庞奥康, 容芷君, 但斌斌, 等. 基于改进 KNN-SVR 的拉矫力缺失数据插补方法[J]. 智能计算机与应用,2026,16(6):10-16. DOI:10.20169/j.issn.2095-2163.24102206

基于改进 KNN-SVR 的拉矫力缺失数据插补方法

庞奥康[1,2], 容芷君[1,2], 但斌斌[1,2], 刘 洋[3], 李学霖[1,2]

(1 武汉科技大学 冶金装备及其控制教育部重点实验室, 武汉 430081; 2 武汉科技大学 机械传动与制造工程湖北省重点实验室, 武汉 430081; 3 武汉钢铁有限公司设备管理部, 武汉 430083)

摘 要: 拉矫力监测数据在采集、传输等环节出现随机缺失使连铸机扇形段状态分析出现误差,为保证拉矫力数据的完整性,引用区间半分法和滑动窗口技术改进 K 近邻支持向量回归(KNN-SVR)的缺失数据插补方法。首先,判断拉矫力特征值的方差和阈值提取正常浇铸工况数据,对提取后的数据进行去噪处理,确保数据插补的质量和可用性;其次,采用区间半分法和滑动窗口优化 KNN-SVR 方法,利用多维拉矫力间的时空相关性,捕捉时序数据全局和局部特征,对不同缺失区间数据进行插补;最后,以某钢厂所采集的拉矫力数据为例,以多种模型评价指标对模型有效性进行评估。结果表明:该方法形成的完整数据集与原始数据集相比具有很好的相似性,在保证数据插补精度的同时降低计算时间,具有较高的工程应用价值。

关键词: 扇形段拉矫力; 数据插补; 时序数据; K 最近邻; 支持向量回归

中图分类号: TF777 **文献标志码**: A **文章编号**: 2095-2163(2026)06-0010-07

Research on imputation methods for tension straightening force missing data based on improved KNN-SVR

PANG Aokang[1,2], RONG Zhijun[1,2], DAN Binbin[1,2], LIU Yang[3], LI Xuelin[1,2]

(1 Key Laboratory of Metallurgical Equipment and Control of Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China; 2 Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China; 3 Equipment Management Department, Wuhan Iron and Steel Co., Ltd., Wuhan 430083, China)

Abstract: The random missing of the tension force monitoring data in the collection and transmission makes the error in the state analysis of the continuous casting machine sector, in order to ensure the integrity of the tension force data, the interval half-split method and the sliding window technology are used to improve the missing data interpolation method of KNN-SVR. Firstly, the variance and threshold of the eigenvalues of the straightening force are judged to extract the normal casting condition data, and the extracted data are denoised to ensure the quality and usability of the data interpolation; secondly, the interval half-split method and the sliding window are used to optimize the KNN-SVR method, and the spatial and temporal correlations between the multi-dimensional straightening force are utilized to capture the global and local features of the temporal data, and the data of the different missing intervals are interpolated; finally, the model validity is assessed with various model evaluation indexes using the tension force data collected from a steel mill as an example. The results show that the complete dataset formed by this method has good similarity with the original dataset. Therefore, the method has high engineering application value through ensuring the interpolation accuracy and reducing the computation time at the same time.

Key words: tension force of the sector segment; data imputation; time series data; KNN; Support Vector Regression

0 引 言

在过去的几十年里,连续铸造是熔融态钢液经强制冷却形成铸坯的过程,是现代钢铁生产流程的核心环节。扇形段机组是铸坯成型的重要部位,是决定铸坯质量和保证生产的关键[1],拉矫力的变化是扇形段运行状态最直观的表现,拉矫力数据缺失所造成的后果也最为严重,恢复难度也远高于其他

基金项目: 国家自然科学基金(51475340); 湖北省重点研发计划(2022BAA059)。

作者简介: 庞奥康(1999—),男,硕士研究生。主要研究方向:信号分析与处理。

通信作者: 容芷君(1974—),女,博士,教授。主要研究方向:产品创新与系统设计,信号分析。Email:15671633803@163.com。

收稿日期: 2024-10-22

数据类型。拉矫力数据一旦缺失，不仅丧失了原始时间位置信息，还会造成扇形段工况数据波动信息残缺。故为确保设备安全可靠运行，利用多传感器监测设备的运行状态[2-3]。由于数据采集设备故障、数据传输过程中的网络中断等现象[4]，据此带来的数据不完整问题对数据质量产生负面影响，导致产品质量不稳、生产成本增加[5]。为解决上述问题，对扇形段拉矫力的缺失数据进行处理则尤显重要。

