

高校智能电表缺失数据修复方法^{*}

陈庆斌¹ 杨耿煌^{1,2} 耿丽清^{1,2} 苏娟³

(1.天津职业技术师范大学自动化与电气工程学院 天津 300222;2.天津职业技术师范大学天津市信息传感与智能控制重点实验室 天津 300222;3.中国农业大学信息与电气工程学院 北京 100083)

摘要:高校运行数据在采集、传输、存储过程中往往会产生数据缺失。对此,提出一种基于改进长短期记忆神经网络—链式方程多重插补法的缺失数据修复方法。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多个填补值,从而产生多个完整数据集,并进行分析优化得到一个最终的完整数据集。为提高缺失值修复精度,在长短期记忆神经网络的预测任务中,采用麻雀搜索算法进行超参数寻优,并结合均值匹配模型对缺失数据进行修复。使用北方某高校2019年数据进行验证,通过无自然缺失算例和自然缺失算例对提出方法进行评估,结果表明,在无自然缺失算例中,整体归因误差为0.106,较其他模型至少降低29.3%,验证了方法的有效性;对11.8%自然缺失率下的数据进行填补,经提出的方法填补之后的数据有效提高了高校后续运行数据的预测精度,间接验证了缺失数据填补的有效性。

关键词:高校运行数据;缺失数据填补;链式方程多重插补;长短期记忆神经网络

中图分类号: TN399 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Method for repairing missing data in college smart meters

Chen Qingbin¹ Yang Genghuang^{1,2} Geng Liqing^{1,2} Su Juan³

(1. School of Automation and Electrical Engineering, Tianjin University of Technology and Education, Tianjin 300222, China; 2. Tianjin Key Laboratory of Information Sensing and Intelligent Control, Tianjin University of Technology and Education, Tianjin 300222, China; 3. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract: Data loss often occurs in the process of collection, transmission and storage of university operation data. In this regard, this paper proposes a missing data repair method based on improved long short-term memory neural network-chain equation multiple interpolation method. The chain equation multiple interpolation method is used to generate multiple filling values for each missing attribute value through iteration, thereby generating multiple complete data sets, and analyzing and optimizing to obtain a final complete data set. In order to improve the accuracy of missing value repair, in the prediction task of long short-term memory neural network, the sparrow search algorithm is used to optimize the hyperparameters, and the mean matching model is used to repair the missing data. The data of a university in the north of China in 2019 are used for verification. The method proposed in this paper is evaluated by non-natural missing examples and natural missing examples. The results show that the overall attribution error of this method is 0.106 in non-natural missing examples, which is at least 29.3% lower than other models, which verifies the effectiveness of this method. The data under the natural missing rate of 11.8% is filled. The data filled by the method proposed in this paper effectively improves the prediction accuracy of the subsequent operation data of colleges and universities, and indirectly verifies the effectiveness of missing data filling.

Keywords: university operation data; missing data filling; multiple interpolation of chain equation; LSTM

收稿日期:2023-11-13

^{*} 基金项目:国家重点研发计划(2022YFB2403002)、天津市高等学校科技发展基金(2022ZD037)、天津市科技计划(23YDTPJC00320)项目资助

0 引言

随着高校内设施的不断完善,校园电力系统运行特性日趋复杂。在校园电力系统中,用电设备类型多、分布广、差异大,导致智能电表数据采集不确定性增加,数据随机缺失率升高,难以满足高校对电力数据进行分析决策的需求。因此,以采集的智能电表数据为研究对象,进行缺失数据修复方法研究,提升校园智能电表数据的完整性便成为了电力系统领域的一个重要研究方向^[1-2]。

机器学习在缺失数据修复应用广泛。文献[3]通过改进随机森林算法(random forest, RF)对缺失数据进行修复;文献[4]通过K-最近邻(K-nearest neighbor, KNN)对不完整信息进行修复;文献[5]通过建立支持向量回归(support vector regression, SVR)模型对缺失数据进行修复。上述研究采用方法均为浅层结构的机器学习算法,忽略了数据随时间变化的影响和数据前后的关联,不能有效处理时序数据缺失问题。近年来,深度学习在各个领域不断发展和应用,循环神经网络(recurrent neural network, RNN)在电力数据时序预测上表现优异,长短期记忆神经网络(long short-term memory, LSTM)解决了RNN随时间反向传播中权重消失的问题,能够自动提取时间维度上的周期特征和模式。文献[6]采用卷积神经网络对时序数据进行特征提取,结合LSTM对连续型缺失数据进行修复;文献[7]通过分析缺失数据前后完整序列,采用LSTM对缺失数据进行修复;文献[8]利用编码器优化聚类,将特征相似数据归为一类,再通过LSTM进行缺失数据修复。可见LSTM在缺失数据修复中的可行性,但LSTM中超参数的选择对其拟合能力具有关键作用,若将其直接应用于缺失数据修复,则难以充分挖掘数据间的关联性和时序特征,导致修复精度下降。同时,大多缺失数据修复方法均只产生单一填补数据集,无法充分考虑时序数据的波动性,进而影响缺失数据修复精度。

