井通风仿真系统输入测定参数、矿井原有基础数据, 固定部分风路风量, 进行风流分配, 经过反复调试, 最终得到潞新二矿矿井通风参数完整数据集并记录至Excel表格中。

潞新二矿矿井通风参数完整数据集见表1。

表1 潞新二矿矿井通风参数数据集

Table 1 Data set of mine ventilation parameters of Luxin No.2 mine

巷道序号	断面积/m <sup>2</sup>	测段风量/(m <sup>3</sup> ·s <sup>-1</sup> )	测段风速/(m·s <sup>-1</sup> )	实际阻力差值/Pa	百米摩擦风阻/(N·s <sup>2</sup> ·m <sup>-8</sup> )
1	9.09	5.89	0.65	52	0.015 178
2	7.40	81.30	10.99	277	0.010 950
⋮	⋮	⋮	⋮	⋮	⋮
101	14.00	38.43	2.75	34	0.017 800
⋮	⋮	⋮	⋮	⋮	⋮
200	12.20	42.71	3.50	268	0.007 270
201	11.60	12.46	1.07	17	0.010 890
⋮	⋮	⋮	⋮	⋮	⋮
312	10.22	9.58	0.94	13	0.009 560
313	9.71	14.02	1.44	131	0.009 560

(3) 随机生成缺失数据。完整数据集中包含313条巷道、5种属性参数, 共1 565条数据, 使用自定义函数随机生成缺失数据。

#### 3.2 插补试验与结果分析

为验证模型对矿井通风参数缺失数据插补的可行性和准确性, 还原实际测量中矿井通风参数数据集

失的随机性,利用自定义函数对原始完整数据集进行随机缺失数据处理。通过改变自定义函数中不同参数的值来设置不同缺失属性、不同缺失率、不同迭代次数的 3 组试验。试验不分先后顺序,对“断面积”、“测段风量”等属性进行缺失值插补试验时,结果显示迭代次数为 6 时模型基本收敛。为排除迭代次数对试验的影响,试验一、二中迭代次数均设置为 6 次,部分迭代次数试验在试验三中阐述。

### 3.2.1 试验 1

通过 `def missing_data_generate()` 自定义函数中 `missing_columns` 参数指定数据缺失列。文中以“断面积”和“测段风量”为例。