在工业领域，监测数据的缺失较为普遍，对依赖完整而高质量的数据集构造健康监测模型造成阻碍[6]。现有的插补方法主要分为基于概率的方法[7-8]、基于插值的方法[9-10]、基于相似性的方法[11-13]及基于回归的方法[14-17]。其中，基于插值的方法，使用多项式对测量值进行时间依赖性建模插补缺失值，结构简单、易实施，但在不均匀数据分布下易引入偏差。基于相似性的方法(如 k-最近邻(KNN)和均值插补)根据特征相关性计算缺失样本和完整样本之间的相似性，然后使用高度相似的完整样本均值来插补缺失值。然而，忽略时间相关性会导致信息丢失和精度降低。基于回归的方法，可以系统地考虑特征相关性和时间相关性，得益于对高度非线性问题优异的建模能力，在提取复杂的相关特征方面表现出色，从而有较提升了插补精度。文献[18]总结了 KNN 的理论基础和在多个领域的应用情况，并指出了需要进一步研究和改进的方向。文献[19]提出了迭代 KNN 插补方法，解决了数值属性和分类属性问题。文献[20]提出了基于均值漂移聚类和海洋捕食者算法的参数自适应支持向量回归(SVR)方法，适用于高速滚齿工艺参数预测等领域。文献[21]结合了 KNN 在利用历史数据信息方面的优势和 SVR 在处理非线性问题方面的强大能力，提高航班滑出时间预测的准确性。

基于以上分析，针对扇形段拉矫力电流信号具有非线性、时变性的特点，现有多数插补模型基于原始数据训练，限制了预测精度，且计算复杂度高、耗时长，难以在准确性和效率之间取得平衡。本文提出一种基于改进 KNN-SVR 的拉矫力缺失数据插补方法。基于专家领域挖掘拉矫力数据不同维度上时序关系，利用数据的分布特征、趋势变化、数据间幅值变化范围及持续时长识别缺失数据；引用区间半分法和滑动窗口技术优化 KNN 识别相似数据点，提供精确的初始估计和高质量的数据基础，结合 SVR 的强大非线性建模能力，实现更准确和高效的缺失数据插补；应用某钢厂扇形段拉矫力真实历史数据集进行实验，验证所提方法的有效性。

1　拉矫力基础数据现状分析

连铸机由 14 台扇形分段组成，不同扇形段对总拉矫力的影响各异。浇铸周期分为 4 个阶段：开始浇铸、正常浇铸、换中包和停止浇铸。每台扇形段由变频电机驱动，通过电机扭矩推算每台扇形段的拉矫力[22]，计算公式如下：

$$F_{\text{refi}} = M_{\text{refi}} \times i_i / r_i \tag{1}$$

其中，F_{refi} 表示扇形段拉矫力；M_{refi} 表示电机的实际扭矩；i_i 表示齿轮齿数比；r_i 表示驱动辊直径。

扇形段拉矫力属于多维时序数据，记扇形段个数为 p，即有 p 个变量。将扇形段拉矫力按采集的时间顺序排列形成拉矫力序列 $T_i = [t_{i1}, t_{i2}, \cdots, t_{im}]$，其中 $i = 1, 2, \cdots, p$。n 为拉矫力序列的长度，即 n 个采样数据为一组，多维拉矫力数据集可以表示为：

$$\boldsymbol{I} = [I_1, I_2, \cdots, I_p]' = \begin{bmatrix} I_{11} & \cdots & I_{1n} \\ \vdots & \ddots & \vdots \\ I_{p1} & \cdots & I_{pn} \end{bmatrix} \tag{2}$$

由此便可得出扇形段原始数据集。

原始扇形段拉矫力数据集示例见表 1。从表 1 中可以看出时序拉矫力数据具有多工况、连续采样、时变性等特点，样本数据每 5 s 从 14 个扇形段中采集一次数据，其中主要为正常数据，也掺杂一定量缺失数据和异常数据，异常数据主要包含故障数据、突变数据等情况。本文利用序列间相关性信息，对异常数据进行识别及处理，保证拉矫力数据的精准性。

表 1　原始扇形段拉矫力数据集示例

Table 1　Example of the original sector segment tension and correction force dataset

Created Time	S1M011	…	S2M142	S1 JzLength	S1Speed	S1Lali	S2 JzLength	S2Speed	S2Lali
2022/05/24 23:09:02	31.32	…	18.39	2 329.1	1.199 06	764.548	2 322.30	1.198 81	859.645
2022/05/24 23:09:07	30.85	…	26.16	2 329.2	1.198 79	768.034	2 322.40	1.199 08	925.216
2022/05/24 23:09:12	29.27	…	32.99	2 329.3	1.198 80	768.034	2 322.50	1.198 53	925.216
2022/05/24 23:09:17	26.88	…	34.66	2 329.4	1.199 32	748.990	2 322.60	1.199 22	946.901
2022/05/24 23:09:22	27.46	…	32.37	2 329.5	1.199 72	748.990	2 322.71	1.199 49	946.901
…	…	…	…	…	…	…	…	…	…