鉴于此,本文提出了一种基于改进LSTM网络的链式方程多重插补方法。该方法为每个缺失数据找到多个填补值,经过有限次迭代形成多个完整数据集,从多个数据集中选择一个最优的数据集用于后续工作。在预测部分,针对LSTM中超参数的选择问题,采用麻雀搜索算法(sparrow search algorithm, SSA)^[9-10]进行最佳超参数组合的寻优,得到超参数最佳的LSTM网络;在填补部分,不同数据集采用不同的填补方法;在数据集选择部分,采用赤池信息准则(akaike information criterion, AIC)^[11]和贝叶斯信息准则(Bayesian information criterion, BIC)^[12]进行最优数据集的选取。为验证填补方法的可行性,先对完整数据进行随机缺失处理,利用本文填补方法填补,再利用本文对自然缺失数据进行填补。经过对比实验和验证,证明了上述方法可行性,可显著提升填补的缺失数据准确率。

1 多重插补(multiple imputation, MI)法

MI法^[13-14]是一种用于处理缺失数据的统计方法,利用不完整数据集的现有数据对缺失值进行多次插补,生成多个完整数据集来估计缺失值。相比于单一填补,MI考虑缺失值不确定性对分析结果造成的影响,可获得更精确填补结果。

2 数据修复模型

2.1 改进LSTM网络

传统的LSTM网络通过人工调整参数,效率和准确性较低,SSA局部搜索能力强,收敛速度快,使用SSA对LSTM网络进行参数寻优,可自动选择最佳的超参数组合,从而提高模型性能,改进LSTM网络流程如图1所示。

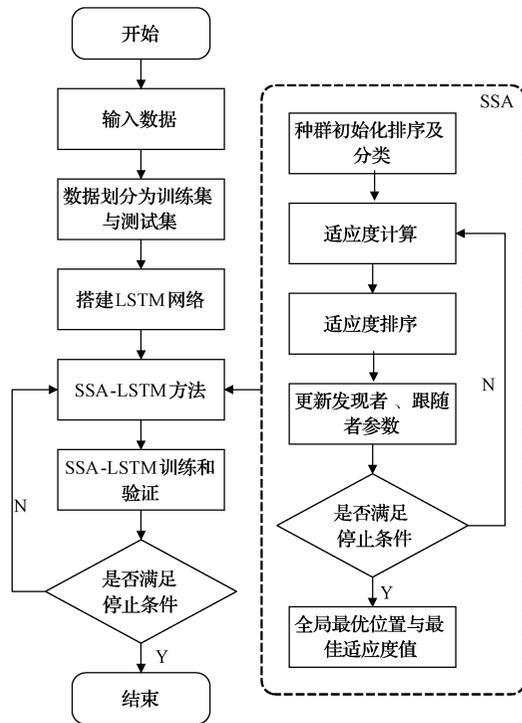


图1 改进LSTM网络流程

Fig. 1 Flow chart of improve LSTM network

2.2 基于改进LSTM-链式方程多重插补法的缺失数据修复方法

在MI的基础上,通过链式方程进行迭代优化,基于改进LSTM网络的链式方程多重插补方法共分为3个阶段,分别为使用改进LSTM网络进行预测,通过预测值中相近的值对应的原始数据集中的值进行数据填充和通过最优原则选取最佳数据集。缺失数据修复方法具体流程如图2所示。

