自动化学报
ACTA AUTOMATICA SINICA

面向全量测点耦合结构分析与估计的工业过程监测方法

赵健程 赵春晖

An Industrial Process Monitoring Method Based on Total Measurement Point Coupling Structure Analysis and Estimation

ZHAO Jian-Cheng, ZHAO Chun-Hui

在线阅读 View online: https://doi.org/10.16383/j.aas.c220090

您可能感兴趣的其他文章

复杂工业过程质量相关的故障检测与诊断技术综述

Review of Quality-related Fault Detection and Diagnosis Techniques for Complex Industrial Processes

自动化学报. 2017, 43(3): 349-365 https://doi.org/10.16383/j.aas.2017.c160427

基于局部 整体相关特征的多单元化工过程分层监测

Hierarchical Monitoring for Multi-unit Chemical Processes Based on Local-global Correlation Features

自动化学报. 2020, 46(9): 1770-1782 https://doi.org/10.16383/j.aas.c190671

基于动态特性描述的变量加权型分散式故障检测方法

Dynamic Feature Characterization Based Variable-weighted Decentralized Method for Fault Detection

自动化学报. 2021, 47(9): 2205-2213 https://doi.org/10.16383/j.aas.c180276

基于稀疏残差距离的多工况过程故障检测方法研究

Fault Detection of Multi-mode Processes Employing Sparse Residual Distance

自动化学报. 2019, 45(3): 617-625 https://doi.org/10.16383/j.aas.c170389

一种基于数据可靠性和区间证据推理的故障检测方法

A Fault Detection Method Based on Data Reliability and Interval Evidence Reasoning

自动化学报. 2020, 46(12): 2628-2637 https://doi.org/10.16383/j.aas.c180518

基于近邻主元得分差分的故障检测策略

Fault Detection Strategy Based on Principal Component Score Difference of Nearest Neighbors

自动化学报. 2020, 46(10): 2229-2238 https://doi.org/10.16383/j.aas.c180163

面向全量测点耦合结构分析与估计的工业过程监测方法

赵健程 [1] 赵春晖 [1]

摘 要 实际工业场景中, 需要在生产过程中收集大量测点的数据, 从而掌握生产过程运行状态. 传统的过程监测方法通常仅评估运行状态整体的异常与否, 或对运行状态进行分级评估, 这种方式并不会直接定位故障部位, 不利于故障的高效检修. 为此, 提出一种基于全量测点估计的监测模型, 根据全量测点估计值与实际值的偏差定义监测指标, 从而实现全量测点的分别精准监测. 为克服原有的基于工况估计的监测方法监测不全面且对测点间耦合关系建模不充分的问题, 提出多核图卷积网络 (Multi-kernel graph convolutional network, MKGCN), 通过将全量传感器测点视为一张全量测点图, 显式地对测点间耦合关系进行建模, 从而实现全量传感器测点的同步工况估计. 此外, 面向在线监测场景, 设计基于特征逼近的自迭代方法, 从而克服在异常情况下由于测点间强耦合导致的部分测点估计值异常的问题. 所提出的方法在电厂百万千瓦超超临界机组中引风机的实际数据上进行验证, 结果显示, 与其他典型方法相比, 所提出的监测方法能够更精准地检测出发生故障的测点.

关键词 自迭代特征替换, 多核图卷积网络, 全量测点估计, 故障检测

引用格式 赵健程, 赵春晖. 面向全量测点耦合结构分析与估计的工业过程监测方法. 自动化学报, 2024, **50**(8): 1517−1538

DOI 10.16383/j.aas.c220090

An Industrial Process Monitoring Method Based on Total Measurement Point Coupling Structure Analysis and Estimation

ZHAO Jian-Cheng[1] ZHAO Chun-Hui[1]

Abstract In the actual industrial scenario, it is necessary to collect a large number of data from measuring points in the production process, so as to master the operational state of the production process. Traditional process monitoring methods usually only evaluate whether the overall operation state is abnormal or not, or carry out hierarchical evaluation of the state. These methods do not directly locate the fault location, which is not conducive to the efficient maintenance of the fault. Therefore, in this paper, a monitoring model based on total measurement point estimation is proposed, and the monitoring indicators are defined according to the deviation between the estimated value and the actual value of total measurement points, so as to realize the separate and accurate monitoring of total measurement points. In order to overcome the problems of incomplete monitoring and insufficient modeling of coupling relationship between measuring points in the original monitoring method based on condition estimation, a multi-kernel graph convolution network (MKGCN) is proposed. By treating the measuring points as a graph of the total measurement points, the coupling relationship between measuring points is explicitly modeled, thus realizing the synchronous estimation of total measuring points. In addition, for the on-line monitoring scenario, a self-iteration method based on feature approximation is designed to overcome the issue of abnormal estimation of some measurement points due to the strong coupling between measurement points under abnormal system state. The method proposed in this paper is verified on the actual data of induced draft fan in 1 000 MW ultra-supercritical thermal power unit of power plant. The results show that the monitoring method proposed in this paper can detect the fault measuring points more accurately than other typical methods.

Key words Self-iterative feature replacement, multi-kernel graph convolutional network (MKGCN), total measurement point estimation, fault detection

Citation Zhao Jian-Cheng, Zhao Chun-Hui. An industrial process monitoring method based on total measurement point coupling structure analysis and estimation. *Acta Automatica Sinica*, 2024, **50**(8): 1517−1538

收稿日期 2022-02-07 录用日期 2022-09-06

Manuscript received February 7, 2022; accepted September 6, 2022

国家自然科学基金杰出青年基金 (62125306), 国家自然科学基金重点项目 (62133003) 资助

Supported by National Natural Science Foundation of China for Distinguished Young Scholars (62125306) and National Natural Science Foundation of China (62133003)

本文责任编委 杨浩

Recommended by Associate Editor YANG Hao

1. 浙江大学控制科学与工程学院 杭州 310027

1. College of Control Science and Engineering, Zhejiang University, Hangzhou 310027

工业过程运行状态监测对于提高生产效率、保证生产安全具有重要意义. 复杂工业环境下运行工况的识别与复杂工业系统智能建模也是工业人工智能的重要研究方向之一[1]. 由于工业过程的连续不间断运行, 导致故障在不同层级间传播和演化, 使得工业过程的监测和诊断成为一个复杂的问题[2]. 为了监控工业生产设备的工作状态, 通常需要对大量测点进行监测, 从而保证工业生产过程的经济、高效、安全. 同一生产过程中, 众多测点之间存在机理上的相关性, 且测点的值和测点间的关系反映了设备性能和健康状况.

由于工业生产流程的复杂性, 基于第一性原理对工业生产过程建立准确的机理模型比较困难, 因此数据驱动的方法近年来获得了更多的关注[3−4]. 从历史的过程数据中提取信息进行建模, 从而进行运行工况监测的数据驱动方法已成为运行工况研究的热点领域之一[5]. 具体来说, 基于统计的过程监测方法通常通过建立监测模型, 定义并计算监测统计量, 根据统计量是否超过控制限判断过程的运行状态, 如基于主成分分析 (Principal component analysis, PCA)[6−7]、协整分析 (Cointegration analysis, CA)[8]、慢特征分析 (Slow feature analysis, SFA)[9−10]、独立成分分析 (Independent component analysis, ICA) 的方法[11] 等. 神经网络[12−14]、贝叶斯网络[15] 等方法也广泛应用于工业生产过程监测、异常检测场景. 此外, 自编码器 (Autoencoder, AE) 结构作为一种有效的异常检测方法, 在图像异常检测[16]、时间序列异常检测[17−19] 等多领域异常检测方面取得了广泛的应用. 然而, 一般的监测方法通常仅判断过程整体的运行状态故障与否, 或对运行状态健康程度进行分级评估. 这种方式并不能直接定位具体故障的故障部位或传感器测点, 需要进一步使用故障隔离、故障根因追溯[20−21] 等方法定位具体发生故障的部位, 明确报警的具体测点, 从而便于一线人员进一步检修. 此外, 对于一线操作人员, 复杂的监测指标是难以理解的. 相对而言, 直接给出测点的估计值, 更有利于一线操作人员理解并进行故障定位.

大量过程数据、工艺数据为判断工业过程运行状态、明确故障原因提供了丰富的数据支持[22]. 本文中描述的测点估计, 定义为以各个测点传感器当前及历史测量得到的数据作为输入, 重构出当前时刻在正常运行状态下各个测点的估计值. 在对测点进行准确估计的基础上, 通过测点估计值与实测值间出现的显著偏差, 为一线人员故障检修提供直接的参考. 对于工业过程中的测点估计任务, 一种经典有效的方法是多变量状态估计方法 (Multivariate state estimation technique, MEST), 在多个领域得到广泛应用[23−27]. MEST 方法使用系统正常运行条件下收集到的足够多的历史数据构建过程记忆矩阵, 使用当前实际测量得到的数据与过程记忆矩阵中的各个历史状态计算相似度, 并利用相似度对历史状态进行加权求和, 从而得到当前系统状态的估计值. 然而, 这种方法一方面对过程记忆矩阵的要求较高, 需要其包含该系统全部正常运行范围和系统所有的特殊状态, 没有对系统内变量间潜在的耦合关系进行有效的建模与学习, 泛化能力较差; 另一方面缺少一种可靠的纠错机制, 系统中的某些变量出现异常可能导致当前状态与过程记忆矩阵中某些本不该相似的状态之间相似性提高, 使估计值出现异常, 从而无法准确判断当前状态的异常情况.

另一种可行的测点估计方法是软测量方法. 对于某一个具体的测点而言, 利用其他测点的信息对其进行估计, 这与一般的软测量任务的范式是一致的. 随着大量工业传感器和工业物联网技术的使用, 大量的过程数据被累积, 为数据驱动的软测量方法提供了支持. He 等[4] 结合偏最小二乘算法 (Partial least squares, PLS) 和极限学习机 (Extreme learning machine, ELM), 对精对苯二甲酸工艺过程的关键变量进行预测. Fan 等[28] 结合多层感知器 (Multilayer perceptron, MLP) 和双 LASSO (Least absolute shrinkage and selection operator) 算法, 对原油蒸馏过程中的重要产物 (煤油) 进行预测. Zhang 等[29] 基于进化极限学习机估计真实丙烯聚合过程变量的熔融指数. 深度学习方法在工业过程软测量领域也得到了有效应用. Ke 等[30] 开发了一种基于长短期记忆网络 (Long short term memory, LSTM) 的软测量模型, 针对硫磺回收装置脱硫系统中 SO_2 和 H_2S 的含量进行预测, 展现了 LSTM 对工业过程动态性进行建模的能力. Yuan 等[31] 设计了一种动态卷积神经网络, 用于在软测量模型中学习分层局部非线性动态特征. Zhu 等[32] 使用卷积神经网络 (Convolutional neural networks, CNN), 应用移动窗口法, 在软测量模型中捕获主导变量和辅助变量间时序相关性. 常树超等[33] 提出一种应用于工业过程软测量时空协同的图卷积网络 (Graph convolutional network, GCN), 使用 LSTM 捕获时序特征, 并使用图神经网络显式地建模各个测点间的相关性. 一般的软测量方法通常仅建立少数、部分测点作为输出的模型. 然而, 只对少数测点进行监测不能够全面充分地反映过程的运行状态, 且忽略了部分指标之间的相关性. 一种可能的策略是轮换建立以当前时刻部分指标为输出, 其余全部指标为输入的模型. 然而, 这种方法每次仅以单个指标

为目标, 对于以多测点估计为目标的任务而言, 没有更深入地建模和挖掘测点之间的耦合关系. 此外, 这种方法对大量测点的信息进行重复的特征提取操作, 且需建立多个模型, 费时费力.

随着图神经网络的蓬勃发展, 一些基于图的方法, 例如图卷积网络[34], 在多种类型的任务中取得良好的效果. 基于图卷积方法, 能够同时更新图中全部节点的特征. 若将众多传感器测点视为一张过程全量测点图, 就能够使用单个图卷积模型实现所有测点特征的同步更新. 基于图的方法通常需要确定图的结构, 即邻接矩阵. 然而, 由于工业生产过程的复杂性, 工业测点之间的关系是部分甚至全部未知的. 也就是说, 基于测点建立的生产过程全量测点图, 其图结构是未知的. 一些确定邻接矩阵的方法已经被提出, 例如基于稀疏编码的方法[33]、基于相关性度量的方法[35], 这两种方式都不是端到端的. Wu 等[36] 在多变量时间序列预测任务中提出一种用于获得图邻接矩阵的图学习层, 基于学习到的节点特征和稀疏性约束能够端到端地学习邻接矩阵. 然而, 上述方法均仅确定单一的邻接矩阵, 对一个内部机理复杂的工业过程而言, 单个邻接矩阵未必能够充分描述众多测点间的耦合关系.

当工业生产过程处于正常状态时, 一个能够充分建模测点之间相关关系的估计模型无疑是好的. 然而, 当某些测点测量的物理量因为某种原因出现异常变化时, 可能会导致模型对于正常的测点的估计同时出现异常, 即产生误报警; 或导致模型对于部分异常的测点的估计出现异常, 致使无法辨识出部分异常的测点, 即产生漏报警, 这也是基于软测量或估计的方法进行异常监测所面对的主要难点.

本文所述的全量测点, 定义为工业过程中某一对象可获得并用于建模的全部测点. 本文将图的邻接矩阵视为一种图的结构特征估计算子, 称其为邻接核 (Adjacency kernel), 从而构建了多核图卷积网络 (Multi-kernel graph convolutional network, MKGCN). 并通过多组可学习的邻接核得到全量测点图不同通道的特征, 从而面向工业场景下图结构未知的任务实现了更充分、更鲁棒的建模. 在提出多核图卷积网络的基础上, 本文在模型训练过程中提出特征逼近约束, 并引入自迭代方法, 通过估计特征逼近和自迭代替换异常特征, 逐步消除异常工况下异常测点对模型估计其他测点造成的干扰, 从而在异常情况下也能实现对全测点的良好估计. 在得到全量测点的估计值后, 通过测点估计值和实际值的偏差情况判断测点在当前时刻是否出现异常. 在训练阶段, 基于模型估计值和实际值的偏差, 确定每个测点的控制限; 在应用阶段, 根据控制限和偏差情况确定测点是否出现异常, 实现全量测点的监测.