2 数据预处理

2.1 提取正常浇铸工况数据

图 1 显示了拉矫力在不同浇铸阶段的变化:开始浇铸时拉矫力波动上升;正常浇铸时拉矫力波动较小;换中包时拉矫力先升后降;停止浇铸时拉矫力迅速降至零。异常和数据缺失多发生在正常浇铸阶段,表现为拉矫力的异常变化。故提取正常浇铸阶段,研究拉矫力缺失数据的插补方法。

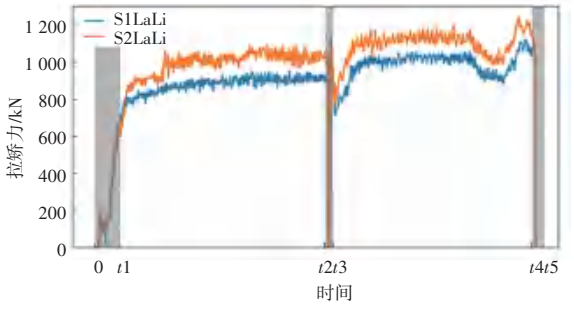


图 1 时序拉矫力一个工作周期内的变化

Fig. 1 The variation of temporal tension and correction force within one working cycle

提取正常浇铸阶段的拉矫力数据方法如下:

(1)设定滑动窗口大小 n, 对拉矫力数据进行加窗处理。

(2)计算窗口内总拉矫力和拉速的方差。

(3)判断总拉矫力的最小值是否大于等于 600 kN,拉速方差是否小于等于 0.5,总拉矫力方差是否小于等于 800,条件全部满足则记录对应时间标签,提取正常浇铸阶段的数据。

以上所有参数均是通过大量历史数据统计而得。

2.2 拉矫力数据去噪处理

受结晶器振动、辊子转动等影响,拉矫力值存在一定的波动。Savitzky-Golay 滤波算法能有效去噪并保留信号特征,提高数据插补准确性。选择合适的窗口大小和多项式阶数对滤波效果很关键,以确保数据能真实反映连铸过程。SG 滤波算法不同参数的对比如图 2 所示,选用多组参数组合进行对比分析,最终通过标准差确定当窗口大小为 15、阶数为 3 时,去噪效果最佳。

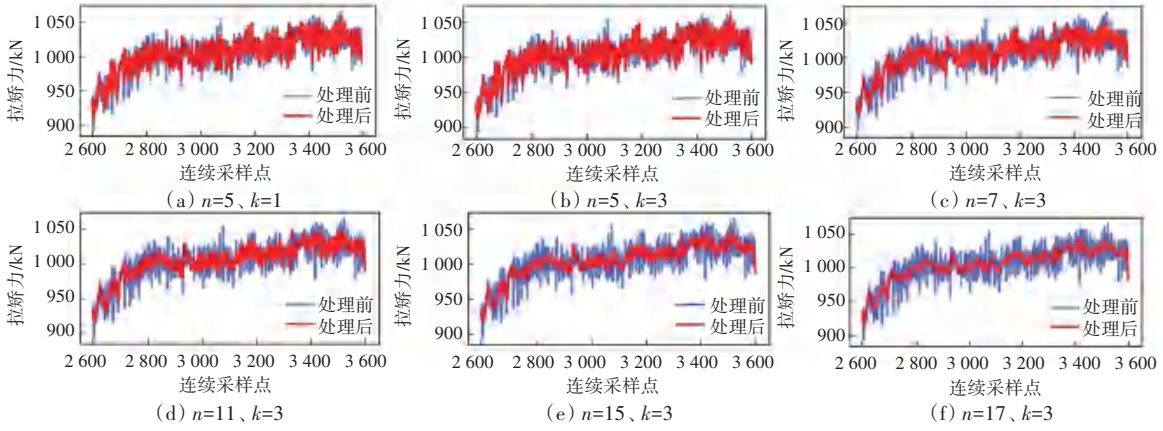


图 2 SG 滤波算法不同参数的对比图

Fig. 2 Comparison of SG filtering algorithm with different parameters

2.3 拉矫力缺失数据辨识

在去除噪声后,通过识别当前时刻拉矫力及拉速与历史时刻拉矫力及拉速是否存在明显的差异来判断扇形段的运行状况。通过分析数据的趋势、幅值变化及持续时间来识别缺失数据。辨识缺失数据如图 3 所示,比较当前拉矫力及拉速和历史数据前 10 个样本拉矫力及拉速均值的差异,如果两流拉矫力趋势相同且幅值变化稳定,则认为设备运行正常。如果数据异常持续超过 2 min,则视为重要故障;若不超过 2 min,则视为可剔除的无用信号。