模型 1: 数据缺失列为“断面积”,数据缺失率为 5%,迭代次数为 6。

模型 2: 数据缺失列为“测段风量”,数据缺失率为 5%,迭代次数为 6。

模型 1、2 插补后数据集在分析合并前后的对比

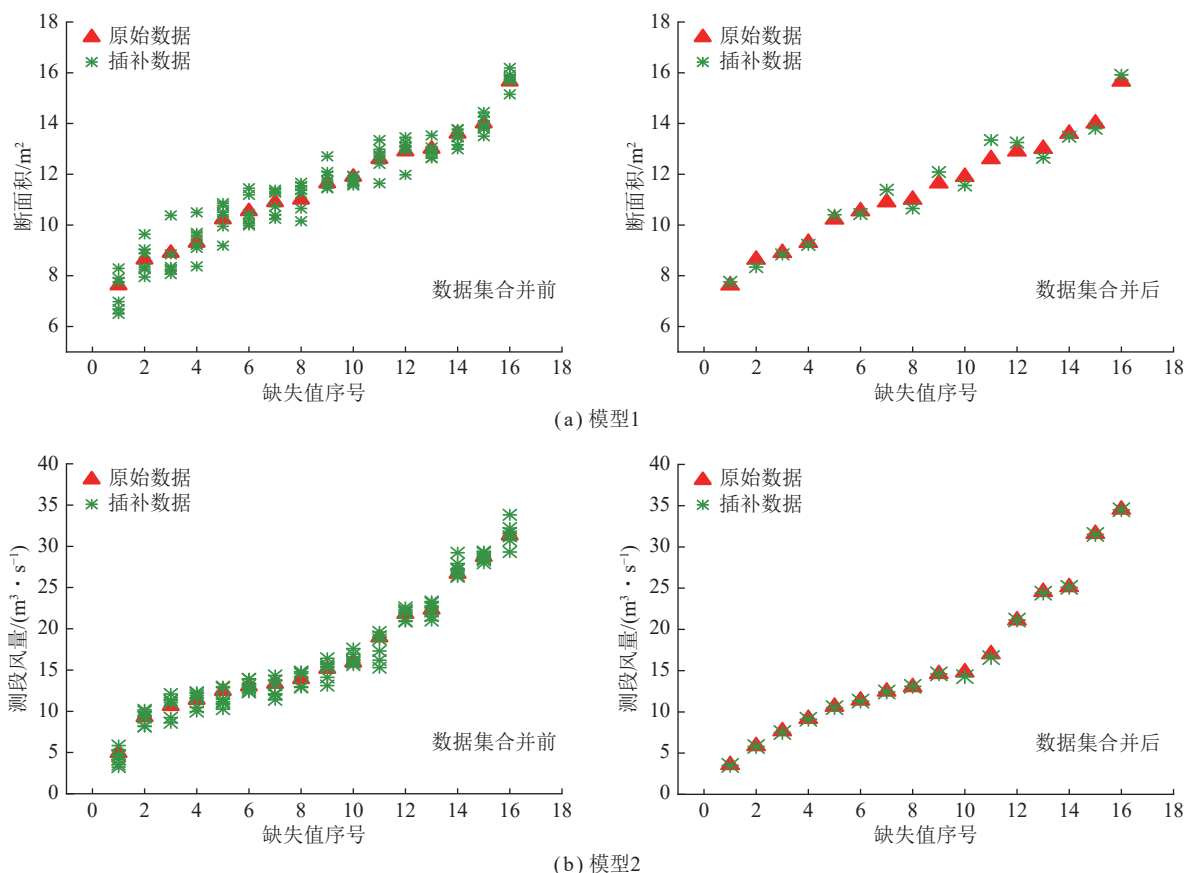


图 5 模型 1、2 数据集合合并前后插补效果

Fig.5 Completion effect of model 1、2 dataset before and after merging

### 3.2.3 试验 3

通过插补函数 `mice()` 中参数 `iterations` 指定数据集在插补过程中的迭代次数。以“百米摩擦风阻”为例,设最终迭代次数为 10 次,编写程序输出每次迭代

效果如图 5 所示。

### 3.2.2 试验 2

通过 `def missing_data_generate()` 自定义函数中参数 `missing_rate` 指定数据缺失率。文中以“断面积”和“测段风量”的数据缺失率为 10%、15% 为例。结果与试验 1 中结果也可做对比试验。