本文的主要贡献点如下:

1) 提出一种基于测点间耦合结构分析与测点误差估计的全量测点同步监测理念. 通过将众多测点视为一张工业生产全量测点图的节点, 提出一种多核图卷积网络, 从而对测点间的耦合结构进行显式的建模与充分的学习. 通过全量测点特征的同步更新, 基于单模型实现了全部测点的同步估计与监测.

2) 通过在多核图卷积网络的基础上引入基于特征逼近的自迭代方法, 构建一种可靠的纠错机制. 应用训练好的基于多核图卷积网络的工况估计模型, 在自迭代过程中逐步消除异常特征的衍生影响, 缓解了系统工作异常的情况下异常测点对其他测点的估计造成的干扰, 在多个测点同时产生异常时仍能保持全量测点估计值的准确性.

1　相关方法

在本节中, 为便于更好地理解本文, 对涉及到的相关方法进行介绍.

1.1　图卷积网络

近年来, Kipf 等[34] 提出的图卷积网络在多种面向图结构数据的任务上得到广泛的应用. 图卷积网络能够有效地抽取图结构数据的特征. 定义图 $G = (V, E)$, 其中, V 是图中节点的集合, E 是图中边的集合. 定义图的邻接矩阵为 A, 是一个 $n \times n$ 的矩阵, n 为图中节点个数. 如果 $e_{ij} \in E$, 则 $a_{ij} \in A = 1$, 否则 $a_{ij} \in A = 0$. 对于加权邻接矩阵, e_{ij} 为 $[0, 1]$ 之间的值. 定义矩阵 $\tilde{A}$ 和度矩阵 $\tilde{D}$ 如下:

$$\tilde{A} = A + I_n \tag{1}$$

$$\tilde{D}_{ii} = \sum_{j=1}^{n} \tilde{a}_{ij} \tag{2}$$

其中, $\tilde{a}_{ij} \in \tilde{A}$, I_n 为大小为 $n \times n$ 的单位矩阵. 定义 $H^{(l)}$ 为第 l 层的隐藏特征, 则图卷积的定义为

$$H^{(l+1)} = \xi\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{3}$$

其中, 当 $l = 0$ 时, $H^{(0)}$ 为初始的特征; $W^{(l)}$ 为可训练的参数矩阵, 用于对特征进行线性变换; ξ 为 Tanh 激活函数. $\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ 项实现邻接矩阵的标准化. $H^{(l)}$ 为更新前图的节点的特征矩阵, 大小为 $n \times m$, m 为输入时每个节点的特征维度. $H^{(l+1)}$ 为更新后图的节点的特征矩阵.

1.2 LSTM 层与全连接层

LSTM 首先由 Hochreiter 等[37] 和 Gers 等[38] 提出, 是循环神经网络的一种变体, 用于解决序列长期依赖问题. LSTM 的结构如图 1 所示. 每个 LSTM 单元根据当前时刻的输入 $\boldsymbol{x}_t$ 和前一时刻的细胞状态 $\boldsymbol{c}_{t-1}$、隐藏状态 $\boldsymbol{h}_{t-1}$, 计算当前时刻的隐藏状态 $\boldsymbol{h}_t$、细胞状态 $\boldsymbol{c}_t$. 与简单循环神经网络相比, LSTM 更有效地学习数据中存在的长期依赖关系, 在许多有挑战性的序列处理任务中获得最先进的表现[39−40].

LSTM 的计算式为

$$\boldsymbol{i}_t = \sigma\left(W_{ii}\boldsymbol{x}_t + \boldsymbol{b}_{ii} + W_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_{hi}\right) \tag{4}$$

$$\boldsymbol{f}_t = \sigma\left(W_{if}\boldsymbol{x}_t + \boldsymbol{b}_{if} + W_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_{hf}\right) \tag{5}$$

$$\boldsymbol{o}_t = \sigma\left(W_{io}\boldsymbol{x}_t + \boldsymbol{b}_{io} + W_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{b}_{ho}\right) \tag{6}$$

$$\boldsymbol{g}_t = \xi\left(W_{ig}\boldsymbol{x}_t + \boldsymbol{b}_{ig} + W_{hg}\boldsymbol{h}_{t-1} + \boldsymbol{b}_{hg}\right) \tag{7}$$

$$\boldsymbol{c}_t = \boldsymbol{f}_t * \boldsymbol{c}_{t-1} + \boldsymbol{i}_t * \boldsymbol{g}_t \tag{8}$$

$$\boldsymbol{h}_t = \boldsymbol{o}_t * \xi\left(\boldsymbol{c}_t\right) \tag{9}$$

$$\boldsymbol{y}_t = \boldsymbol{h}_t \tag{10}$$

其中, σ 为 Sigmoid 激活函数, $*$ 表示 Hadamard 乘积 $\boldsymbol{b}_{ii}$, $\boldsymbol{b}_{hi}$, $\boldsymbol{b}_{if}$, $\boldsymbol{b}_{hf}$, $\boldsymbol{b}_{io}$, $\boldsymbol{b}_{ho}$, $\boldsymbol{b}_{ig}$, $\boldsymbol{b}_{hg}$, W_{ii}, W_{hi}, W_{if}, W_{hf}, W_{io}, W_{ho}, W_{ig}, W_{hg} 为可训练参数. LSTM 用于提取单个指标历史数据中包含的时序特征. 在本文中, 使用了双向的长短期记忆网络 (Bi-directional LSTM, BiLSTM)[41], 如式 (11) 所示, 其中, *concat* 表示拼接操作, $\overrightarrow{\boldsymbol{h}}_t$ 为输入 X 时序正向输入 LSTM 层得到的特征, $\overleftarrow{\boldsymbol{h}}_t$ 为输入 X 时序反向输入 LSTM 层得到的特征, $\overleftrightarrow{\boldsymbol{h}}_t$ 为 BiLSTM的输出, 即将输入 X 分别时序正向和时序反向输入 BiLSTM 层, 再将输出特征进行拼接.

$$\overleftrightarrow{\boldsymbol{h}}_t = concat(\overrightarrow{\boldsymbol{h}}_t, \overleftarrow{\boldsymbol{h}}_t) \tag{11}$$

双向的长短期记忆网络是一种常用的 LSTM 的变体, 通过正序和逆序两个方向提取数据中的时序特征, 从而更充分地对序列进行特征提取, 对于多种类型的序列数据都有很好的效果.

此外, 全连接层 (Fully connected layer, FC) 定义为

$$F^{(i)} = \xi\left(F^{(i-1)}W^{(i)} + \boldsymbol{b}^{(i)}\right) \tag{12}$$

其中, $F^{(i-1)}$ 是前一层网络的输出, $W^{(i)}$ 和 $\boldsymbol{b}^{(i)}$ 是可训练参数.

2 面向全量测点耦合结构分析的估计模型

本节介绍面向全量测点估计的多核图卷积模型的整体结构、多核图卷积层的实现、模型训练过程与特征逼近约束.

2.1 估计模型整体结构

在本文提出的全量测点同步估计模型中, 每个测点的估计值并非由其自身当前及历史的测量值而来, 而是利用其他测点当前及历史的测量值进行估计, 这与一般的软测量任务的范式是一致的. 在工业生产过程中, 某些测点因测量、传输、处理等环节发生短时故障的情况时有发生, 因此使用基于软测量的范式对模型的实际落地应用是十分有益的, 可以避免因为测点自身测量值缺失导致无法产生估计值. 由于传感器测点均部署在同一生产过程中, 彼此存在机理上的联系, 故而存在一定程度的相关性, 因此这种估计方式是合理的.

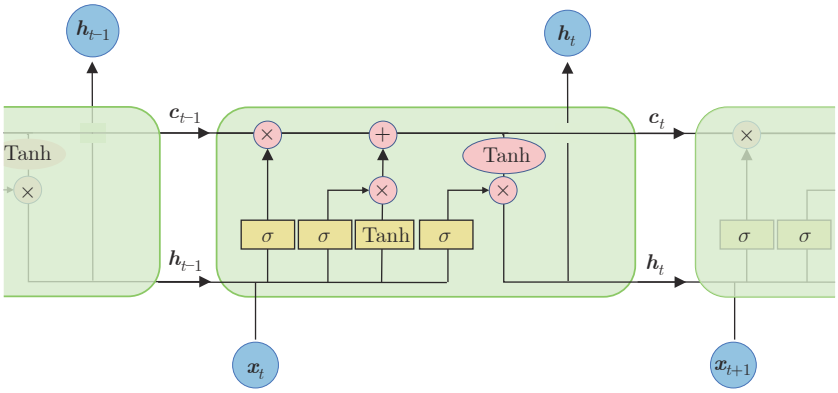


图 1 LSTM 内部结构

Fig. 1 Internal structure of LSTM

面向全量测点估计的多核图卷积模型结构如图 2 所示. 基于 MKGCN 层的估计模型结构, 由 4 个模块构成, 包括多通道特征提取模块、基于 MKGCN 层的特征聚合模块、估计模块以及特征逼近模块. 在每个时刻, 模型的输入 X 为实际部署的传感器提供的各个测点的数据. X 的大小为$[n, len]$, 其中 n 为测点数, 将所有测点视为一张工况全量测点图中的节点, n 即为图的节点数; len 为时间窗口长度, 即 X 包含长度为 len 的 t, $t-1$, $t-2$, $\cdots$, $t-len+1$ 时刻所采集到的全量测点的测量数据. 模型输出假定系统在正常情况下, 全量 n 个测点在当前 t 时刻的估计值. 对于每一个具体的测点, 模型使用其他测点的特征, 估计其 t 时刻的值.

在多输入通道特征提取模块, 对于输入 X, 使用多通道特征提取模块提取多通道输入的特征. 本文使用两个输入通道, 即 $c_{\text{in}}=2$, 分别采用双向长短期记忆层和全连接层提取原始数据中的特征, 得到大小为 $[n, 2\times ld]$ 的特征矩阵 $H_1^{(1)}$、$H_2^{(1)}$. 其中, 特征矩阵的上标表示网络层数, 下标表示特征矩阵所属的通道, 单个方向 LSTM 网络输出的特征长度为 ld, 全连接层输出的特征长度为 $2\times ld$. 值得一提的是, 在每个时刻, 每个测点仅分别通过 BiLSTM 和 FC 进行一次特征提取. 由此得到两个输入通道的特征图作为多核图卷积层的输入. 使用多个通道的输入是为了从不同的角度进行特征提取, 也可以采用不同的方式进行设计, 例如不同时间窗口长度的输入通道、不同频率的输入通道等, 可根据实际对象特性进行相应的选取与设计.

在基于 MKGCN 层的特征聚合模块中, 基于本文提出的多核图卷积层实现了全量测点图中节点特征的聚合与更新, 更新后图中每个节点的特征是由其他节点的原特征聚合更新而来, MKGCN 层聚合得到的节点特征矩阵为 MK. 注意, 更新后每个节点的特征不包含其自身原有特征. 将代表多通道过程全量测点图的特征矩阵输入 MKGCN 层, 得到多个通道的输出图.

在特征逼近模块, 对 MKGCN 层输出的各个测点的 c_{out} 个通道、长度为 $4\times ld$ 特征, 应用基于无偏置的全连接层实现的特征逼近层, 将多输出通道的特征压缩为单输出通道的特征, 从而与两个输入通道输入的特征进行特征逼近. 特征逼近模块一方面使得 MKGCN 层具有从其他测点特征重构当前每个测点特征的能力, 更主要的作用是为在线监测阶段的基于特征逼近的自选代方法提供支持.

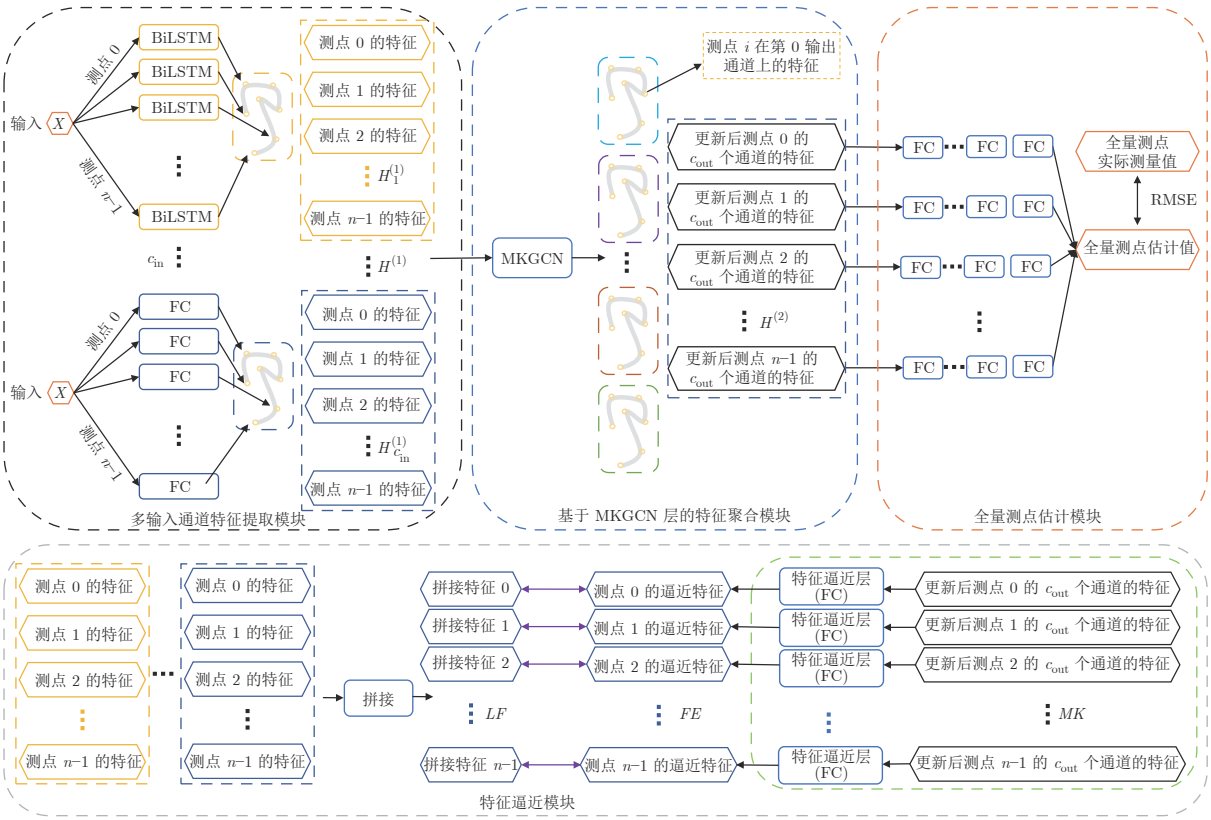


图 2　面向全量测点估计的多核图卷积模型结构

Fig. 2　Structure of multi-kernel graph convolution model for total measurement points estimation

在全量测点估计模块, 针对每一个测点, 基于MKGCN 层输出的特征, 使用由全连接层构成的全量测点估计模块, 得到每个测点 t 时刻的估计值. 值得一提的是, 这里的估计值只是初步的估计值, 在后文中会使用自迭代方法进行修正, 得到更加可靠的估计值, 尤其对于系统异常情况是十分有必要的.