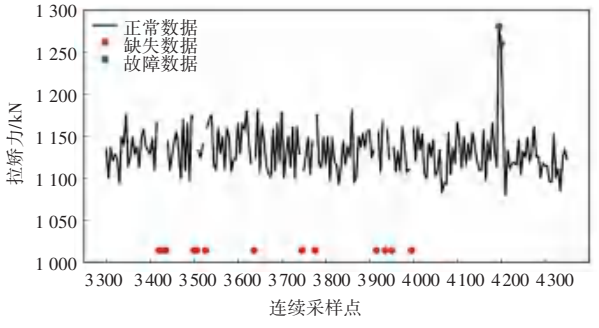


图 3 辨识缺失数据图

Fig. 3 Diagram of missing data identification

具体步骤如下：

(1)计算历史拉矫力数据前 10 个样本数据拉矫力及拉速的均值。

(2)判断当前时刻拉矫力偏差是否超过样本拉矫力均值的 15%，若在范围内，则说明设备运行正常，通过判断是否存在空值输出正常数据标签和缺失数据标签；若超出范围，判断当前时刻的拉速是否超过样本拉速均值的 20%。若在范围内，同上；若超出范围，转入下一步。

(3)判断持续周期是否在 2 min 内变动，若超过 2 min，认为蕴含着重要的异常信息，标记为故障信号并保存，反之剔除异常点形成空白数据点。

以上所有步骤参数均通过大量历史数据统计获得。

3　基于区间半分法和滑动窗口改进的 KNN-SVR 的缺失数据插补

扇形段拉矫力电流信号具有非线性、时变性的特点，现有多数插补模型基于原始数据训练，限制了预测精度，且计算复杂度高、耗时长，难以在准确性和效率之间取得平衡。基于以上问题，结合 k 最近邻(kNN)方法和支持向量回归(SVR)方法来插补缺失值。将样本从原始空间映射到更高维的特征空间，使样本在这个特征空间内线性可分。首先，利用 k 近邻匹配算法，在完备数据集中搜索与待插补数据块中时刻 t 状态向量相近的 k 个完备数据块，基于 k 个完备数据块，构建训练集，对训练集进行归一化处理。其次，应用 SVR 来进行回归预测，预测时考虑了与目标值的偏差，通过使用径向基核函数来映射数据到高维空间，在该空间中找到最优的拟合曲面。但当数据集过大时，kNN-SVR 需要计算每个样本与其它样本之间的距离，计算成本较高，且对所有邻居赋予相同的权重，这可能会受到异常值的影响。更重要的是，kNN-SVR 主要依赖于全局数据分布来预测缺失值，未充分考虑数据的时间性和空间性，会限制其在数据插补上的精确性。

根据上述 kNN-SVR 算法的问题，引入区间半分法和滑动窗口算法解决计算量大且精度不足的问题。区间半分法通过动态调整搜索区间，能够更精确地定位缺失数据的可能取值范围，结合滑动窗口技术，可以在数据的局部区域内进行插补，利用邻近数据的相关性，从而提高插补值的准确性。改进后的 KNN-SVR 算法步骤如下。

(1)计算相似度，选择前 k 个最相似的数据。计算公式为：

$$\mathrm{Sim}(i,j)=\frac{1}{N}\sum_{k=1}^{N}\frac{1}{1+|x_{ik}-x_{jk}|} \tag{3}$$

其中，$\mathrm{Sim}(i,j)$ 表示相似度；x_{ik} 表示目标数据 i 在条件 k 下的表达值；x_{jk} 是参考数据 j 在条件 k 下的表达值；N 表示条件的总数。

(2)区间半分法初步定位。设置合适的半区间最小长度 l^* 和相似度阈值 θ^*。

初步定位相似数据切换点，假设共有 s 个切换点，标记为 $K=\{k_0,k_1,k_2,\cdots,k_s,k_{s+1}\}$，其中 $k_0=1$，$k_{s+1}=N$；

将相邻的时间序列段进行合并处理，i 从 0 开始，计算序列 $X_{k_i,k_{i+1}}=\{x(k),k=k_i,\cdots,k_{i+1}\}$ 和序列 $X_{k_{i+1},k_{i+2}}=\{x(k),k=k_{i+1},\cdots,k_{i+2}\}$ 之间的相似度指标。若两序列段相似，移除切换点 k_{i+1}，对切换点集合 K 重新整理，令 $s=s-1$，重复此过程，直到无法再合并相邻序列段为止。