模型 3: 数据缺失列为“断面积”,数据缺失率为 10%,迭代次数为 6。

模型 4: 数据缺失列为“断面积”,数据缺失率为 15%,迭代次数为 6。

模型 5: 数据缺失列为“测段风量”,数据缺失率为 10%,迭代次数为 6。

模型 6: 数据缺失列为“测段风量”,数据缺失率为 15%,迭代次数为 6。

模型 3~6 插补后数据集在分析合并前后的对比效果如图 6 所示。

后数据均值。

模型 7: 数据缺失列为“百米摩擦风阻”,数据缺失率为 10%,迭代次数为 6。

模型 8: 数据缺失列为“百米摩擦风阻”,数据缺

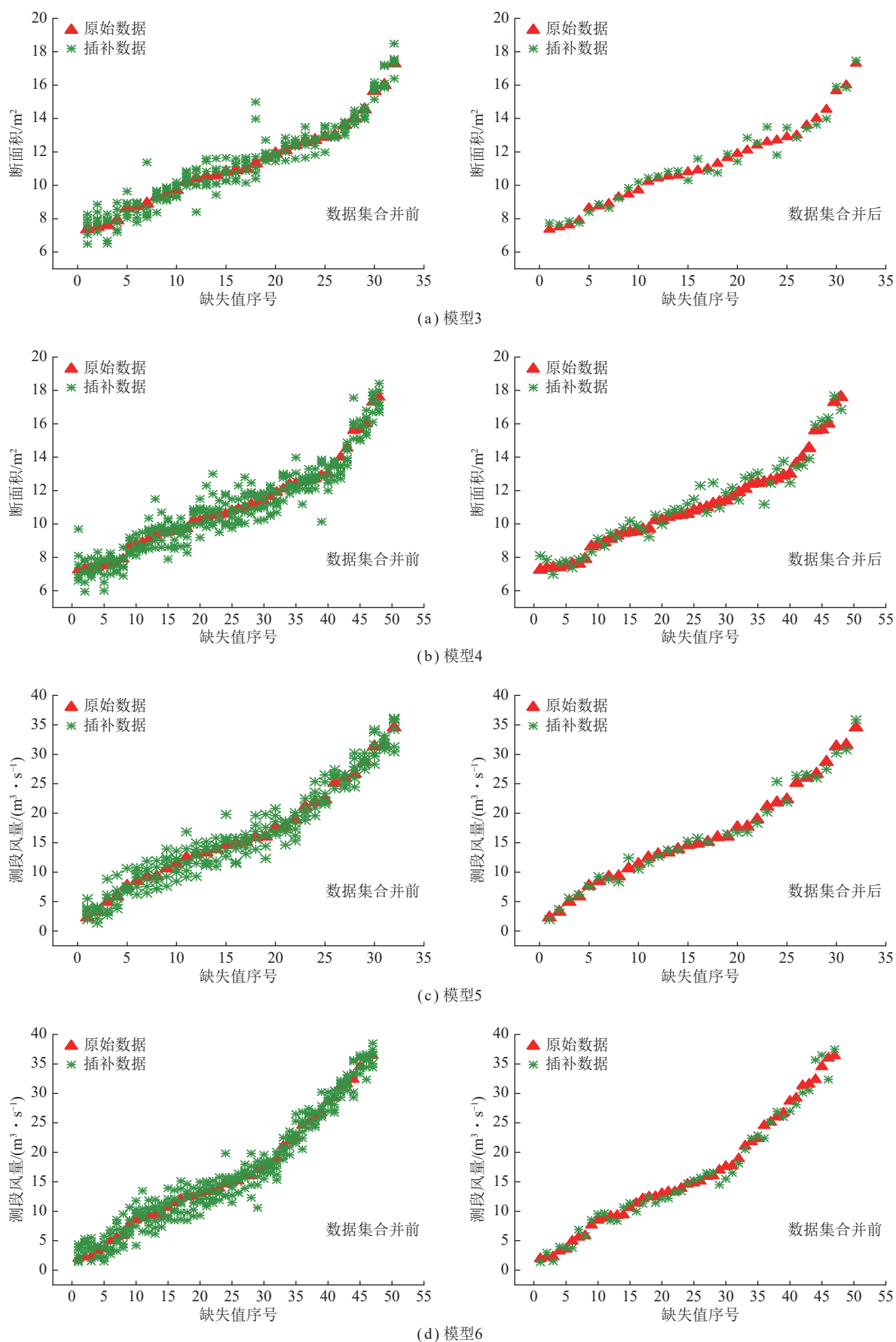


图6 模型3、4、5、6数据集合并前后插补效果

Fig.6 Completion effect of model 3, 4, 5, 6 dataset before and after merging



失率为 10%, 迭代次数为 10。

均值收敛情况如图 7 所示, 设置 datasets=6, 即在插补过程中同时生成 6 个完整数据集, 观察每个完整数据集的均值收敛情况可以得出, 在迭代次数为 6 次时, 均值基本收敛, 一定的迭代次数后, 均值一定收敛。

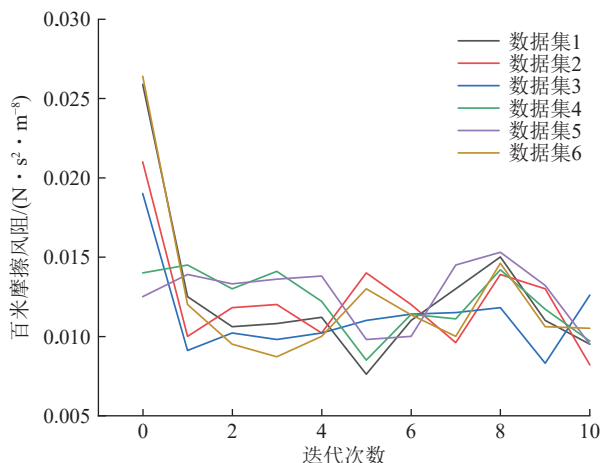


图 7 均值收敛示意

Fig.7 Mean convergence plot

根据 3 组实验, 潞新二矿矿井通风参数缺失数据补全模型评价标准见表 2。

表 2 评价标准

Table 2 Criteria for evaluation

模型编号	均方误差	解释方差	决定系数	平均绝对误差
1	0.044 5	0.988 309	0.990 019	0.180 000
2	0.325 4	0.993 045	0.993 743	0.480 000
3	0.171 1	0.971 029	0.972 544	0.350 000
4	0.305 9	0.943 267	0.955 951	0.470 000
5	1.012 1	0.985 740	0.985 721	0.780 000
6	1.592 4	0.985 635	0.982 897	1.120 000
7	$1.2 \times 10^{-7}$	0.985 583	0.978 552	0.000 535
8	$1.9 \times 10^{-7}$	0.985 328	0.976 334	0.000 897