2.2 多核图卷积层

在 MKGCN 层中, 输入通道数为 c_{in}, 节点数为 no_{in}, 每个节点特征长度为 fe_{in} 的图的特征矩阵 $H^{(l)}$, 大小为 $[c_{\text{in}}, no_{\text{in}}, fe_{\text{in}}]$. 输出特征矩阵 $H^{(l+1)}$, 通道数为 c_{out}, 大小为 $[c_{\text{out}}, no_{\text{out}}, fe_{\text{out}}]$. 在这一过程中, 进行图中全部节点特征的聚合与更新. 对于本文中的场景, 输入通道和输出通道的节点数相同, 均为全量测点数 n, 即 $no_{\text{in}} = no_{\text{out}} = n$.

由于工业过程工作机理的复杂性、未知性、随机性, 测点之间的作用关系是部分甚至全部未知的, 从而全量测点图的结构是未知的. 因此对全量测点图的邻接矩阵初始化为对角线为 0, 其余元素全为 1 的矩阵, 如式 (13) 所示. 即将全量测点图初始化为没有自环 (Self-loop) 结构的全连接图, 使得对应于每个测点的节点聚合来自其他所有测点的信息. 可根据机理已知的情况, 调整初始化的方式, 即

$$A = Ones_{n\times n} - I_n \tag{13}$$

其中, $Ones_{n\times n}$ 为元素全为1、大小为 $n \times n$ 的矩阵. 若 A 中的元素 $a_{ij} = 1$, 则变量 j 的信息用于预测变量 i, 即节点 j 的信息流入节点 i. 然而, 邻接矩阵中非对角线的值全为 1, 不能反映测点间相关性的强弱, 因此定义邻接核 $\hat{A}$ 为

$$\hat{A} = A \odot \bar{A} \tag{14}$$

其中, $\bar{A}$ 为形状和 A 相同, 元素均为可学习参数的矩阵; $\odot$ 表示矩阵点乘. 通过将参数矩阵 $\bar{A}$ 和邻接矩阵 A 相乘, 保留 A 初始化的信息, 也使得图中节点间连接权重可被学习, 使得邻接核 $\hat{A}$ 成为一种估计图结构特征的算子. 然而, 对于一个复杂的任务, 仅使用一个邻接核 $\hat{A}$ 并不鲁棒. 在卷积神经网络中[42], 通常使用多个卷积核提取不同的特征, 得到多个通道的输出图, 实现视觉特征的有效提取. 类似地, 在 Transformer 模型中[43], 也引入多头注意力机制来加强模型的表征能力. 受此启发, 本文将图的邻接矩阵视为一种图结构特征估计算子, 构建多核图卷积层, 并通过不同的可学习的邻接核得到全量测点图不同通道的特征, 从而面向图结构未知的任务实现了更充分、更鲁棒的建模. 采用多个包含多邻接核 $\hat{A}$ 的邻接核组更鲁棒地捕获多通道的输入图中包含的复杂过程变量间的耦合关系, 得到多个通道的输出图. 从而得到 MKGCN 层的定义为

$$H_{ij}^{(l+1)} = \hat{A}_{ij} H_i^{(l)} \quad i = 1, 2, \cdots, c_{\text{in}};\ j = 1, 2, \cdots, c_{\text{out}} \tag{15}$$

$$H_j^{(l+1)} = \left(\sum_{i=1}^{c_{\text{in}}} H_{ij}^{(l+1)}\right) W_j \quad j = 1, 2, \cdots, c_{\text{out}} \tag{16}$$

$$H^{(l+1)} = \xi\left(concat\left(H_j^{(l+1)}\right)\right) \quad j = 1, 2, \cdots, c_{\text{out}} \tag{17}$$

其中, $\hat{A}_{ij}$ 为对应于第 i 个输入通道、第 j 个输出通道的邻接核. 这里所描述的通道的概念类似于图像中的 RGB 通道的概念. 其中, 对输入特征矩阵 $H_i^{(l)}$ 左乘邻接核 $\hat{A}_{ij}$, 并在输入通道的维度上叠加, 从而实现节点间多个输入通道的特征聚合; 右乘权重矩阵 W_j, 与经典的图卷积网络一致, 从而实现每个节点自身特征的变换和增强; 两者从不同的维度共同作用实现全量测点图中节点特征的聚合和更新. MKGCN 层的计算过程如图 3 所示 (图中以 $c_{\text{in}} = 3$, $c_{\text{out}} = 2$ 为例).

此外, 与一般的卷积神经网络类似, 在一般性的图学习任务中, MKGCN 也能够多堆叠层使用, 如图 4 所示. 中间层子图的大小受邻接核的大小控制, 当 $a \neq b$, 即邻接核不初始化为方阵时, 图中的节点数会发生变化. 此时中间隐藏层的图中节点不再具有物理意义, 邻接核的物理意义也发生了变化, 不再具备邻接矩阵的物理意义, 而是仅作为一种图结构特征估计和变换的算子, 可初始化为式 (18), 即邻接核中的元素全部初始化为 1.

$$A = Ones_{a\times b} \tag{18}$$

MKGCN 层受到卷积神经网络和图神经网络的启发, 结合多通道技巧, 能够有效聚合和更新图中节点的信息, 且使得模型更加鲁棒, 具有更强的建模能力. 与一般的图卷积方法相比, MKGCN 通过引入可学习的邻接核, 能够适应缺乏图的邻接矩阵先验知识的任务, 且整个流程是端到端的.

2.3 特征逼近

在估计模型的训练过程中, BiLSTM 通道提取出的节点特征矩阵为 $H_1^{(1)}$, FC 通道提取出的特征矩阵为 $H_2^{(1)}$, MKGCN 层聚合得到的节点特征矩阵为 MK. 特征逼近策略如式 (19) ~ (21) 所示:

$$LF = concat(H_1^{(1)}, H_2^{(1)}) \tag{19}$$

$$FE = concat(FC_i(MK_i)) \tag{20}$$

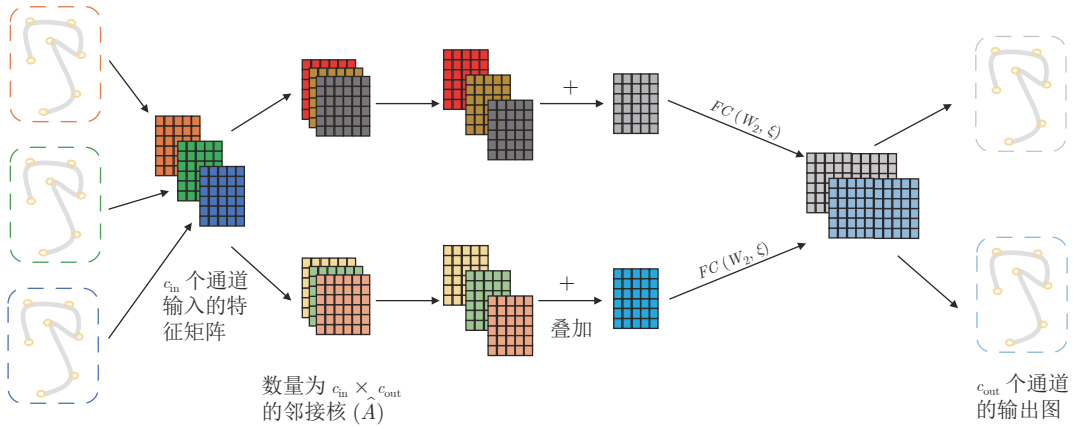


图 3　MKGCN 层的计算过程

Fig. 3　Calculation process of MKGCN layer

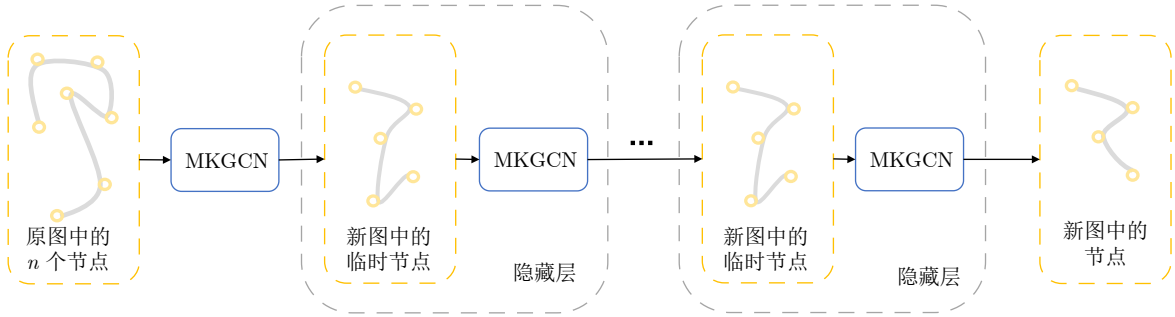


图 4　MKGCN 层的堆叠使用

Fig. 4　The stacking use of MKGCN layers

$$\min J = \left(\underset{i,j}{\text{mean}}\left(|LF_{ij} - FE_{ij}|\right)\right)$$
$$i = 0, \cdots, n-1;\ j = 1, \cdots, 4 \times ld \quad (21)$$

在模型训练阶段, 对 MKGCN 层聚合得到的多通道的节点特征矩阵 MK, 使用各个测点对应的特征逼近层 (FC) 将 c_{out} 个通道的特征压缩为单通道特征后, 将压缩得到的特征与原始数据由 LSTM 和 FC 两个通道提取出的特征再沿特征维度拼接得到的特征进行逼近. 即对于每个节点而言, 其利用其他节点的信息通过 MKGCN 尽可能重构出从其本身原始数据直接提取出的特征.

2.4　模型训练

在模型训练过程中, 一方面优化模型的估计精度; 另一方面保持 MKGCN 层聚合出的更新后节点特征与 FC 层和 BiLSTM 层提取出的原始数据中的特征进行逼近. MKGCN 层、全连接层、BiLSTM 层均包含可学习参数. 权重采用 Xavier 方法[44] 进行初始化, 偏置初始化为 0. 使用均方根误差 (Root mean square error, RMSE) 指标与特征逼近损失的和作为损失函数进行训练. RMSE 指标定义为

$$RMSE(Y, \hat{Y}) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad (22)$$

其中, y_i 为某个测点在 i 时刻真实值, $\hat{y}_i$ 为模型得到的某个测点在 i 时刻的估计值.

$$loss = RMSE + \lambda \times J \quad (23)$$

在训练过程中, 观察到特征逼近损失与估计误差 RMSE 相比有数量级的差异. 为了在整个训练的过程中始终保持特征逼近的目标, 设置超参数 λ 随训练轮数呈线性增加.

$$\lambda = epoch + 10 \quad (24)$$

其中, $epoch$ 为当前训练进行到的轮数.

2.5　方法总结

本方法适用于多输入多输出的复杂工业设备全量测点的同步估计. 根据已知的机理信息, 可对过程全量测点图结构, 即邻接核的初始化方式进行调整. 基于 MKGCN 层的估计模型通过引入多邻接

核技巧构建多通道输出, 一方面提高模型的鲁棒性和表征能力; 另一方面避免了基于图的方法需要图的邻接矩阵作为先验的要求, 使得整个过程为端到端的, 易于实现. 本文提出的估计模型捕获了测点时序特性和测点间的耦合关系, 高效地实现了全量测点同步估计. 基于此, 在后文中将进一步提出基于估计偏差的控制限, 并应用于过程监测.

3 基于特征逼近的自迭代方法与监测方法

3.1 基于特征逼近的自迭代方法

测点代表的物理量之间存在生产机理上的相关性, 模型学习到这一相关性对于正常运行情况下的估计是十分有帮助的. 然而, 在应用阶段, 当异常情况发生时, 可能会因此导致模型对于正常测点的估计出现异常, 即产生误报警; 或导致模型对其他异常测点的估计出现异常, 即产生漏报警. 因此, 本文设计了一种基于特征逼近的自迭代方法, 其核心思想是对重构误差较大的测点, 通过使用 MKGCN 层聚合重建得到的相对正常的测点特征不断替换从原始数据提取出的异常特征, 从而消除掉异常测点对其他测点进行估计产生的干扰.