(3)滑动窗口优化切换点。利用滑动窗细化状态转换点，确定状态转换点集合 $K=\{k_0,k_1,k_2,\cdots,k_s,k_{s+1}\}$。考虑序列 $X_{k_i,k_{i+2}}$，其中 k_i 和 k_{i+2} 是边界点，而 k_{i+1} 是该序列段内识别出的转换点。序列 $X_{k_i,k_{i+2}}$ 按照时间顺序被划分为 3 个部分：$X^l_{k_i,k_{i+1}-l^*/2}$、$X^m_{k_{i+1}-l^*/2,k_{i+1}+l^*/2}$、$X^r_{k_{i+1}+l^*/2,k_{i+2}}$，上标‘$l$’、‘$m$’和‘$r$’分别表示左、中、右三个子序列。当 3 个子序列的长度都超过 l^* 时，计算 X^l 和 X^m 之间的夹角 $\theta_{l,m}$，以及 X^m 和 X^r 之间的夹角 $\theta_{m,r}$。

① 如果满足 $\theta_{l,m}>\theta_{m,r}$，且 $\theta_{l,m}>\theta^*$，真实的切换点在 $k_{i+1}-l^*/2$ 的左侧。以 ω^* 为步长向左平移，子序列变为 X^l，X^m 和 X^r，判断是否依然满足 $\theta_{l,m}>\theta_{m,r}$，且 $\theta_{l,m}>\theta^*$。

若不满足或某段子序列长度为 l^*，则停止右移，最新的 $k_{i+1}-l^*/2$，作为更新的切换点；

② 同理，若 $\theta_{l,m}>\theta_{m,r}$，且 $\theta_{l,m}>\theta^*$，则真实的切换点在 $k_{i+1}-l^*/2$ 的右侧。向右移动的滑动窗口加区间半分法如图 4 所示，以 ω^* 为步长向右平移，子序列变为 X^l，X^m 和 X^r，判断是否依然满足 $\theta_{l,m}>\theta_{m,r}$，且 $\theta_{l,m}>\theta^*$，若不满足或某段子序列长度小于 l^*，则停止右移，最新的 $k_{i+1}-l^*/2$ 作为更新的切换点。

令 $i=i+1$，重复步骤，直至所有的序列被遍历完成。

(4)构建并训练 SVR 模型。使用选定的 K 个数据作为训练数据，构建 SVR 模型。SVR 模型的优化问题为：

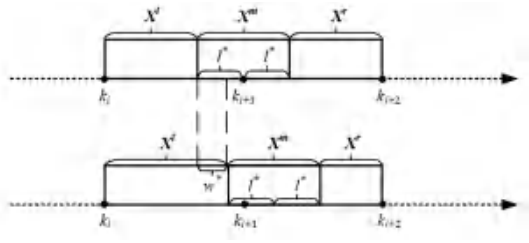


图 4 向右移动的滑动窗口加区间半分法

Fig. 4 Sliding window moving to the right with interval semidivision method

$$\min_{\alpha} \frac{1}{2}\sum_{i,j=1}^{K}\alpha_i\alpha_j y_i y_j K(x_i,x_j) - \sum_{i=1}^{K}\alpha_i y_i \quad (4)$$

约束条件为：

$$0 \leqslant \alpha_i \leqslant C, i = 1,2,\cdots,K \quad (5)$$

其中，C 表示正则化参数，K 表示核函数。通过求解上述优化问题，得到拉格朗日乘子α_i。

将求得的拉格朗日乘子代入回归函数：

$$f(x) = \sum_{i=1}^{K}\alpha_i y_i K(x_i,x) + b \quad (6)$$

其中，b 表示偏置项。

(5)缺失值插补。对于缺失表达水平值对应的位置 x_j，代入回归函数 $f(x)$ 计算缺失值的插补值：

$$f(x_j) = \sum_{i=1}^{K}\alpha_i y_i K(x_i,x_j) + b \quad (7)$$

区间半分法通过强制的二分策略，有效避免了复杂的搜索过程，初步确定状态转换点。在此基础上，滑动窗口进一步优化了状态转换点的定位。由于区间半分法已经将真实状态转换点与识别出的状态转换点之间的偏差限制在 l^* 以内，滑动窗口的搜索范围被大幅缩小。通过将 2 种方法相结合，改进后的 KNN-SVR 的插补方法同时保证了效率和精度。

4 效果验证

为验证所提出方法的有效性，提高扇形段机组健康检测效果，以一个周期内的连铸机扇形段实际运行数据为依据，通过分析基于现场专家经验的正常浇铸工况提取方法，采用时域特征中的方差特征提取正常工况数据。考虑到噪声对拉矫力异常幅值波动的干扰，分别使用小波阈值去噪与 SG 滤波去噪进行对比，结果如图 5 所示。由图 5 可知，经过 SG 滤波去噪处理之后的数据比原来更加光滑，并且还保持了数据的有效信息。

为了真实反映实际工业场景的数据缺失情况，对预处理后的正常工况数据进行筛选，选用不包含缺失的数据集，通过数据异常清洗与人工删除设置所需的缺失数据，随机分为不同区间的缺失数据类型，如图 6 所示。