分析模型 1~6 的插补结果与评价指标可知: “断面积” 缺失率由 5% 增加到 15% 时, 均方误差增加 0.261 4, 解释方差下降 0.045, 决定系数下降 0.035, 平均绝对误差增加 0.29; “测段风量” 数据缺失率由 5% 增加到 15% 时, 均方误差增加 1.267, 解释方差下降 0.007 41, 决定系数下降 0.010 8, 平均绝对误差增加 0.64。结果表明模型中预测数据与原始真实值匹配程度良好, 分散程度相近, 验证了本文插补方法针对不同缺失属性、不同缺失率的插补可行性与准确性。在对“百米摩擦风阻”进行迭代次数试验时, 迭代次数

为 6、10 时, 均值收敛程度良好, 该结果表明在迭代一定次数时, 模型具有收敛性。

## 4 结 论

(1) 根据试验结果分析, 潞新二矿矿井通风参数缺失数据插补模型可以对矿井通风参数缺失数据集进行插补, 验证了模型有效性。

(2) 提出的基于随机森林链式方程多重插补法的矿井通风参数缺失数据插补方法可以轻松处理混合类型的数据, 自主学习参数之间的相关性, 使用随机森林的方法来预测缺失值, 不需要太多的超参数调整, 对矿井通风参数进行“黑盒”插补, 降低了插补复杂性, 提高了插补精度。

(3) 模型在试验中为了模拟实际数据缺失的不确定性, 对缺失数据的属性与缺失率进行了不同的试验, 结果表明, 随着数据缺失率的升高, 插补模型依然保持较高的插补精度, 说明插补模型具有一定的稳定性。对初始完整数据集插补时的迭代次数进行分析时, 发现经过一定的迭代次数后, 相关性一定会收敛。

## 参考文献(References):

- [1] 卢新明, 尹红. 矿井通风智能化理论与技术[J]. 煤炭学报, 2020, 45(6): 2236–2247.  
LU Xinming, YIN Hong. The intelligent theory and technology of mine ventilation[J]. Journal of China Coal Society, 2020, 45(6): 2236–2247.
- [2] 周福宝, 魏连江, 夏同强, 等. 矿井智能通风原理、关键技术及其初步实现[J]. 煤炭学报, 2020, 45(6): 2225–2235.  
ZHOU Fubao, WEI Lianjiang, XIA Tongqiang, et al. Principle of intelligent ventilation in mine, key technology and its preliminary implementation[J]. Journal of China Coal Society, 2020, 45(6): 2225–2235.
- [3] 周福宝, 辛海会, 魏连江, 等. 矿井智能通风理论与技术研究进展[J]. 煤炭科学技术, 2023, 51(1): 313–328.  
ZHOU Fubao, XIN Haihui, WEI Lianjiang, et al. Research progress of mine intelligent ventilation theory and technology[J]. Coal Science and Technology, 2023, 51(1): 313–328.
- [4] HUANG Haojian. An efficient method to classification with missing data[J]. Academic Journal of Computing & Information Science, 2021, 4(8): 63–66.
- [5] 赵志文, 高敏. 缺失数据下随机系数自回归模型的参数估计[J]. 统计与决策, 2022, 38(1): 16–20.  
ZHAO Zhiwen, GAO Min. Parameter estimation of random coefficient autoregressive model with missing data[J]. Statistics & Decision, 2022, 38(1): 16–20.
- [6] 余嘉茵, 何玉林, 崔来中, 等. 针对大规模数据的分布一致缺失值插补算法[J]. 清华大学学报(自然科学版), 2023, 63(5): 740–753.  
YU Jiayin, HE Yulin, CUI Laizhong, et al. A distributionally consistent missing value interpolation algorithm for large-scale data[J].