基于特征逼近的自迭代方法如图 5 所示, 仅以单个测点异常的情况为例. 图中虚线表示特征的复制操作, 且仅画出一个输入通道, 其余通道省略. 图中测点的值或特征处的符号“×”数量越多, 表示该测点的值或特征越偏离系统正常运行状态下得到的

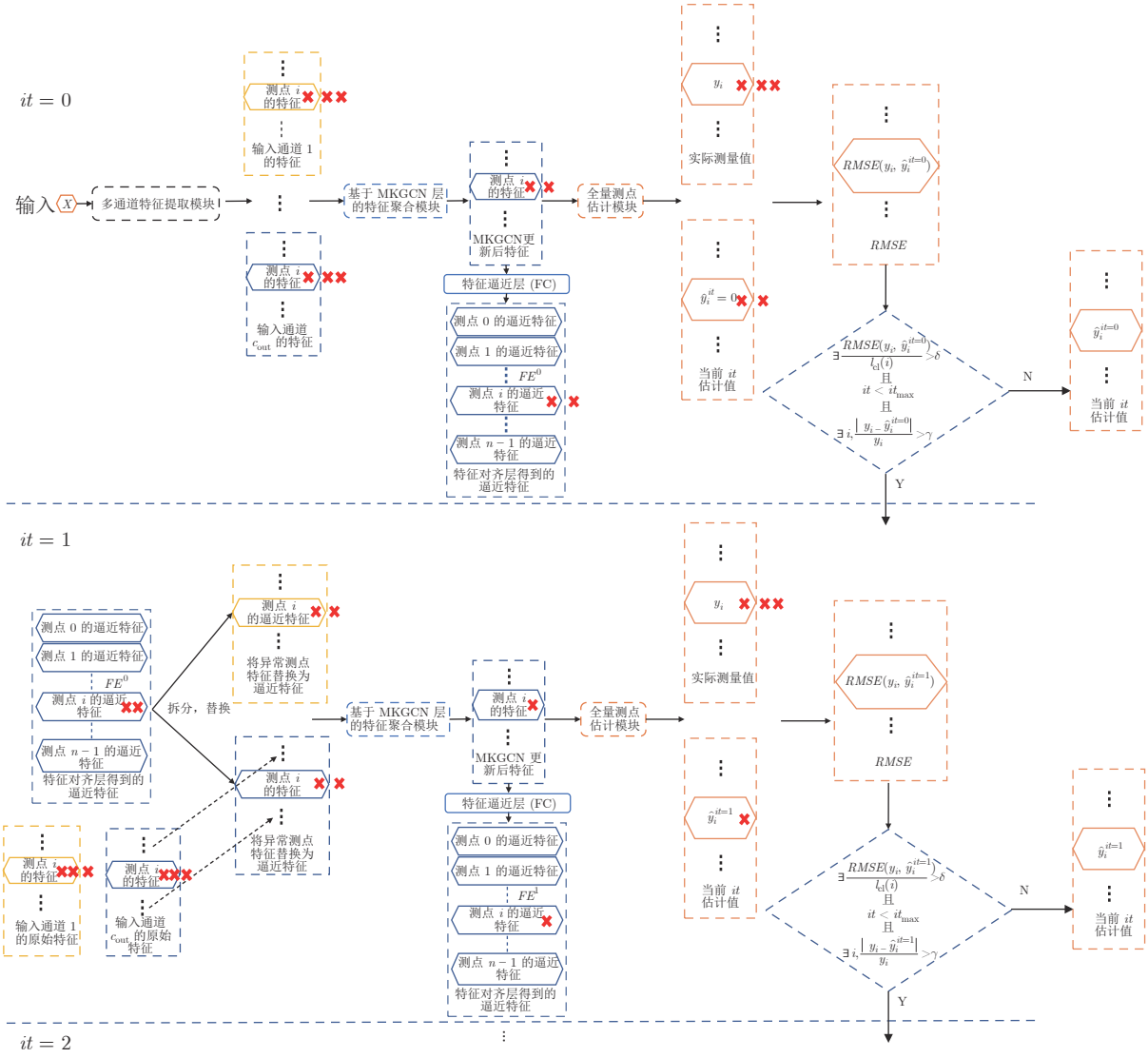


图 5 自迭代方法

Fig. 5 Self-iterative method

值或特征. 测点 i 的控制限 $l_{\rm cl}(i)$ 的计算方式见第 3.3 节. δ, γ 为设定的超参数, $\hat{y}_i^{it}$ 为当前时刻测点 i 在第 it 次迭代时的估计值.

当迭代次数 $it=0$ 时, 基于原始输入数据, 使用面向全量测点估计的多核图卷积模型得到当前迭代的估计值. 当异常工况下某些测点出现显著异常, 或者因某些干扰因素导致某些测点的实测值暂时缺失时, 某些测点的估计误差超过了控制限, 说明此时系统可能发生了异常. 需要对异常测点在系统正常情况下的值进行估计, 从而精准判断出现异常的位置和异常程度. 由于异常的测量值也被用于估计其他测点假定在系统正常情况下的估计值, 即异常的测点值对其他测点值的估计产生了干扰, 需要触发自迭代方法, 对于估计值进行纠错, 从而在优化异常测点的估计值的同时降低对正常测点的误报情况. 在自迭代方法中, 随着迭代过程的进行, 出现异常或缺失的测点的特征被其他测点估计出的该测点特征进行替换, 从而逐步消除该异常测点特征对其他测点估计的干扰. 随着迭代次数的增加, 存在异常的特征中包含的异常信息被逐步丢弃, 从而得到当前时刻正常情况下各个测点的估计值.

本文采用了两种可调控的阈值去控制最大的迭代次数, 从而满足系统的实时性要求. 一方面, 定义了最大迭代次数 $it_{\max}$, 当特征迭代次数超过设定的最大迭代次数 $it_{\max}$ 时, 停止迭代; 另一方面, 定义了估计变化阈值 γ, 当两次迭代得到的估计值变化小于阈值 γ 时, 说明此时迭代已经趋于收敛, 可以停止迭代. 一般来说, 在迭代次数较少时收敛效果已经较好, 当迭代次数更多时能够得到更加准确的估计值, 但也会增加估计耗时. 因此, 具体的迭代次数和阈值根据传感器采样的频率和控制要求选定.

值得一提的是, 该自迭代方法不适用于基于过程记忆矩阵的 MEST 类方法. 对于 MEST 类方法, 当模型估计出现异常时, 由于上一次迭代产生的估计值已经偏离了正常情况, 再次迭代过程中, MEST 方法重新计算相似度时会使得当前估计状态与部分本不相似的状态的相似程度进一步提高, 导致估计值异常. 对于本文提出的方法, 由于测点间的耦合关系通过 MKGCN 层中的参数进行建模, 且自迭代过程中 MKGCN 层的参数是固定的, 因此会随着迭代过程的进行, MKGCN 层聚合出的节点特征迭代替换异常的节点特征, 只保留符合测点间耦合关系的特征, 从而逐步消除异常节点特征对于其他节点估计值的干扰. 基于特征逼近的自迭代方法步骤如算法 1 所示.

算法 1. 基于特征逼近的自迭代方法

1) $it=0$, 得到初始估计值 $\hat{y}_0^0, \hat{y}_1^0, \cdots, \hat{y}_{n-1}^0$, 初始特征矩阵 FE^0
2) 定义迭代过程中临时特征为 TMP
3) for it in $[0, 1, \cdots, it_{\max}]$:
4) $\quad TMP = LF$
5) $\quad flag = \text{False}$
6) $\quad$ for i in n
7) $\quad\quad$ if $\frac{RMSE(y_i, \hat{y}_i^{it})}{l_{\rm cl}(i)} > \delta$
8) $\quad\quad\quad TMP_{i,} \cdots = FE_{i,}^{it} \cdots$
9) $\quad\quad\quad flag = \text{True}$
10) $\quad$ if not $flag$
11) $\quad\quad$ break
12) $\quad$ else:
13) $\quad\quad \hat{y}^{it+1}, FE^{it+1} = \text{MKGCN}(TMP)$
14) $\quad$ if $\forall i$, $\frac{\left|\hat{y}_i^{it+1} - \hat{y}_i^{it}\right|}{\left|\hat{y}_i^{it}\right|} < \gamma$
15) $\quad\quad$ break

3.2　对自迭代多核图卷积模型的进一步讨论

在本节中, 将对自迭代方法中异常特征的消除和正常特征的保留这两个问题从理论方面作进一步的讨论.

1) 异常特征的消除

首先, 讨论在自迭代过程中, 多核图卷积模型是如何消除异常特征的. 回顾 Kipf 等[34] 提出的图卷积网络:

$$H^{(l+1)} = \xi\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \tag{25}$$

图卷积网络具有过平滑现象, 深层的 GCN 会使得节点特征相近, 难以区分. 文献 [45] 表明, 图卷积网络是一种特殊形式的拉普拉斯平滑. 对于一个没有二部分量的图, 当 GCN 层数过多时, 会使得同一连通分量的节点的特征收敛于同一个值, 即

$$L = D - A \tag{26}$$

$$L_{\rm sym} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}} \tag{27}$$

$$\mathbf{1}_j^{(i)} = \begin{cases} 1, & v_j \in C_i \\ 0, & v_j \notin C_i \end{cases} \tag{28}$$

$$\lim_{m\to+\infty}(I - \beta L_{\rm sym})^m \boldsymbol{w} = D^{-\frac{1}{2}}\left[\mathbf{1}^{(1)}, \mathbf{1}^{(2)}, \cdots, \mathbf{1}^{(k)}\right]\boldsymbol{\theta} \tag{29}$$

其中, C_i 表示图中第 i 个连通分量, v_j 表示图中第 j 个节点, β 为损失权重, $\beta \in (0, 1]$, $\boldsymbol{w} \in \mathbf{R}^n$, $\boldsymbol{\theta} \in \mathbf{R}^k$. 对于本文提出的多核图卷积层, 首先考虑一种简化的情况, 即 $c_{\rm in}=1$, $c_{\rm out}=1$, 此时其退化为

$$H^{(l+1)} = \xi\left(\hat{A}H^{(l)}W^{(l)}\right) \tag{30}$$

可以注意到, 式 (25) 与式 (30) 具有类似的形式. 区别在于, GCN 中采用的卷积算子 $\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ 为正则化的图邻接矩阵, 而在本文提出的工况估计模型中, MKGCN 的邻接核 $\hat{A}$ 是学习得到的, 且主对角线元素始终为 0. 由于训练得到的权重参数一般不为 0, 因此可认为图中只有一个连通分量, 即测点图是全连通的. 考虑到本文采用的 Tanh 激活函数, 在原点附近满足 $\mathrm{Tanh}(x) \approx x$, 在自变量较大时趋于平缓, 值域在 $(-1, 1)$ 之间, 且已对网络的输入输出进行了 min-max 标准化, 权重采用 Xavier 方法进行初始化, 故在本节的讨论中, 认为 MKGCN 层的运算是近似线性的, 近似满足叠加原理.

在自迭代过程中, 考虑每次迭代使用的特征, 正常测点的特征均采用其原始特征, 而异常测点的特征采用的是上一次迭代得到的更新特征. 将特征 FE 分解为符合模型的特征 (即各个测点在系统正常情况下数据产生的特征) FE_c 和不符合模型的异常特征 (即测点异常波动产生的特征) FE_{ic}, 即

$$FE = FE_c + FE_{ic} \tag{31}$$

对于自迭代过程, 与循环神经网络类似, 可以将自迭代过程展开为权重一致的 MKGCN 层与线性层的串接. 对 $c_{\mathrm{in}} = 1$ 和 $c_{\mathrm{out}} = 1$ 的情况, 此时 MKGCN 层间无偏置的特征逼近层中原有的压缩多通道的作用消失, 仅起到了缩放的作用, 定义该参数为 ω, 并令 $W \times \omega = W'$. 考虑异常特征 FE_{ic}, 其经过自迭代过程后, 可得

$$H_{ic}^{m} = \hat{A}^{m} FE_{ic} W'^{m} \tag{32}$$

Chiang 等[46] 采用一种强化自连接的技巧, 通过在正则化后的卷积算子上再增加一个单位矩阵 I, 使得图卷积的过程中更多地保留自身特征, 能够一定程度上缓解图卷积的过平滑问题. 本文的邻接核 $\hat{A}$ 与之相反, 由于强制约束主对角线元素为 0, 一定程度上加剧了过平滑现象, 且由于图全连通, 当自迭代次数充分多, 即 m 充分大时, 一般会使得 H_{ic}^{m} 中的元素趋于 0. 可以从矩阵范数的角度理解这一点, 由于矩阵乘积的范数小于等于矩阵范数的乘积, 即

$$\|\hat{A}^{m}\|_1 \leq \|\hat{A}\|_1^{m} \tag{33}$$

当 $\|\hat{A}\|_1 < 1$, 且 $m \to \infty$ 时, $\|\hat{A}\|_1^{m} \to 0$, 进而 $\|\hat{A}^{m}\|_1 \to 0$, 即 $\hat{A}^{m}$ 中元素趋近于 0. 进而可以得知, 多次通过 MKGCN 层得到的特征 H_{ic}^{m} 中元素趋于 0, 即起到了消除异常特征的作用. 为了保证 $\|\hat{A}\|_1 < 1$, 本文建议当模型效果不佳时, 对训练好的邻接核进行检验. 如不满足, 则在网络层间引入权重的正则化约束, 重新训练.

2) 正常特征的保留

接下来, 讨论在自迭代过程中, 多核图卷积模型是如何保留正常特征的. 一方面, 对于重构误差未超限的测点, 在自迭代过程中, 始终使用其原始特征; 另一方面, 在训练过程中, 引入特征逼近约束, 即理想情况下有

$$FE_c = \hat{A}FE_cW' \tag{34}$$

即通过逼近约束使得自迭代过程中, 符合模型的特征尽可能保持不变. 这两方面共同作用使得自迭代过程中符合模型的特征尽可能不损失.