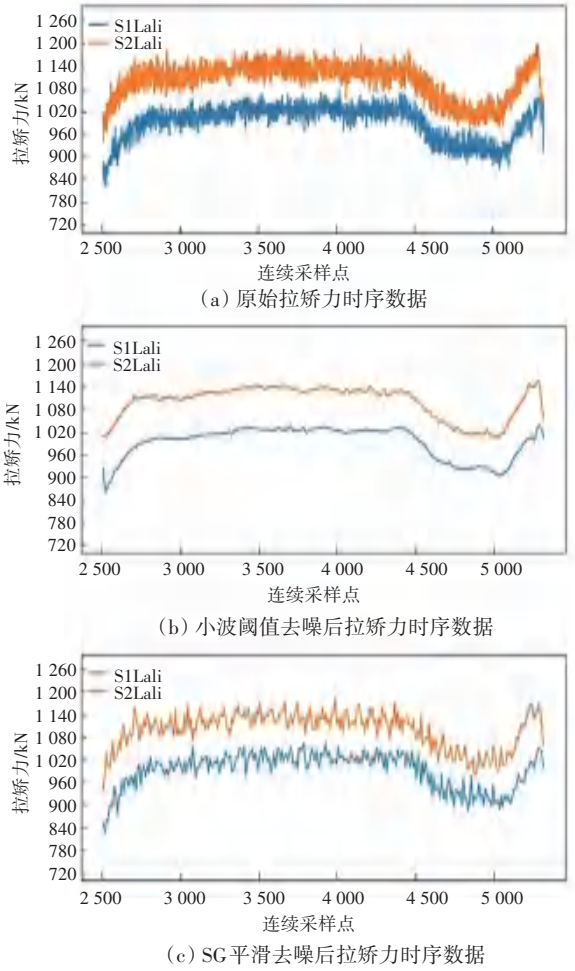


图 5 去噪前后效果对比

Fig. 5 Comparison of pre-and pos-denoising effects

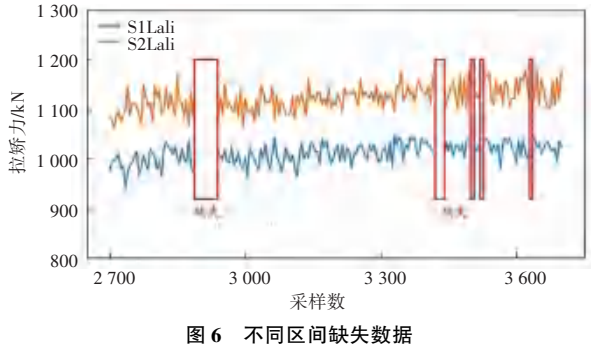


图 6 不同区间缺失数据

Fig. 6 Missing data in different intervals

针对不同的缺失区间，根据拉矫力的自相似性，使用改进后的 kNN-SVR 实现缺失数据插补，通过 kNN 选择相似数据，基于区间半分法结合滑动窗口将相似数据集根据两流拉矫力的方差划分为多个完备数据块，选择方差最大的特征作为分割维度，划分完备数据块，将缺失区间邻近的数据样本作为训练数据，建立回归模型并训练，最后对缺失部分以模型