基于改进LSTM网络-链式方程多重插补法的缺失数据修复方法步骤如下。

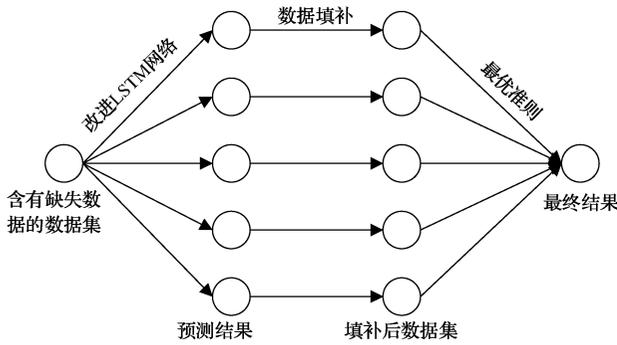


图2 缺失数据修复方法流程

Fig. 2 Flow chart of missing data repair method

步骤1)对含有缺失值的数据集,在当前属性中随机选择一些值对当前属性的缺失值进行填补,得到初始的随机填补的完整数据集。

步骤2)在初始的完整数据集中,对所有不完整属性进行缺失值的改进 LSTM 预测。例如,初始的完整数据集中 A、B、C 3 种属性,A 属性中存在缺失数据,则利用步骤1)填补后的 B、C 属性作为特征矩阵,A 属性作为标签列,采用改进 LSTM 网络对 A 属性进行预测,利用预测值中与缺失属性对应的预测值的值所对应的原始数据中的值进行填补,再将 A、C 属性作为特征矩阵,B 属性作为标签列,采用改进 LSTM 网络对 B 属性进行预测,依此类推。

步骤3)在步骤2)得到 A 属性的预测值后,不使用由改进 LSTM 产生的预测值,而是使用与改进 LSTM 预测值中相近值所对应的原始数据中的值进行填补,填补方法如式(1)所示。

$$y = \frac{1}{m+1} \sum_{i=1}^{m+1} y_i \quad (1)$$

式中: y 为缺失值; y_i 为第 i 个与预测值中相近值所对应的原始数据中的值; m 为当前已生成的数据集个数。

步骤4)重复执行步骤2)和3),直至填补完所有空缺值,得到1个完整数据集。

步骤5)重复执行上述步骤,得到多个完整数据集,对多个数据集采用 AIC 准则和 BIC 准则进行分析,得到最终的填补结果。

3 基于网络拓扑的能量守恒约束

校园内通过总表记录共用变压器的运行数据,食堂、教室、照明等由分表记录各自的运行数据,高校网络拓扑如图3所示。

图3中,总表记录的数据应等于各分表记录数据和其他损耗之和,符合能量守恒定律,如式(2)所示。

$$d_z = \sum_{i=1}^N d_i + \sum_{j=1}^{N-M} d_j + e_{\text{other}} \quad (2)$$

式中: d_z 为总表的数据; N 为分表的数量; M 为缺失该数据的分表数量; e_{other} 为线路损耗。

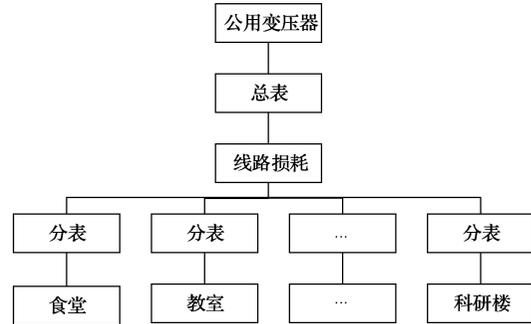


图3 高校网络拓扑图

Fig. 3 Network topology diagram of campuses

由式(2)可知,在同一时间和属性下,缺失值等于总表数据与其他完整分表数据和其他损耗的差值,如式(3)所示。

$$\sum_{i=1}^N d_i = d_z - \sum_{j=1}^{N-M} d_j - e_{\text{other}} \leq d_z - \sum_{j=1}^{N-M} d_j \quad (3)$$

因此,在同一时间和属性下,缺失值的上界为总表数据与其他完整分表数据的差值,当缺失数据的修复值超过上界时,则将修复值调整为总表数据与其他完整分表数据的差值。

4 算例分析

为验证模型的实际修复效果,本文将选取2个算例进行分析,其中算例1为高校运行数据中无自然缺失的数据,所有缺失数据均具有对应的真实值,可直接验证模型的有效性;算例2为高校运行数据中存在自然缺失的数据,通过能量守恒约束,并与历史数据中具有相同运行情况的无自然缺失的数据进行比对,可间接验证模型的有效性。

4.1 数据来源

本文以北方某高校2019年实际运行数据为例进行验证,共包含2个公用变电站,即总表2个,共包含宿舍、教学楼、食堂等38个设施,即分表38个,数据采样间隔为1h,其中,数据包含5种属性,分别是电压、电流、有功功率、无功功率和电度,其中部分数据示例如表1所示。

表1 部分高校运行数据示例

Table 1 Examples of operational data from some campuses

采样点	电压/V	电流/A	有功功率/kW	无功功率/kW	电度/(kW·h)
1	393.6	422.0	274.1	31.2	6 148 540.0
2	393.9	433.0	284.9	32.1	6 148 830.0
⋮	⋮	⋮	⋮	⋮	⋮
8 759	394.6	560.0	394.0	41.3	8 526 660.0
8 760	395.2	480.0	336.2	21.3	8 527 030.0