综上, 讨论了 $c_{\mathrm{in}} = 1$ 和 $c_{\mathrm{out}} = 1$ 情况下自迭代过程的效果. 在 $c_{\mathrm{in}} \neq 1$ 或 $c_{\mathrm{out}} \neq 1$ 的一般情况下, 由于 MKGCN 层近似满足叠加原理, 对于 $c_{\mathrm{in}} \neq 1$ 的情况, 可视为多个单通道情况的叠加; 对于 $c_{\mathrm{out}} \neq 1$ 的情况, 特征逼近层的输出特征即为多个单输出通道的线性组合. 对于一般情况下, 当模型效果不理想时, 本文仍建议对学习到的邻接核进行检查. 需要注意的是, 自迭代方法也并不能保证特征中异常的部分完全去除、正常的部分完全保留. 这主要由于以下原因造成: 1) 所使用的工况估计模型并没有有效学习到系统正常情况下测点间耦合关系, 从而无法准确区分正常特征与异常特征. 这也是 MKGCN 层的意义所在, 通过多通道的设计, 对测点间关系进行更加充分的建模; 2) 系统本身并非真正线性, 而是存在非线性因素; 3) 未报警的测点特征可能也包含异常因素, 尽管在每次迭代过程中完全使用更新后的特征而非保留部分原始特征, 可能有助于解决这一问题, 然而这也会带来潜在的风险. 因为特征逼近为软约束, 正常特征在经过 MKGCN 和特征逼近层时也会产生少许损失, 因此实际情况中需要权衡以上因素.

3.3 基于估计偏差的监测指标与在线监测步骤

考虑来自实际生产过程的验证数据中也可能包含少量噪声, 故采用核密度估计 (Kernel density estimation, KDE)[47] 方法确定控制限. 定义第 j 个测点在 i 时刻测量值与估计值的偏差 l_i 为

$$l_i = \left|y_i^j - \hat{y}_i^j\right| \quad i = 1, 2, \cdots, m; \quad j = 1, 2, \cdots, n \tag{35}$$

其中, n 为测点总数, y_i^j 为第 j 个测点的当前 i 时刻的值. 基于训练好的模型, 分别对每一个测点在仅包含系统正常运行数据的验证集上估计偏差情况进行核密度估计, 得到第 j 个测点的误差控制限 l_{cl}^j, 即

$$\hat{f}_h = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{I - I_i}{h}\right) \tag{36}$$

$$l_{\mathrm{cl}}^{j} = KDE\left(\left\{l_i^j\right\}_{i=1}^{m}, significant_{\mathrm{level}}\right) \tag{37}$$

其中, $\hat{f}_h$ 为根据偏差数据估计得到的概率密度函数, 核函数 K 选用高斯核函数. 在本文中, 设定带宽 $h = 0.01$, 显著性参数 $significant_{\mathrm{level}} = 0.98$, 该值可根据实际调整.

如果在训练和测试阶段模型能够对处于正常状态的测点进行准确有效的估计, 则当应用阶段的模型估计值与实际测量值之间产生的估计偏差 RMSE 指标超过模型训练阶段得到的误差控制限 l_{cl} 时, 说明当前该生产过程可能存在异常, 需要触发报警. 基于自迭代多核图卷积模型的工况监测模型的应用步骤如算法 2 所示.

算法 2. 基于 MKGCN 模型的工况监测模型算法

模型开发阶段

1) 从分布式控制系统和历史记录中收集训练数据, 数据包括工业设备运行过程中全量测点的数据. 在训练阶段, 仅使用工业设备正常运行时的数据进行建模和验证. 将数据 min-max 规范化到 [0, 1], 对每一个指标序列 $x_1, x_2, \cdots, x_n$, 规范化方法为

$$y_i = \frac{x_i - \min\limits_{1\le j\le n}\{x_j\}}{\max\limits_{1\le j\le n}\{x_j\} - \min\limits_{1\le j\le n}\{x_j\}}$$

2) 基于收集到的数据, 对参数使用 Xavier 方法进行初始化, 根据式 (23) 定义的损失函数, 使用反向传播算法训练基于 MKGCN 的估计模型.

3) 在验证集上, 对训练好的估计模型, 计算估计偏差, 使用 KDE 估计得到控制限.

在线应用阶段

4) 采用与步骤 1) 相同的方法收集 t 时刻及 $t-1, t-2, \cdots, t-len+1$ 时刻长度为 len 的测量数据, 并使用与训练阶段相同的方式对数据进行标准化. 根据训练好的基于 MKGCN 层的估计模型, 对当前 t 时刻的全量测点值进行估计, 得到估计值.

5) 使用自迭代算法迭代替换异常特征, 对测点估计值进行优化.

6) 根据估计值和当前各个测点的传感器得到的实测值, 计算偏差, 根据控制限确定当前状态是否超限, 如超限则触发报警.

4　案例研究

在本节中, 基于某燃煤电厂中引风机的实际生产数据, 对本文提出的基于 MKGCN 与自迭代策略的全量测点监测模型进行验证. 分别介绍所使用的数据、模型实现以及取得的效果.

4.1　数据说明

引风机是电厂中百万千瓦超超临界机组中的重要辅机之一, 通过改变静叶角度, 即进口挡板开度, 控制炉膛压力, 将烟气吸入烟道, 使锅炉内形成负压, 从而增加氧气, 使燃料燃烧更充分. 一旦引风机发生故障, 可能导致主机停机, 影响电厂正常供电. 该引风机共有 33 个测点 (变量), 对应的物理量如表 1 所示.

本文采用引风机正常运行时的数据训练模型并测试估计效果; 使用同时包含正常和异常工况的数据检验工况监测效果. 数据进行了 min-max 标准化, 如式 (38) 所示 .

$$y_i = \frac{x_i - \min\limits_{1\le j\le n}\{x_j\}}{\max\limits_{1\le j\le n}\{x_j\} - \min\limits_{1\le j\le n}\{x_j\}} \tag{38}$$

将用于建模的数据按照时间顺序, 前 80% 划分

表 1　引风机测点对应表

Table 1　Measuring points of induced draft fan

测点编号	物理量	测点编号	物理量	测点编号	物理量
0	功率信号三选值	11	引风机水平振动	22	引风机油箱温度
1	进气温度	12	引风机后轴承温度 1	23	引风机中轴承温度 1
2	引风机电机定子线圈温度 1	13	引风机后轴承温度 2	24	引风机中轴承温度 2
3	引风机电机定子线圈温度 2	14	引风机后轴承温度 3	25	引风机中轴承温度 3
4	引风机电机定子线圈温度 3	15	引风机键相	26	炉膛压力
5	引风机电机水平振动 1	16	引风机静叶位置反馈	27	引风机出口风温
6	引风机电机水平振动 2	17	引风机前轴承温度 1	28	引风机入口压力
7	引风机电机轴承温度 1	18	引风机前轴承温度 2	29	引风机出口风压
8	引风机电机轴承温度 2	19	引风机前轴承温度 3	30	引风机静叶开度指令
9	引风机电流	20	引风机润滑油温度	31	总燃料量
10	引风机风垂直振动	21	引风机润滑油压力	32	炉膛压力

为训练集, 后 20% 划分为测试集, 从而符合实际的应用场景. 所用的数据均为每 3 min 采样一次. 用于建模的数据均在正常工况下实际采集得到, 采样时间长度约为 8 天, 内含 4000 个采样点, 包含了该引风机的主要工作功率范围 (400 ~ 1 000 MW), 以及多次负载缓慢或快速上升、下降时的数据. 此外, 一段包含正常运行情况和异常运行情况的长度为 300 的数据 (下文称为监测数据集) 用于评估模型的监测性能. 根据测点值是否存在异常升高或者下降, 测点的故障分为高报警和低报警两种. 在该监测数据集中, 由于环境温度波动、冷却效果不理想、润滑油温高, 出现了短期异常工况, 测点 1、4、8、12、13、14、20、27 产生高报警异常, 其余 25 个测点相应的变量正常. 定义变量集合 V, 集合内元素为全部 33 个变量; 定义异常变量集合 F, 集合内元素为 8 个出现异常的变量; 定义正常变量集合 N, 集合内元素为全部正常变量. 为了保证系统的监测可靠性, 本文所提方法在建模阶段使用的数据应尽可能包括测试数据中可能出现的健康状态. 若实际应用过程中发现模型的监测可靠性出现下降, 可使用新收集到的数据对模型进行调整与更新, 从而提升模型对于当前状态的适应性与可靠性.

前文提到反映工业设备运行状态的测点间具有一定的相关性. 首先使用了斯皮尔曼等级相关系数 (Spearman′s rank correlation coefficient, SRCC)[48] 的绝对值去评估两个指标之间的相关性. 两个指标之间的斯皮尔曼相关性等于这两个指标的秩值之间的皮尔逊相关性. SRCC 基于秩序列, 是非参数的, 适合非线性的对象. 斯皮尔曼等级相关系数的绝对值定义为

$$\rho_S(\boldsymbol{r},\boldsymbol{s})=\left|\frac{\sum\limits_{i=1}^{N}(r_i-\overline{r})(s_i-\overline{s})}{\sqrt{\sum\limits_{i=1}^{N}(r_i-\overline{r})^2\sum\limits_{i=1}^{N}(s_i-\overline{s})^2}}\right| \tag{39}$$

其中, $\boldsymbol{r}$, $\boldsymbol{s}$ 代表任意两个不相同的测点 $\boldsymbol{p}$, $\boldsymbol{q}$ 的阶次序列. $\boldsymbol{r}$ 中的一项 r_i 是将 $\boldsymbol{p}$ 升序排列后 p_i 的阶次, $\boldsymbol{s}$ 中的一项 s_i 是将 $\boldsymbol{q}$ 升序排列后 q_i 的阶次.

如上文所述, 本文使用斯皮尔曼等级相关系数来评估数据集中各个测点之间的相关性, 画出 SRCC 热力图矩阵如图 6 所示. 从热力图矩阵可以看出, 对角线周围呈现出多个矩形高热度区域, 这是正常的. 从表 1 中可以看出, 这是由于多个物理意义相近的变量, 其编号也靠近, 而在正常工况下它们的波动情况显然也应该是近似的. 例如, 当设备处于高负荷工作状态时, 多个轴承的温度都会同步升高.

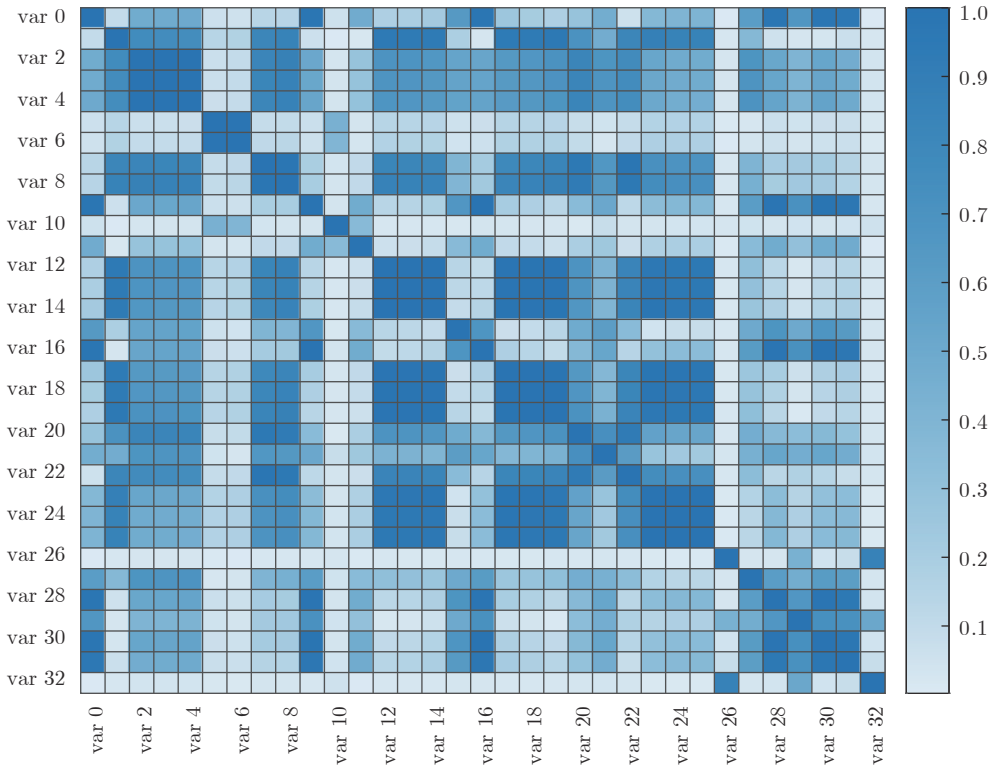


图 6 训练数据中测点间相关性

Fig. 6 Correlation between measuring points on training data

4.2　全量测点误差估计和监测实验

参照 Feng 等[35] 的实验设计, 将本文所提出的基于 MKGCN 的全量测点估计模型和 7 种常用的参考模型进行比较. 值得一提的是, 本文期望得到每个测点的估计值, 从而为一线人员提供可直观理解的有效参考, 故选择的对比方法也需要能够提供每个测点的估计值. 所选择的对比方法包括: 偏最小二乘回归 (Partial least squares regression, PLSR)[49]、极限学习机 (ELM)[50]、全连接网络 (FC)[44]、长短期记忆网络 (LSTM)[38]、一维卷积网络 (Conv1D)[51]、图卷积网络 (GCN)[34] (GCN 的实现同样不引入自环特性, 从而与 MKGCN 方法保持一致) 和多变量状态估计方法 (MEST)[27]. 包含多种常用的应用于软测量任务的线性、非线性、基于图卷积网络层的全量测点同步估计模型, 以及经典的多变量状态估计方法.