- Journal of Tsinghua University (Natural Science Edition), 2023, 63(5): 740–753.
- [7] MA X, HAN Y, QIN H, et al. KNN data filling algorithm for incomplete interval-valued fuzzy soft sets[J]. *International Journal of Computational Intelligence Systems*, 2023, 16(1): 1–15.
- [8] LU Y. Analysis of image restoration based on EM algorithm[J]. *Journal of Physics: Conference Series*, 2022, 2242(1): 012045.
- [9] 王凤梅, 胡丽霞. 一种基于近邻规则的缺失数据插补方法[J]. 计算机工程, 2012, 38(21): 53–55, 62.
- WANG Fengmei, HU Lixia. A missing data imputation method based on neighbor rules[J]. *Computer Engineering*, 2012, 38(21): 53–55, 62.
- [10] LYDERSEN Stian. Multiple imputation of missing data [J]. *Tidsskrift for den Norske laegeforening: Tidsskrift for praktisk medicin, nyraekke*, 2022, 142(2): doi: 10.4045/tidsskr.21.0772.
- [11] 陈娟, 王献雨, 罗玲玲, 等. 缺失值插补效果: 机器学习与统计学习的比较[J]. 统计与决策, 2020, 36(17): 28–32.
- CHEN Juan, WANG Xianyu, LUO Lingling, et al. Comparison of machine learning and statistical learning in the imputation of missing values[J]. *Statistics & Decision*, 2020, 36(17): 28–32.
- [12] 刘露诗. 基于机器学习与缺失值插补技术的海底硫化物成矿定量预测 [D]. 长春: 吉林大学, 2022: 37–43.
- LIU Lushi. Quantitative prediction of seafloor sulfide mineralization based on machine learning and missing value interpolation technique [D]. Changchun: Jilin University, 2022: 37–43.
- [13] AWAN S E, BENNAMOUN M, SOHEL F, et al. Imputation of missing data with class imbalance using conditional generative adversarial networks[J]. *Neurocomputing*, 2021, 453: 164–171.
- [14] 刘泽功. 利用通风系统调风和阻力测定求算复杂通风网路分支风阻[J]. 煤矿安全, 1991, 22(1): 1–7.
- LIU Zegong. Calculation of branch wind resistance of complex ventilation network by air regulation and resistance measurement of ventilation system[J]. *Safety in Coal Mines*, 1991, 22(1): 1–7.
- [15] 司俊鸿, 陈开岩. 基于 Tikhonov 正则化的矿井通风网络测风求阻法[J]. 煤炭学报, 2012, 37(6): 994–998.
- SI Junhong, CHEN Kaiyan. Measuring airflow & evaluating resistance model of the mine ventilation network based on Tikhonov regularization[J]. *Journal of China Coal Society*, 2012, 37(6): 994–998.
- [16] 邓立军. 矿井通风阻力系数反演研究 [D]. 阜新: 辽宁工程技术大学, 2014: 41–55.
- DENG Lijun. Research on inversion of mine ventilation resistance coefficient [D]. Fuxin: Liaoning University of Engineering and Technology, 2014: 41–55.
- [17] 刘剑, 李雪冰, 陈廷凯, 等. 矿井定常湍流脉动对通风阻力测试影响的理论分析[J]. 中国安全生产科学技术, 2016, 12(5): 22–25.
- LIU Jian, LI Xuebing, CHEN Tingkai, et al. Theoretical analysis on influence of steady turbulence fluctuation on ventilation resistance measurement in mine[J]. *Journal of Safety Science and Technology*, 2016, 12(5): 22–25.
- [18] 李雨成, 李俊桥, 邓存宝, 等. 基于角联子网的风量反演风阻病态改良算法[J]. 煤炭学报, 2019, 44(4): 1147–1153.
- LI Yucheng, LI Junqiao, DENG Cunbao, et al. Improved algorithm of air quantity calculating resistance based on diagonal subnetwork [J]. *Journal of China Coal Society*, 2019, 44(4): 1147–1153.
- [19] 倪景峰. 矿井通风仿真系统可视化研究 [D]. 阜新: 辽宁工程技术大学, 2004: 93–105.
- NI Jingfeng. Research on the visualization of mine ventilation simulation system [D]. Fuxin: Liaoning University of Engineering and Technology, 2004: 93–105.
- [20] GRUND S, LÜDTKE O, ROBITZSCH A. Multiple imputation of missing data in multilevel models with the R package mdmb: A flexible sequential modeling approach[J]. *Behavior Research Methods*, 2021, 53(6): 2631–2649.
- [21] 温廷新, 苏焕博. 基于链式多重插补的 WOA-ELM 煤与瓦斯突出预测模型[J]. 中国安全生产科学技术, 2022, 18(7): 68–74.
- WEN Tingxin, SU Huanbo. WOA-ELM prediction model of coal and gas outburst based on multiple imputation by chained equations[J]. *Journal of Safety Science and Technology*, 2022, 18(7): 68–74.
- [22] 倪景峰, 李振, 乐晓瑞, 等. 基于随机森林的阻变型通风网络故障诊断方法[J]. 中国安全生产科学技术, 2022, 18(4): 34–39.
- NI Jingfeng, LI Zhen, LE Xiaorui, et al. Resistance variant fault diagnosis method of ventilation network based on random forest[J]. *Journal of Safety Science and Technology*, 2022, 18(4): 34–39.