4.2 评价指标

为更好地评估模型的实际修复效果,本文选用平均绝对误差(mean absolute error,MAE)作为每个属性的修复误差,如式(4)所示。

$$e_{MAE} = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (4)$$

式中: n 为样本数量; x_i 为实际值; \hat{x}_i 为修复值。

为评价模型在多个属性上修复的整体效果,本文使用广义平均绝对偏差(generalized mean absolute deviation,GMAD)来衡量某个属性上的修复效果,如式(5)所示。

$$e_{GMAD} = \frac{1}{n} \sum_{i=1}^n \frac{|x_i - \hat{x}_i|}{x_{max} - x_{min}} \quad (5)$$

式中: x_{max} 和 x_{min} 分别为缺失值所在属性所有真实值的最大值和最小值。

随后,本文通过累加多个属性上的 GMAD 来表示整体归因误差 g_{sum} ,如式(6)所示。

$$g_{sum} = \sum_{i=1}^m e_{GMAD,i} \quad (6)$$

式中: m 为参与修复的属性个数; $e_{GMAD,i}$ 为第 i 个属性上的 GMAD 结果。

4.3 无自然缺失算例分析

算例采用数据来源整个月数据(12月),为验证模型对高校缺失数据填补的可行性和准确性,本文采用自定义函数对无自然缺失的完整数据集进行随机缺失数据处理,模拟自然缺失是为了对训练后模型的填补精度进行直接量化。通过改变自定义函数中不同参数的值来设置不同缺失率,分别为5%、10%、15%、20%和25%。

在不同缺失率下,采用本文提出模型进行缺失数据填补,在每个缺失率下分别产生5个不同的完整数据集,分别用①、②、③、④、⑤进行表示,数据集①的缺失数据填补方式为采用一个相近的值对应原始数据进行填补,数据集②的缺失数据填补方式为采用2个相近的值对应原始数据的均值进行填补,依此类推,如式(1)所示。

在不同缺失率下的不同数据集缺失数据填补完成后,在实际应用中只能选择填补效果最好的数据集进行应用。对此,本文采用AIC准则和BIC准则进行最优数据集的选取,AIC准则既考虑了模型的拟合情况,又考虑了模型复杂度的影响,AIC值越小,越接近真实情况;BIC准则加入了对样本数量的考虑,可有效避免模型复杂度过高的情况出现。不同缺失率下不同数据集的AIC、BIC值计算结果如表2所示,其中最小值加粗表示。

表2 不同数据集 AIC、BIC 计算结果

Table 2 Calculation results of AIC and BIC on different datasets

数据集	5%		10%		15%		20%		25%	
	AIC	BIC								
①	5 713.25	5 736.88	5 924.25	5 947.31	6 079.98	6 103.04	6 322.99	6 346.05	6 181.05	6 204.11
②	5 737.42	5 760.47	5 793.68	5 816.74	5 921.35	5 944.41	6 058.76	6 081.82	6 051.43	6 074.49
③	5 827.69	5 850.75	5 760.84	5 783.90	5 891.82	5 914.88	6 017.69	6 040.75	5 997.39	6 020.35
④	5 731.23	5 754.29	5 790.44	5 813.50	5 869.72	5 892.78	5 981.63	6 004.69	5 985.01	6 008.07
⑤	5 786.67	5 809.73	5 766.41	5 789.47	5 849.21	5 872.27	5 984.44	6 007.50	5 952.95	5 976.01

由表2可知,在5%缺失率下数据集①的修补效果最好,即采用一个相近的值对应原始数据进行填补,在10%缺失率下数据集③的修补效果最好,即采用3个相近的值对应原始数据进行填补,在15%和25%缺失率下数据集⑤的修补效果最好,即采用5个相近的值对应原始数据进

行填补,在20%缺失率下数据集④的修补效果最好,即采用4个相近的值对应原始数据进行填补。同时,本文对进行误差分析,分别计算各不同缺失率下各完整数据集的MAE和 g_{sum} ,具体计算结果如表3所示,其中最小误差加粗表示。