基于 MKGCN 的全量测点估计模型的模型结构实现如表 2 所示, 输入通道设计为分别来自 BiLSTM 和 FC 的两个通道, 输出通道为超参数, 通过网格搜索确定. 同样基于图的 GCN 模型为单通道输入、单通道输出, 实现方式如表 3 所示. 由于模型的性能容易受到实现方式的影响, 使用公开的 Python 包实现了 PLSR、ELM 方法, 使用 PaddlePaddle 框架实现了 FC、LSTM、Conv1D、GCN、MKGCN 网络, 如表 4 所示. 其中 nc 为保留的组件数量; E 为特征数; α 为损失权重; ld 为隐藏层维度 (即特征长度); oc 为邻接核组数, 即输出图通道数. 所有深度神经网络方法中的层数都设置为 5. 除非特别指定, 否则选择 Tanh 作为激活函数. 将 BiLSTM 的时间窗步数和 Conv1D 的核大小 len 设置为 4. 此外本文复现了经典的 MEST 方法. 为了保证实验的公平性, 所有模型中影响性能的主要超参数采用网格搜索的方法进行确定.

FC、LSTM、Conv1D、GCN 和 MKGCN 是深度神经网络, 使用自适应矩估计 (Adam) 方法进行训练. 每种深度神经网络均训练 200 轮. 初始学习率设置为 0.001, 学习率每 25 轮衰减为原来的 50%, 批次大小设置为 64. 其中, 除了 GCN、MKGCN、MEST 方法为同时对全量测点进行估计外, 其余对比模型对某一个测点为主导变量、其余测点为辅助变量的情况, 均进行重新训练, 故所需的模型数与实际的测点数相同. 在 MEST 方法的实现过程中, 可能由于矩阵奇异导致无法正常求解逆矩阵, 因此本文对求逆的部分增加了矩阵 ρI, 其中 $\rho = 0.5$, I 为单位阵, 从而防止需要求逆的矩阵奇异引起估计值显著异常.

RMSE、平均绝对误差 (Mean absolute error, MAE) 指标可用于评估每个指标的预测值和真实值间的距离. 定义真实值序列为 $\boldsymbol{y} = \{y_1, y_2, \cdots, y_m\}$, 预测值序列为 $\hat{\boldsymbol{y}} = \{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_m\}$, 其中, m 为序列长度.

表 2　基于 MKGCN 层的工况估计模型结构

Table 2　Structure of working condition estimation model based on MKGCN layer

序号	网络层	数目	参数	激活函数
1	BiLSTM	n	[input_size = len, hidden_size = ld]	None
	FC	n	[input_size = len, output_size = $2 \times ld$]	
2	MKGCN	1	[$c_{\text{in}} = 1$, $no_{\text{in}} = n$, $fe_{\text{in}} = 2 \times ld$, $c_{\text{out}} = oc$, $no_{\text{out}} = n$, $fe_{\text{out}} = 4 \times ld$]	Tanh
3	FC 0	n	[input_size = $4 \times ld$, output_size = $2 \times ld$]	Tanh
4	FC 1	n	[input_size = $2 \times ld$, output_size = 1]	Tanh
5	FC 2	n	[input_size = oc, output_size = 1]	None
6	特征逼近层 (FC)	n	[input_size = oc, output_size = 1]	None

表 3　基于 GCN 的工况估计模型结构

Table 3　Structure of working condition estimation model based on GCN

序号	网络层	数目	参数	激活函数
1	BiLSTM	n	[input_size = len, hidden_size = ld]	None
2	GCN	1	[in_feature = $2 \times ld$, out_feature = $4 \times ld$]	Tanh
3	FC 0	n	[input_size = $4 \times ld$, output_size = $2 \times ld$]	Tanh
4	FC 1	n	[input_size = $2 \times ld$, output_size = ld]	Tanh
5	FC 2	n	[input_size = ld, output_size = 1]	None

表 4 模型实现和参数网格搜索范围

Table 4 Model implementation and parameter grid search range

方法	Python 包	超参数	超参数调整范围
PLSR	scikit-learn	nc	$nc = \{5, 10, 15, 20, 25\}$
ELM	D.C. Lambert	E, α	$E = \{50, 100, 150, 200, 250\}$, $\alpha = \{0.1, 0.3, 0.5, 0.7, 0.9\}$
FC	PaddlePaddle	ld	$ld = \{8, 16, 32, 64, 128\}$
BiLSTM	PaddlePaddle	ld	$ld = \{8, 16, 32, 64, 128\}$
Conv1D	PaddlePaddle	ld	$ld = \{8, 16, 32, 64, 128\}$
GCN	PaddlePaddle	ld	$ld = \{8, 16, 32, 64, 128\}$
MKGCN	PaddlePaddle	ld, oc	$ld = \{8, 16, 32, 64, 128\}$, $oc = \{2, 4, 8, 16, 32\}$

RMSE 指标定义如式 (22) 所示, 其基于 L_2 范数距离评估了软测量的预测误差.

MAE 指标定义如下:

$$MAE\left(\boldsymbol{y}, \hat{\boldsymbol{y}}\right) = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i| \tag{40}$$

MAE 基于 L_1 范数距离评估了软测量的预测误差.

本文使用误报率、漏报率、F1 值评估模型的监测性能, 其中 F1 值定义如式 (41) 所示. $Accuracy$, $Recall$, $False_{\rm p}$, $False_{\rm n}$ 分别为正确率、召回率、误报率、漏报率.

$$\mathrm{F1} = \frac{Accuracy \times Recall \times 2}{Accuracy + Recall} \tag{41}$$

$$Recall = 100\% - False_{\rm n} \tag{42}$$

$$Accuracy = 100\% - False_{\rm p} \tag{43}$$

使用网格搜索得到的各种模型的最优超参数如表 5 所示.

测试数据上不同模型的估计结果如表 6 和表 7 所示 (表中数据为各测点性能指标的平均值), 表中 N 表示没有出现异常的测点集合, F 表示存在异常的测点集合. 可以看出, 在多种深度神经网络模型中, 本文提出的基于 MKGCN 模型在最优超参数下所需的模型参数数量较少. 且只需建立单个模型即可估计全部测点, 而无需建立与测点数相同数量的模型. MKGCN 模型通过引入多通道特性替代模型提高估计精度的过程中对于高维特征的依赖, 从而节约了模型所需参数, 模型更加轻量, 且能通过

表 5 网格搜索结果与深度神经网络方法在最优超参数下总参数量

Table 5 Grid search results and total parameters of depth neural network method with optimal hyperparameters

方法	最优超参数	模型数	总参数量
PLSR	$nc = 15$	n	/
ELM	$E = 200$, $\alpha = 0.9$	n	/
MEST	/	/	/
FC	$ld = 128$	n	5×10^5
BiLSTM	$ld = 128$	n	6.9×10^6
Conv1D	$ld = 128$	n	9×10^5
GCN	$ld = 64$	1	9.8×10^6
MKGCN	$ld = 8$, $oc = 32$	1	1.8×10^5

表 6 测试数据上不同模型的工况估计结果 (RMSE)

Table 6 Results of different working condition estimation models on test data (RMSE)

变量	PLSR	ELM	FC	BiLSTM	Conv1D	GCN	MEST	MKGCN
var $\in N$	0.042	0.064	0.059	0.052	0.060	0.042	0.005	0.044
var $\in F$	0.046	0.076	0.059	0.049	0.082	0.049	0.006	0.046

表 7 测试数据上不同模型的工况估计结果 (MAE)

Table 7 Results of different working condition estimation models on test data (MAE)

变量	PLSR	ELM	FC	BiLSTM	Conv1D	GCN	MEST	MKGCN
var $\in N$	0.034	0.052	0.049	0.043	0.051	0.034	0.004	0.036
var $\in F$	0.039	0.066	0.050	0.041	0.070	0.043	0.005	0.039

单模型对全量测点进行同步估计, 每个变量特征的复用也起到了节约模型参数的作用.

由表 8 和表 9 可知 (表中数据为各测点性能指标的平均值), 基于 MKGCN 的估计模型尽管也存在一定程度的误报警和漏报警现象, 但对正常测点和异常测点取得了显著优于其他对比方法的监测效果, 其监测效果如图 7 所示. 其中对于每一个测点, 第 1 行图中的短横线为该测点的重建误差, 实线为误差控制限; 第 2 行图中的实线为实测值, 短横线为模型得到的测点的估计值. GCN 模型的监测效果也较好, 这说明基于图的方法由于显式地考虑了测点间的耦合关系, 且倾向于使得估计结果整体最优, 从而在故障发生时能够在准确地反映变量出现异常的同时尽量避免了异常变量和正常变量估计的相互影响. 与 GCN 方法相比, MKGCN 方法在提升了对异常测点的监测准确性的同时, 显著降低了对于正常测点的误报现象. 对于 MEST 方法, 当故障发生时, 由于系统特性的改变, 导致观测向量偏离正常工作空间; 由于异常测点干扰了相似性的计算, 无法使用记忆矩阵中历史向量的组合准确估计出异常情况下各个测点, 尤其当系统多变量异常时, 计算相似度的过程倾向于"迎合"异常的测点, 从而对异常测点出现了显著的漏报现象, 如图 8 所示. 对于其他对比方法, 由于在建模过程中每个模型仅以单测点估计精度为目标进行优化, 缺乏对于系统当前整体状态的建模和估计, 导致异常发生时正常测点的估计和异常测点的估计容易出现相互干扰, 都不能实现准确估计, 从而出现了更多的误报、漏报现象.

4.3　与自编码器的性能对比

本文提出面向全量测点的基于估计的监测模型, 从输入输出的角度看, 与自编码器是类似的. 应用自编码器进行时间序列异常检测的核心思想在于使用正常数据训练的自编码器, 学习正常数据内的

表 8　监测数据上各监测指标 (var $\in N$)

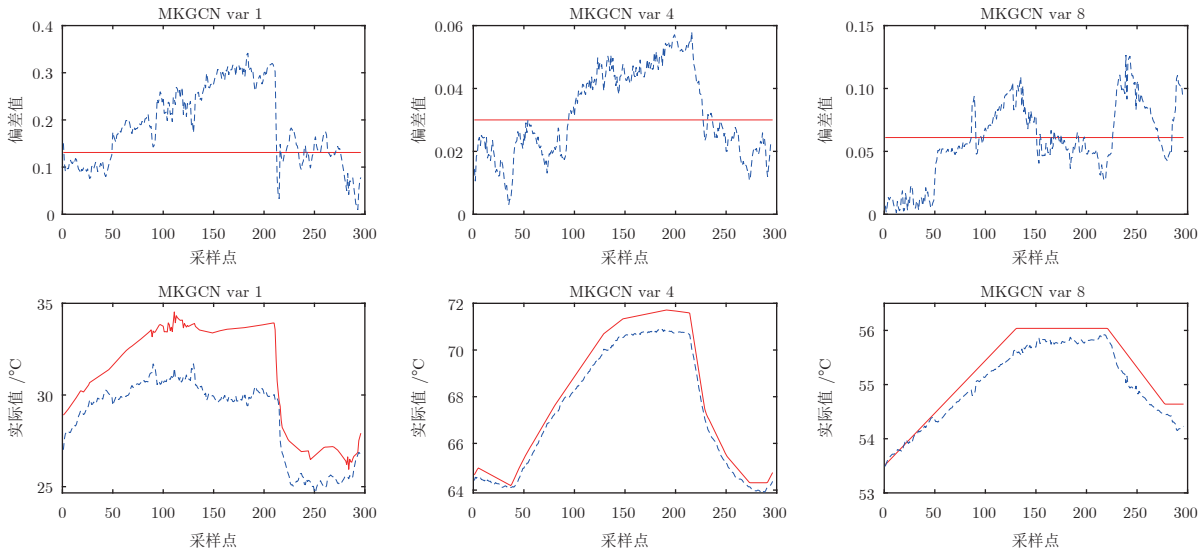
Table 8　Monitoring indicators on monitoring data (var $\in N$)

指标	PLSR	ELM	FC	BiLSTM	Conv1D	GCN	MEST	MKGCN
$False_{\mathrm{p}}$	13.267	29.573	34.267	27.392	42.581	23.568	2.853	4.500
$False_{\mathrm{n}}$	0	0	0	0	0	0	0	0
F1	92.895	82.648	79.324	84.131	72.951	86.642	98.553	97.698

表 9　监测数据上各监测指标 (var $\in F$)

Table 9　Monitoring indicators on monitoring data (var $\in F$)

指标	PLSR	ELM	FC	BiLSTM	Conv1D	GCN	MEST	MKGCN
$False_{\mathrm{p}}$	15.958	30.583	31.375	32.390	37.162	10.769	0	10.769
$False_{\mathrm{n}}$	24.250	5.042	5.917	1.140	1.774	6.968	33.208	1.056
F1	79.681	80.203	79.362	80.302	76.644	91.092	80.090	93.836