预测结果进行插补。结果如图 7 所示。由图 7 可知，所提算法在不同缺失区间仍可进行插补，保证了算法在多场景下的可用性。

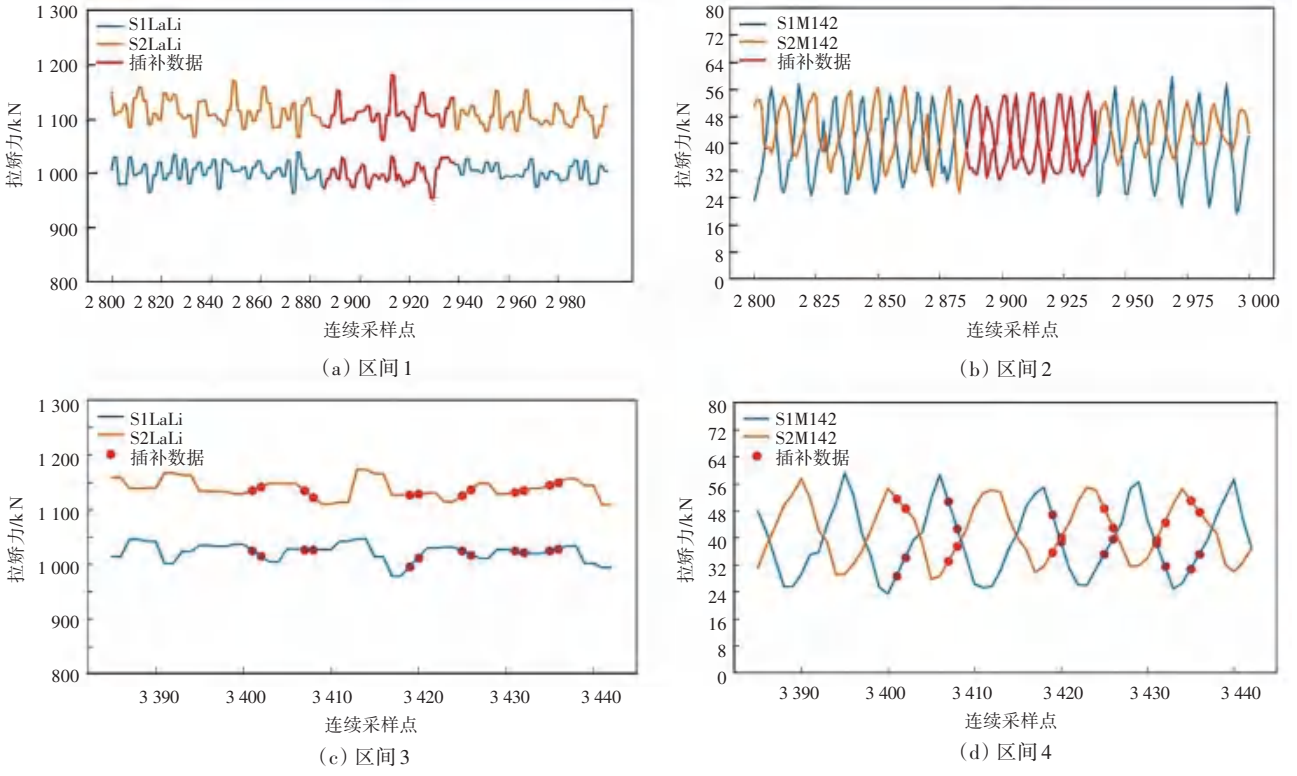


图 7　不同缺失区间数据插补

Fig. 7　Interpolation of data in different missing intervals

为对比说明提出方法的有效性，采用 kNN、SVR、kNN-SVR、改进后的 kNN-SVR、长短期记忆网络(LSTM)5 种恢复方法对缺失数据进行恢复。插补精度通过计算均方根误差(RMSE)、平均绝对百分比误差(MAPE)、决定系数(R^2)和数据恢复所需时间 t 等量化指标进行评估。不同方法插补结果对比如图 8 所示。

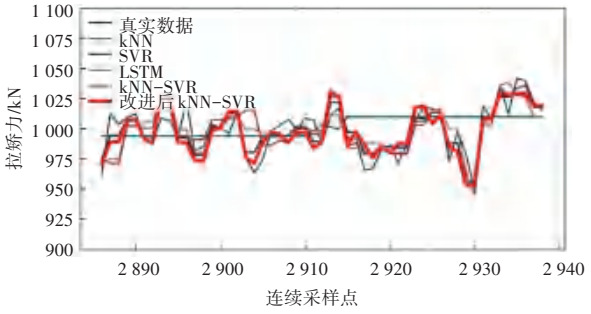


图 8　不同方法插补结果对比

Fig. 8　Comparison of interpolation results using different methods

由图 8 可知，尝试利用不同的方法插补局部缺失值，由于缺失区间过大，无法捕捉到数据的整体分布特性，对于数据波动与突变不能有效地进行展现，kNN 和 SVR 的插补效果最差；kNN-SVR 算法与 LSTM 算法要优于 kNN 和 SVR 算法，但也存在较多不相似数据干扰，插补后数据与真实数据相差较大；从总体上看，改进后的 kNN-SVR 算法插补效果最佳，所得插补数据与真实数据基本一致，仅存在微小跳变。

不同方法误差指标结果见 2。根据表 2 可以看出，改进后的 kNN-SVR 数据插补方法所得的插补效果最好，提出方法的 RMSE、MAPE 和时间最小，R^2 最大。上述结果进一步验证了所提出的数据插补方法的有效性，可以辅助连铸机扇形段管理者提高拉矫力数据质量，开展更精准的相关数据研究。

表 2　不同方法误差指标对比

Table 2　Comparison of error indicators for different methods

插补方法	RMSE	MAPE	R^2	时间/s
kNN	18.774	1.405	-0.174	3.36
SVR	10.928	0.865	0.602	10.71
kNN-SVR	9.955	0.689	0.670	5.85
改进 kNN-SVR	4.679	0.375	0.872	1.46
LSTM	8.420	0.681	0.764	4.37