表3 不同数据集误差计算结果

Table 3 Error calculation results for different datasets

数据集	5%		10%		15%		20%		25%	
	e_{MAE}	g_{sum}	e_{MAE}	g_{sum}	e_{MAE}	g_{sum}	e_{MAE}	g_{sum}	e_{MAE}	g_{sum}
①	518.17	0.014	2 769.29	0.086	3 017.80	0.140	3 938.48	0.222	4 960.63	0.263
②	958.76	0.029	2 541.44	0.163	2 554.06	0.115	3 500.71	0.162	4 950.27	0.228
③	829.87	0.039	2 333.29	0.158	2 545.81	0.111	3 239.30	0.154	4 682.54	0.219
④	736.93	0.034	2 386.20	0.159	2 586.55	0.110	3 174.76	0.150	4 646.89	0.217
⑤	692.57	0.027	2 440.54	0.160	2 523.78	0.106	3 195.10	0.153	4 635.51	0.215

由表3可知,不同缺失率下的最优数据集误差最小,表明通过AIC准则和BIC准则判断最优数据集是合理的,这对存在自然缺失数据集插补后如何选择最优数据集提供了可靠的判据。

为验证本文提出方法的有效性,本文构建了多种模型

进行实验与结果的对比。本文考虑了一些应用广泛的机器学习模型,如RF和KNN等,也构建了具有强大学习能力的深度学习模型,如多层感知机(multi-layer perceptron,MLP)和LSTM等。本文提出方法与上述模型在15%缺失率数据集下的不同属性的实验结果如表4所示。

表4 不同模型的缺失填补效果比较

Table 4 Comparison of gap filling effects among different models

模型	电压/V		电流/A		有功功率/kW		无功功率/kW		电度/(kW·h)		g_{sum}
	e_{MAE}	e_{GMAD}									
KNN	1.147	0.070	41.34	0.060	26.05	0.054	8.924	0.082	71 061	0.218	0.485
SVR	1.394	0.086	24.34	0.035	25.58	0.053	9.045	0.083	84 300	0.259	0.517
RF	0.814	0.050	31.77	0.046	24.58	0.051	8.322	0.076	67 406	0.207	0.431
MLP	0.287	0.017	31.09	0.045	27.18	0.056	4.653	0.042	19 707	0.060	0.223
LSTM	0.301	0.018	23.03	0.033	9.618	0.020	3.210	0.029	15 993	0.049	0.150
本文方法	0.248	0.015	12.94	0.018	8.014	0.016	1.825	0.016	12 595	0.038	0.106

由表4可知,在15%模拟自然缺失率下,本文提出方法的填补结果更接近真实值,在电流、电压、有功功率、无功功率和电度5种属性上均优于其他模型,而且相较于排名第2的模型在结果上领先较多,其中,电度误差相较于其他属性大是因为电度本身是百万级数据,基数很大。参照MAE指标,在电流上领先0.17,在电压上领先0.43等。 g_{sum} 指标是将模型在所有属性上的GMAD进行求和,以表示模型在所有属性缺失值上的综合归因效果,本文提出方法取得了0.106的填充效果,排名第2的是LSTM,结果为0.150,其次是MLP,结果为0.223,可以看出,本文提出方法相较于其他模型更有优势,获得了更优的缺失数据填补效果。

4.4 自然缺失算例分析数据来源

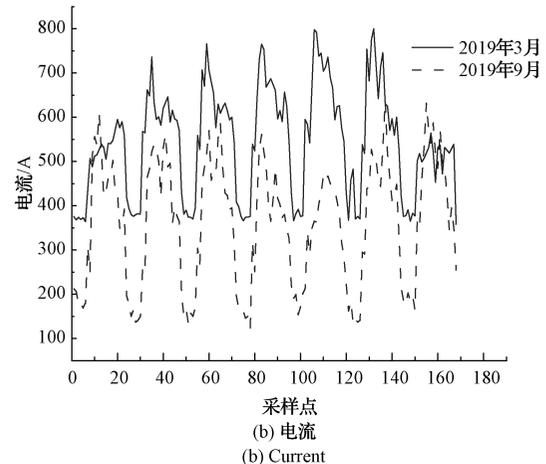
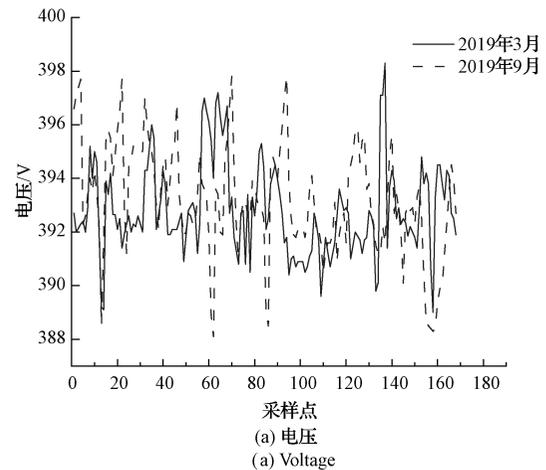
算例采用数据来源整个月数据(3月),该月公用变电站1共有数据3720个,自然缺失率为11.8%,由表2可知,采用3个相近的值对应原始数据进行填补所得完整数据集填补效果最优。