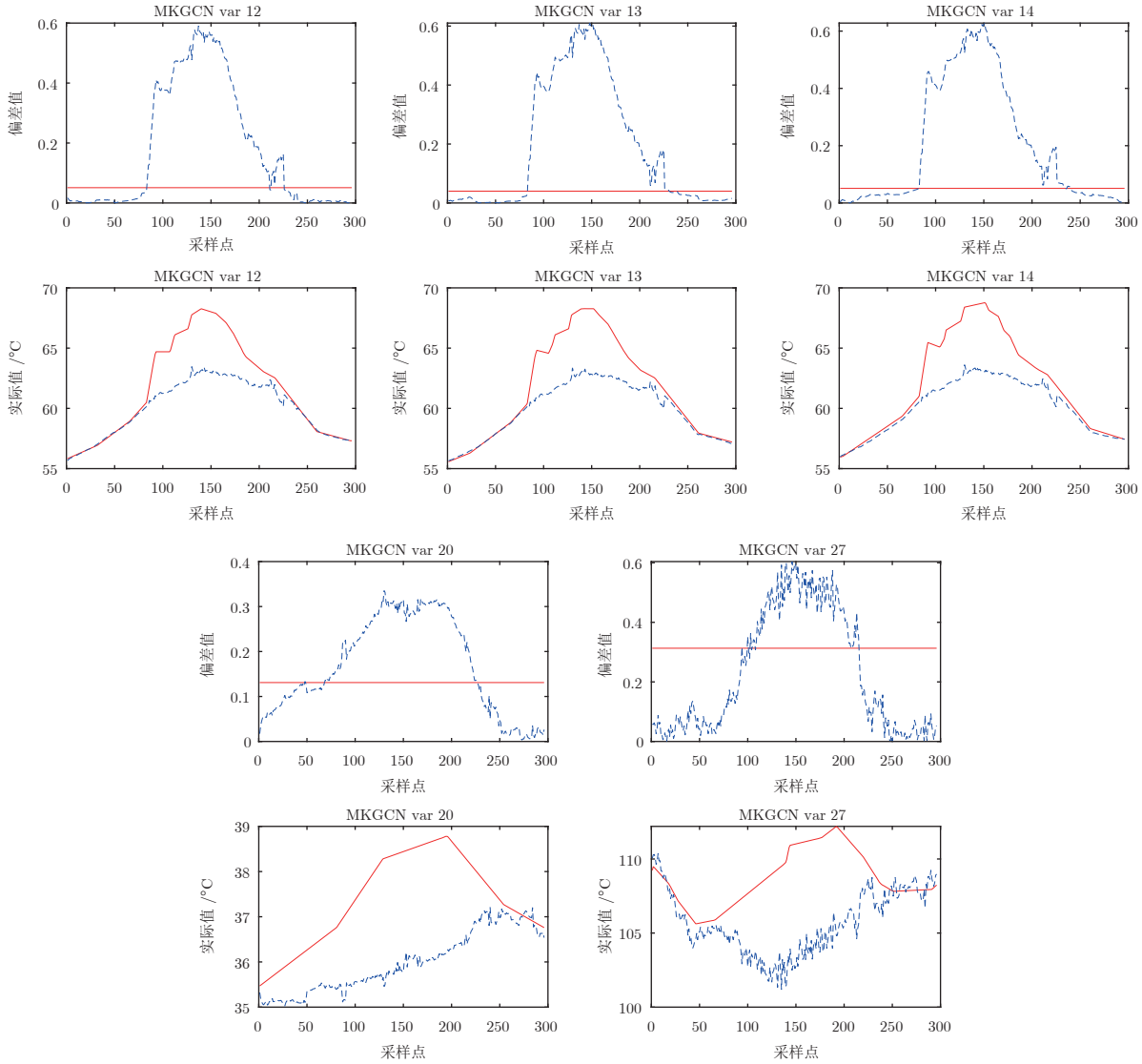


图 7 基于 MKGCN 的模型监测效果图 (var $\in F$)

Fig. 7 Monitoring diagram of model based on MKGCN (var $\in F$)

模式, 故相对于重构异常数据, 重构正常数据时误差更小. 本文提出的模型与自编码器结构相比, 侧重点有所不同. 本文所提出的模型的目标除了对于所有测点的运行状态进行故障监测外, 还需要重建出正常的运行状态, 从而为生产人员提供参考; 而基于自编码器的异常检测模型大多侧重于评估运行状况整体是否出现异常, 而非精准定位每一个测点是否异常.

对于自编码器结构, 与其他对比方法不同, 在对每个测点进行估计时, 其自身当前及历史信息也被用于估计当前测点, 即模型除了捕获测点间的相关性还考虑了测点自身的影响. 为了与本文涉及到的其他方法保持一致性, 模型输入 X 为传感器提供的各个测点的数据, X 的大小为$[n, len]$, 模型输出全部 n 个测点当前 t 时刻的估计值, 即使用当前时刻信息及历史信息重构当前时刻信息. 网络结构如表 10 所示. 同样采用了网格搜索寻找最优超参数, 考虑到测点个数 $n = 33$, 为了保持 AE 中的信息瓶颈特性, 超参数取值范围为 $ld = \{4, 8, 16\}$. 通过网格搜索得到最优超参数 $ld = 16$.

由实验结果图 9 和表 11 可见, 基于自编码器的估计模型对于正常数据能够实现相当准确的重建, 这是十分自然的. 因为对于每一个测点, 当前时刻及历史时刻的全量测点的实测值被用于估计当前时刻的实测值, 直接利用了输入数据中包含的目标输出信息, 使得 AE 模型在数据正常的情况下准确地重建数据相对容易. 然而, 与其他对比方法类似, 当运行状态出现异常时, 尽管 AE 模型能够识别出

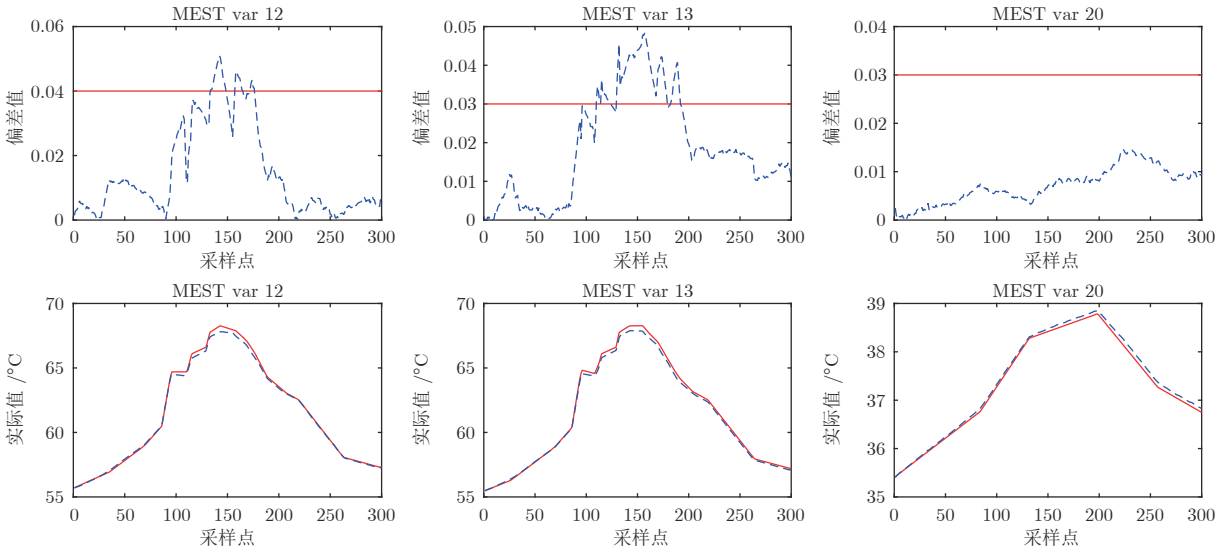


图 8　MEST 方法漏报的部分异常变量

Fig. 8　Some abnormal variables partially missed by MEST method

表 10　基于 AE 的工况估计模型的结构

Table 10　Structure of working condition estimation model based on AE

序号	网络层	数目	参数	激活函数
1	BiLSTM	1	[input_size = len, hidden_size = $2 \times ld$]	None
2	FC 0	1	[input_size = $4 \times ld$, output_size = $2 \times ld$]	Tanh
3	FC 1	1	[input_size = $2 \times ld$, output_size = ld]	Tanh
4	FC 2	1	[input_size = ld, output_size = $2 \times ld$]	Tanh
5	FC 3	1	[input_size = $2 \times ld$, output_size = n]	None

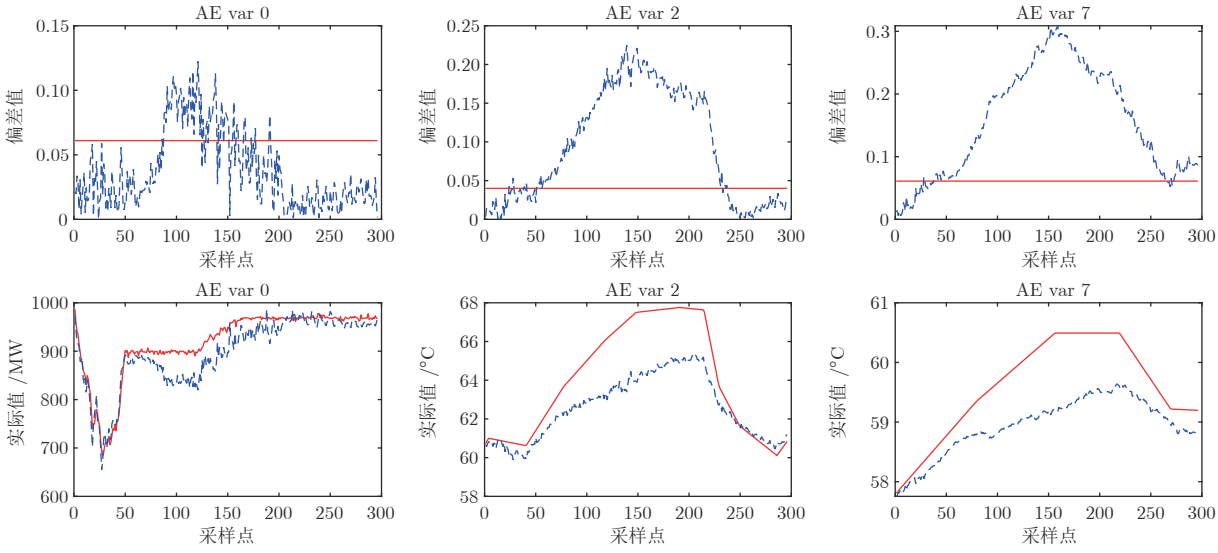


图 9　AE 模型误报的部分正常变量

Fig. 9　Some normal variables with serious false alarm by AE model

当前过程的整体状态出现异常, 表现为众多测点的估计值和实测值之间出现了明显偏差, 产生报警. 但由于测点间估计的相互干扰, AE 模型不能准确地区分异常测点和正常测点, 对于正常测点出现明显的误报警现象, 不能准确地对全量测点进行估计, 部分正常测点被识别为异常.

表 11 AE 与 MKGCN 实验结果对比
(MKGCN 实验结果同表 6 ~ 9)

Table 11 Comparison of experimental results between AE and MKGCN (The experimental results of MKGCN are the same as Tables 6 ~ 9)

指标	AE	MKGCN
RMSE, var $\in N$	0.020	0.044
RMSE, var $\in F$	0.022	0.046
MAE, var $\in N$	0.016	0.036
MAE, var $\in F$	0.019	0.039
$False_{\rm p}$, var $\in N$	38.811	4.500
$False_{\rm n}$, var $\in N$	0	0
F1, var $\in N$	75.922	97.698
$False_{\rm p}$, var $\in F$	35.009	10.769
$False_{\rm n}$, var $\in F$	0.887	1.056
F1, var $\in F$	78.505	93.836

4.4 单通道与多通道的性能对比

为了反映出 MKGCN 层中多通道特性对模型的影响, 令输出通道数 $oc = 1$, 对超参数 ld 在 [8, 16, 32, 64, 128] 范围内重新搜索, 得到 $oc = 1$ 时最优超参数 $ld = 128$, 此时模型参数量为 2.2×10^7. 与多输出通道情况下的性能指标对比如表 12 所示. 在参数显著增长的情况下估计精度并无有效提升, 说明多输出通道特性的引入有助于节约模型参数, 更高效地捕获测点间的关系. 在异常工况下, 多输出通道的监测性能显著优于单输出通道. 说明单输出通道的情况下模型倾向于通过提高特征长度 ld 来提升估计精度, 但是过长的特征导致模型对正常工况产生一定程度的过拟合, 从而在异常工况发生时, 模型的异常检测能力显著下降. 而多输出通道特性的引入能够对测点间的关系更多样性、更鲁棒地建模.

在多核图卷积层中, 采用多组邻接核, 实现了多通道的特征提取. 输出通道 1, 2 及输入通道 1, 2 上的邻接核可视化结果如图 10 所示. 为了便于观察, 对于邻接核矩阵的元素进行了 min-max 标准化. 对比图 6 可以看出, 多通道特性的引入从不同的角度对测点间的关系进行了建模, 且同一输入通道下不同输出通道的邻接核相似但不完全相同.

表 12 单输出通道与多输出通道性能对比

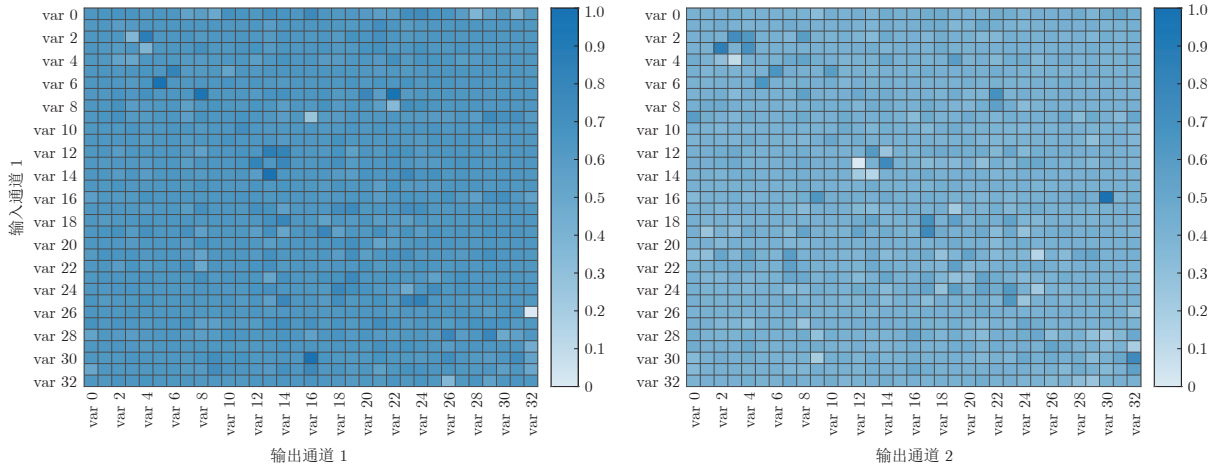
Table 12 Performance comparison between single output channel and multiple output channels

指标	$oc = 1$	$oc = 32$
RMSE, var $\in N$	0.043	0.044
RMSE, var $\in F$	0.055	0.046
MAE, var $\in N$	0.035	0.036
MAE, var $\in F$	0.049	0.039
$False_{\rm p}$, var $\in N$	38.486	4.500
$False_{\rm n}$, var $\in N$	0	0
F1, var $\in N$	76.172	97.698
$False_{\rm p}$, var $\in F$	36.022	10.769
$False_{\rm n}$, var $\in F$	5.237	1.056
F1, var $\in F$	76.385	93.836

此外, 本文展示了不同输入通道数量对于实验结果的影响, 如表 13 所示. 其中 $c_{\rm in}^1$ 表示仅使用 LSTM 输入通道, $c_{\rm in}^2$ 表示仅使用 FC 输入通道, $c_{\rm in}^{1,2}$ 表示同时采用两个输入通道. 可以看出, 同时采用多个不同方式构建的输入通道起到了类似于集成学习的作用, 不同角度提取的特征作为 MKGCN 不同通道的输入有助于更多样化、更充分地对系统进行建模, 这一点在图 10 不同通道邻接核的可视化结果中也得到了体现.