5　结束语

基于扇形段拉矫力缺失数据的分析，可以得到

如下结论：

(1)考虑拉矫力数据间时序分布相似性和数据间趋势一致性等特性，总结分析了基于现场专家经验的拉矫力异常数据处理方法。识别并剔除异常数据，显著提升基础数据质量，完善数据可用信息，对连铸机的稳定、安全、可靠运行具有重要意义。

(2)结合钢厂实际数据情况，采用所提方法对随机去除的监测数据进行恢复，使用区间半分法和滑动窗口改进 KNN-SVR 算法，根据缺失数据邻近数据样本的完整程度选择相似数据，在保证数据插补精度的同时降低计算时间。结果表明，面对不同缺失类型的数据，采用不同方法对恢复缺失数据进行对比，所提插补方法处理后的结果为连铸机扇形段辊子健康检测等提供了可靠的数据基础，具有较好的工程应用价值。

参考文献

[1] 白宇. 连铸过程检测信号的频域分析及应用方法研究[D]. 大连：大连理工大学, 2019.

[2] 都胜朝, 谭志强, 章勤奋. 连铸机扇形段在线监测与诊断技术的研究与应用[J]. 武钢技术, 2017, 55(6)：63-66.

[3] 邹康, 容芷君, 但斌斌, 等. 基于拉矫力电流信号的连铸机状态评估方法研究[J]. 武汉科技大学学报(自然科学版), 2021, 44(3)：176-181.

[4] 曹宸铭. 结构健康监测缺失数据概率分布信息重构方法研究[D]. 哈尔滨：哈尔滨工业大学, 2022.

[5] MIAO Mou, ZHAO Xiaoqiang. Corrigendum to incipient fault detection and diagnosis of nonlinear industrial process with missing data[J]. Journal of the Taiwan Institute of Chemical Engineers, 2022, 134：104314.

[6] 姚邹静, 赵春晖, 李元龙, 等. 面向工业软测量应用的定制化生成对抗数据填补模型[J]. 控制与决策, 2021, 36(12)：2929-2936.

[7] LIU Junyan, KUMAR S, PALOMAR D P. Parameter estimation of heavy-tailed AR model with missing data via stochastic EM [J]. IEEE Transactions on Signal Processing, 2019, 67(8)：2159-2172.

[8] 张韧, 徐志升, 刘巍, 等. 基于信息扩散的海洋稀疏资料插值算法-概率模型[J]. 计算力学学报, 2013(4)：554-558.

[9] 孟杰, 李春林. 基于随机森林模型的分类数据缺失值插补[J]. 统计与信息论坛, 2014, 29(9)：86-90.

[10] 陈小波, 陈程, 陈蕾, 等. 基于改进低秩矩阵补全的交通量数据缺失值插补方法[J]. 交通运输工程学报, 2019, 19(5)：180-190.

[11] DABERDAKU S, TAVAZZI E, CAMILLO B D. A combined interpolation and weighted K-nearest neighbours approach for the imputation of longitudinal ICU laboratorydata [J]. Journal of Healthcare Informatics Research, 2020, 4(2)：174-188.

[12] 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.

[13] 张学文, 于兴伟, 侯鑫宇, 等. 基于时序向量相似性的空间目标群匹配技术研究[J]. 网络安全与数据治理, 2024, 43(2)：29-36.

[14] DENG Wei, GUO Yixiu, LIU Jie, et al. A missing power data filling method based on improved random forest algorithm [J]. Chinese Journal of Electrical Engineering, 2019, 5(4)：33-39.

[15] 陈娟, 王献雨, 罗玲玲, 等. 缺失值填补效果:机器学习与统计学习的比较[J]. 统计与决策, 2020, 36(17)：28-32.

[16] 张正智, 周赛, 王占伟, 等. 面向缺失数据基于 SVR-SVDD 的冷水机组故障检测方法[J]. 建筑科学, 2023, 39(2)：193-201.

[17] 黄灏尹. 缺失数据下基于神经网络的非参数回归及其应用[D]. 长春：吉林大学, 2022.

[18] 毋雪雁, 王水花, 张煜东. K 最近邻算法理论与应用综述[J]. 计算机工程与应用, 2017, 53(21)：1-7.

[19] ZHANG Shichao. Nearest neighbor selection for iteratively KNN imputation [J]. Journal of Systems and Software, 2012, 85(11)：2541-2552.

[20] 曹卫东, 倪建军, 姜博严. 支持大数据的参数自适应支持向量回归方法[J]. 计算机集成制造系统, 2023, 29(2)：511-521.

[21] 冯霞, 孟金双. 基于 KNN 和 SVR 的航班滑出时间预测[J]. 西南交通大学学报, 2017, 52(5)：1008-1014.

[22] 代超, 容芷君, 但斌斌, 等. 扇形段拉矫力数据突变特征提取方法[J]. 计算机集成制造系统, 2025, 31(1)：254-263.