2019年3月处于春季且非疫情期间的正常运行状态,此时,校园处于刚开学状态,各项设施尚未到达满载状态,相同的,2019年9月处于秋季且非疫情期间的正常运行状态,从现实角度考虑,与3月的运行状态整体相同。在对2019年3月中缺失数据填补完成后,选取一周与2019年9月运行曲线进行对比,结果如图4所示。

由图4可知,本文提出方法对2019年3月中缺失数据填补完成后,整体运行趋势与2019年9月相似,其中电压整体处于380~400,属于正常运行波动;电流、有功功率、无功功率和电度波动情况均与2019年9月吻合,间接验证了本文提出方法在实际情况下也具有优良的性能。

由于算例中的自然缺失数据没有真值,无法直接量化不同模型填补的自然缺失数据精度,为对填补模型的误差进行间接量化,进而验证模型的填补效果,采用时间卷积

网络(temporal convolutional network,TCN)^[15]和极限梯度提升算法(extreme gradient boosting,XGBoost)^[16]对数据中的各种属性进行预测。模型的输入是含自然缺失高校运行数据在不同填补模型下填补得到的结果,包括5种属性7d的数据(168个时间采样点),输出是未来1d(24



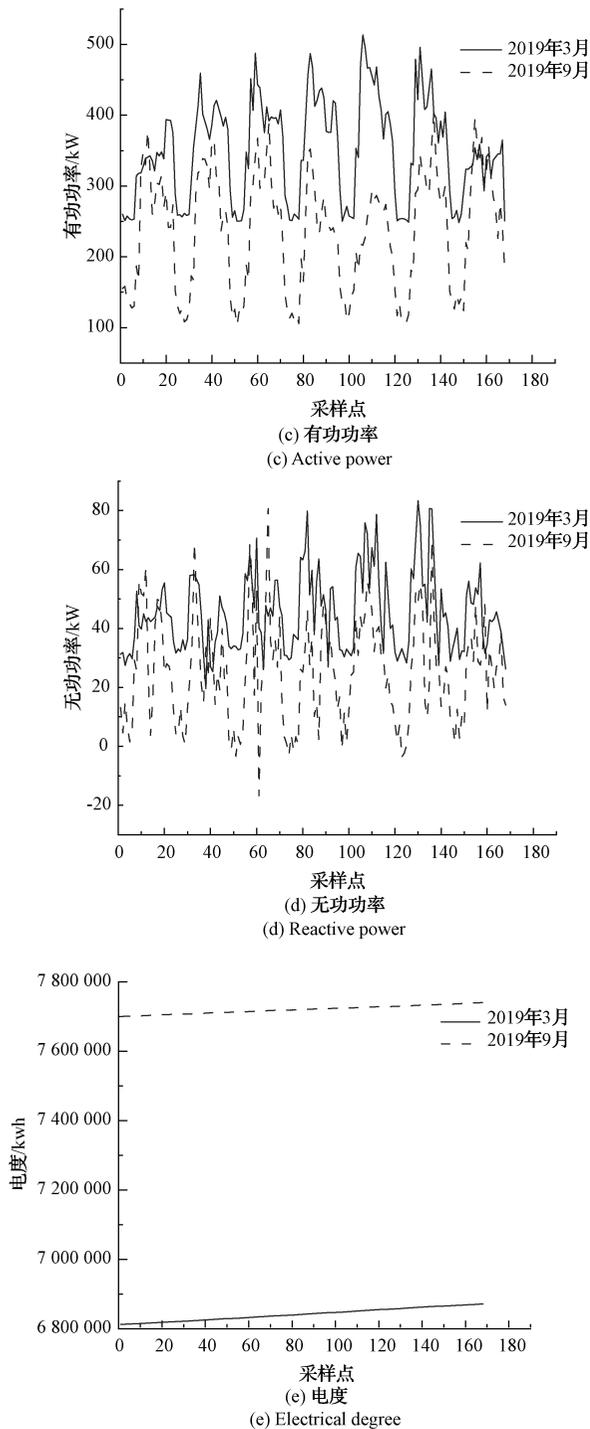


图4 运行曲线比较

Fig. 4 Comparison chart of operating curves

个时间采样点)的数据。其中,用于训练的168个时间采样点来自不同模型的填补数据,而24个预测点的真实数据来源于原始数据集。预测误差结果如图5所示。

由图5可知,本文提出方法的平均MAE为818.49,较其他几种填补模型的误差最低,较次优的LSTM模型降低7.6%。由此可知,本文提出方法通过多次填补得到

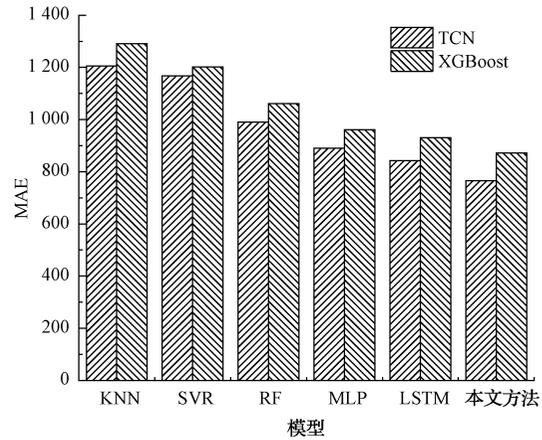


图5 不同方法下不同模型填补数据预测误差结果

Fig. 5 Results of filling data prediction errors with different models under different methods

多个完整数据集,在不同缺失率下选择最优填补方式,可有效对缺失数据进行修复,进而提升预测模型的泛化能力,使得预测模型可以获取更精确的信息,从而提高了预测结果精度,因此,使用本文提出方法填补后的数据进行时序预测的效果优于其他填补模型,间接验证了本文提出方法对缺失数据填补效果最优。

5 结论

本文针对高校运行数据存在缺失的问题,提出了一种基于改进LSTM网络的链式方程多重插补法,并利用北方某高校2019年实际运行数据进行验证,得到如下结论。

1) 本文提出方法可以对高校运行过程中的缺失数据进行填补,验证了模型的有效性。同时,本文提出方法可以处理具有多种属性的数据,自主学习参数之间的相关性,使用改进LSTM网络的方法进行预测,不需要太多超参数调整,降低了填补复杂性,提高了填补精度。

2) 本文通过人为构造不同缺失率下的数据集,采用本文提出方法进行填补,在填补过程中,针对不同的数据集采用不同的填充方式,并通过AIC准则和BIC准则选取最优数据集,通过验证,发现了在高校运行数据中不同缺失率下数据的最佳填充方式,为自然缺失数据中缺失值的填充提供了依据。

3) 基于不同模型填补的数据对高校运行数据进行时序预测,表明本文提出方法可有效对缺失数据进行修复,进而提升预测模型的泛化能力,使得预测模型可以获取更精确的信息,从而提高了预测结果精度,因此,使用本文提出方法填补后的数据进行时序预测的效果优于其他填补模型,间接验证了本文提出方法对缺失数据填补效果最优。

对缺失数据进行填补后,后续将构建校园用电行为特征,建立校园综合能源系统,进而节能减排。

参考文献

- [1] 王永超, 唐求, 马俊, 等. 基于复合核 SVM 的智能电表基本误差预测方法[J]. 电子测量与仪器学报, 2021, 35(10):209-216.
WANG Y CH, TANG Q, MA J, et al. Prediction method of basic error of smart meter based on composite core SVM [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(10): 209-216.
- [2] 徐健, 刘长良, 王梓齐, 等. 基于缺失数据填补的风电齿轮箱状态监测研究[J]. 仪器仪表学报, 2022, 43(9):88-97.
XU J, LIU CH L, WANG Z Q, et al. Research on condition monitoring of wind turbine gearbox based on missing data imputation[J]. Chinese Journal of Scientific Instrument, 2022, 43(9):88-97.
- [3] 吕勤学, 郭杜杜, 李心, 等. 基于优化随机森林算法的浮动车 GPS 数据插补模型[J]. 科学技术与工程, 2022, 22(4):1656-1661.
LYU Q X, GUO D D, LI X, et al. GPS data interpolation model of floating car based on optimized random forest algorithm[J]. Science Technology and Engineering, 2022, 22(4):1656-1661.
- [4] 赵春霞, 赵营颖. 基于多元回归 KNN 的网络数据库不完整信息填充[J]. 计算机仿真, 2021, 38(8):339-343.
ZHAO CH X, ZHAO Y Y. Incomplete information filling of network database based on multiple regression KNN [J]. Computer Simulation, 2021, 38(8):339-343.
- [5] 遇茜, 钱政, 聂志鹏. 风机异常及缺失数据的填补方法研究[J]. 电测与仪表, 2020, 57(23):1-8.
YU Q, QIAN ZH, NIE ZH P. Research on filling method of abnormal and missing data of wind turbines [J]. Electrical Measurement & Instrumentation, 2020, 57(23): 1-8.
- [6] 杨旭, 崔瑞飞, 田超, 等. 基于线性样条和 CNN-LSTM 的北斗卫星缺失数据处理方法[J]. 空间科学学报, 2022, 42(1):163-169.
YANG X, CUI R F, TIAN CH, et al. Linear spline and CNN-LSTM for missing values imputation of beidou satellite radiation dose data [J]. Chinese Journal of Space Science, 2022, 42(1):163-169.
- [7] 宋维, 高超, 赵玥, 等. 基于 LSTM 的活立木茎干水分缺失数据填补方法[J]. 林业科学, 2020, 56(2): 134-141.
SONG W, GAO CH, ZHAO Y, et al. Method of filling the missing water loss data of living plant stem by sequence based on LSTM [J]. Scientia Silvae Sinicae, 2020, 56(2):134-141.
- [8] 卢继哲, 刘宣, 唐悦, 等. 基于聚类和 LSTM 的电力分钟冻结数据缺失值填充方法[J]. 控制工程, 2022, 29(4):611-616.
LU J ZH, LIU X, TANG Y, et al. Missing value treatment for minute freezing data of electricity based on clustering and LSTM[J]. Control Engineering of China, 2022, 29(4):611-616.
- [9] 陈庆斌, 杨耿煌, 耿丽清, 等. 基于改进麻雀搜索算法的配送中心选址研究[J]. 天津职业技术师范大学学报, 2023, 33(2): 14-19.
CHEN Q B, YANG G H, GENG L Q, et al. Study on distribution center location based on improved sparrow search algorithm [J]. Journal of Tianjin University of Technology and Education, 2023, 33(2): 14-19.
- [10] 侯利民, 韦卓庆, 曹生辉, 等. 麻雀搜索算法的 PMSM 固定时间无模型滑模控制[J]. 电子测量与仪器学报, 2023, 37(12):156-165.
HOU L M, WEI ZH Q, CAO SH H, et al. Fixed time model-free sliding mode control of PMSM based on sparrow search algorithm[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(12):156-165.
- [11] CLAUDIA W, LISA M, SUMEYYE U, et al. Prevalence and determinants of overweight and obesity among preschool-aged children from migrant and socioeconomically disadvantaged contexts in Montreal, Canada[J]. Preventive Medicine Reports, 2023, 36:102397.
- [12] DANIEL P, ARKAITZ I, ANA L, et al. Switching to bictegravir/emtricitabine/tenofovir alafenamide (BIC/FTC/TAF) plus darunavir/cobicistat in heavily antiretroviral-experienced, virologically suppressed HIV-infected adults receiving complex regimens[J]. The Journal of Antimicrobial Chemotherapy, 2023, 78(11): 2696-2701.
- [13] ZHAO F, LU Y, LI X N, et al. Multiple imputation method of missing credit risk assessment data based on generative adversarial networks[J]. Applied Soft Computing Journal, 2022, 126:109273.
- [14] ALERYANI, WANG W, IGLESIA L D B. Multiple

- imputation ensembles (MIE) for dealing with missing data[J]. *Computer Science*, 2020, 1(2):273-291.
- [15] 常乐, 汪庆年. 基于优化聚类分解与 XGBOOST 的超短期电力负荷预测[J]. *国外电子测量技术*, 2022, 41(5):46-51.
- CHANG L, WANG Q N. Ultra-short-term power load forecasting based on optimal cluster decomposition and XGBOOST[J]. *Foreign Electronic Measurement Technology*, 2022, 41(5):46-51.
- [16] 严帅, 熊新. 基于 KPCA 和 TCN-Attention 的滚动轴承退化趋势预测[J]. *电子测量技术*, 2022, 45(15):28-34.
- YAN SH, XIONG X. Prediction of rolling bearing degradation trend based on KPCA and TCN-Attention[J]. *Electronic Measurement Technology*, 2022, 45(15):

28-34.

作者简介

陈庆斌, 硕士研究生, 主要研究方向为负荷预测、电力系统。

E-mail: c1185955913@163.com

杨耿煌(通信作者), 教授, 主要研究方向为智能信息处理与控制系统。

E-mail: ygh@tute.edu.cn

耿丽清, 副教授, 主要研究方向为信号检测与处理、智能控制。

E-mail: gengliq2004@126.com

苏娟, 副教授, 主要研究方向为电力系统。

E-mail: sujuan@cau.edu.cn