4.5 基于特征逼近的自迭代方法的作用

表 14 中展示了不进行自迭代 ($it = 0$)、进行少量自迭代 ($it = 5$)、进行充分自迭代 ($it = 50$) 的实



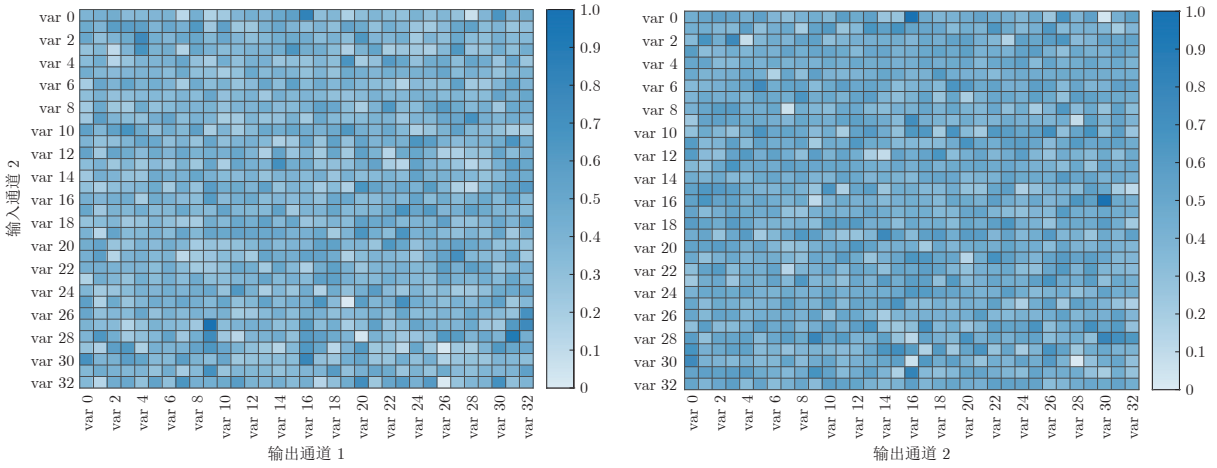


图 10　不同通道的邻接核可视化结果

Fig. 10　Visualization results of adjacency kernels in different channels

表 13　单输入通道与多输入通道性能对比

Table 13　Performance comparison between single input channel and multiple input channels

指标	c_{in}^{1}	c_{in}^{2}	$c_{\text{in}}^{1,2}$
RMSE, var $\in N$	0.084	0.046	0.044
RMSE, var $\in F$	0.044	0.044	0.046
MAE, var $\in N$	0.072	0.038	0.036
MAE, var $\in F$	0.037	0.038	0.039
$False_{\text{p}}$, var $\in N$	22.703	5.527	4.500
$False_{\text{n}}$, var $\in N$	0	0	0
F1, var $\in N$	87.195	97.158	97.698
$False_{\text{p}}$, var $\in F$	33.405	15.372	10.769
$False_{\text{n}}$, var $\in F$	16.765	10.093	1.056
F1, var $\in F$	73.991	87.187	93.836

验结果, 其中 it 为自迭代次数.

在第 3.2 节中, 讨论了自迭代过程中正常和异常特征的变化情况. 当迭代次数充分时, 能够基本消除异常特征对于其他测点估计的负面影响. 图 11 展示了一个异常测点 (测点 12) 和一个正常测点 (测点 25) 在迭代过程中的变化. 自迭代方法通过不断地将相对正常特征替换掉异常特征, 能够有效地改善测点的估计, 从而减少误报警与漏报警.

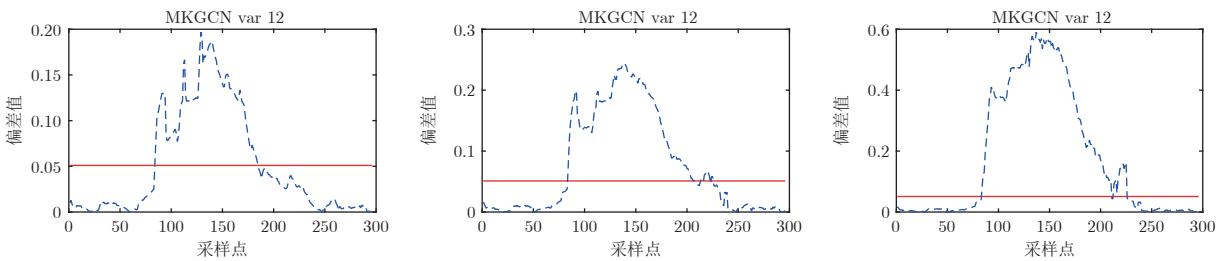
表 14　自迭代效果对比

Table 14　Comparison of self-iteration effect

指标	$it = 0$	$it = 5$	$it = 50$
$False_{\text{p}}$, var $\in N$	11.662	7.527	4.500
$False_{\text{n}}$, var $\in N$	0	0	0
F1, var $\in N$	93.808	96.089	97.698
$False_{\text{p}}$, var $\in F$	11.740	12.289	10.769
$False_{\text{n}}$, var $\in F$	1.732	0.676	1.055
F1, var $\in F$	93.000	93.157	93.837

5　结束语

本文提出一种面向全量测点耦合结构分析的多核图卷积估计模型, 并引入自迭代方法. 基于此, 设计了过程估计与监测方法. 与经典的多变量状态估计方法相比, 本文提出的测点估计与监测方法能够更充分地对测点间的耦合关系进行建模, 提升监测性能, 与一般的监测方法相比更加直观、全面. 通过将众多测点视为一张工业生产全量测点图中的节点, 本文提出了多核图卷积层对测点间的耦合结构进行有效的建模与学习, 基于估计偏差建立了监测指标, 从而实现全量测点的同步估计与监测. 并通过引入基于特征逼近的自迭代方法这一纠错机制,



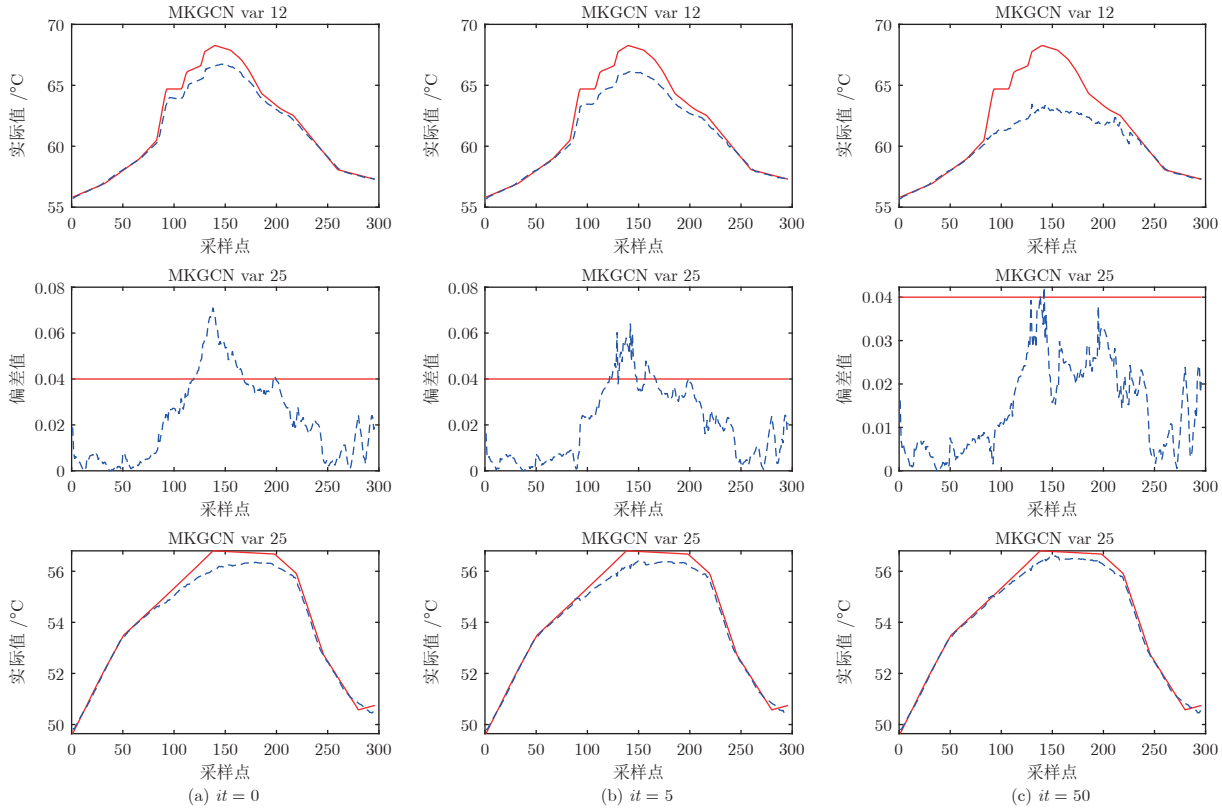


图 11 测点 12 和测点 25 的工况估计值对比

Fig. 11 Comparison of working condition estimated values of measuring point 12 and measuring point 25

有效地消除了异常特征对于正常测点估计的干扰. 通过某燃煤电厂百万千瓦超超临界机组中重要辅机 (引风机) 的实际生产数据, 验证了所提方法在真实场景下的有效性. 所提方法在模型参数较少的情况下显著降低了系统异常情况下对于正常测点的误报情况, 能够对各个测点实现精准估计和有效监测, 对故障发生时快速精准定位故障原因具有重要的现实意义.

References

1 Chai Tian-You. Development directions of industrial artificial intelligence. *Acta Automatica Sinica*, 2020, **46**(10): 2003−2012
(柴天佑. 工业人工智能发展方向. 自动化学报, 2020, **46**(10): 2003−2012)

2 Ma Liang, Peng Kai-Xiang, Dong Jie. Review of root cause diagnosis and propagation path identification techniques for faults in industrial processes. *Acta Automatica Sinica*, 2022, **48**(7): 1650−1663
(马亮, 彭开香, 董洁. 工业过程故障根源诊断与传播路径识别技术综述. 自动化学报, 2022, **48**(7): 1650−1663)

3 Zhao C H. Perspectives on nonstationary process monitoring in the era of industrial artificial intelligence. *Journal of Process Control*, 2022, **116**: 255−272

4 He Y L, Geng Z Q, Zhu Q X. Soft sensor development for the key variables of complex chemical processes using a novel robust bagging nonlinear model integrating improved extreme learning machine with partial least square. *Chemometrics and Intelligent Laboratory Systems*, 2016, **151**: 78−88

5 Zhao Chun-Hui, Hu Yun-Yun, Zheng Jia-Le, Chen Jun-Hao. Data-driven operating monitoring for coal-fired power generation equipment: The state of the art and challenge. *Acta Automatica Sinica*, 2022, **48**(11): 2611−2633
(赵春晖, 胡赟昀, 郑嘉乐, 陈军豪. 数据驱动的燃煤发电装备运行工况监控——现状与展望. 自动化学报, 2022, **48**(11): 2611−2633)

6 Sun X, Marquez H J, Chen T W, Riaz M. An improved PCA method with application to boiler leak detection. *ISA Transactions*, 2005, **44**(3): 379−397

7 You L X, Chen J. A variable relevant multi-local PCA modeling scheme to monitor a nonlinear chemical process. *Chemical Engineering Science*, 2021, **246**: Article No. 116851

8 Zhao C H, Sun H. Dynamic distributed monitoring strategy for large-scale nonstationary processes subject to frequently varying conditions under closed-loop control. *IEEE Transactions on Industrial Electronics*, 2019, **66**(6): 4749−4758

9 Song P Y, Zhao C H. Slow down to go better: A survey on slow feature analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, **35**(3): 3416−3436

10 Zhao C H, Chen J H, Jing H. Condition-driven data analytics and monitoring for wide-range nonstationary and transient continuous processes. *IEEE Transactions on Automation Science and Engineering*, 2021, **18**(4): 1563−1574

11 Fan Ji-Cong, Wang You-Qing, Qin S. Joe. Combined indices for ICA and their applications to multivariate process fault diagnosis. *Acta Automatica Sinica*, 2013, **39**(5): 494−501
(樊继聪, 王友清, 秦泗钊. 联合指标独立成分分析在多变量过程故障诊断中的应用. 自动化学报, 2013, **39**(5): 494−501)

12 Ma L Y, Ma Y G, Lee K Y. An intelligent power plant fault diagnostics for varying degree of severity and loading conditions. *IEEE Transactions on Energy Conversion*, 2010, **25**(2): 546−554

13 Zhao R, Yan R Q, Wang J J, Mao K Z. Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors*, 2017, **17**(2): Article No. 273

14 Shen Y, Abubakar M, Liu H, Hussain F. Power quality disturbance monitoring and classification based on improved PCA and convolution neural network for wind-grid distribution systems. *Energies*, 2019, **12**(7): Article No. 1280

15 Yu J, Rashid M M. A novel dynamic Bayesian network-based networked process monitoring approach for fault detection, propagation identification, and root cause diagnosis. *AIChE Journal*, 2013, **59**(7): 2348−2365

16 Dimokranitou A. Adversarial Autoencoders for Anomalous Event Detection in Images [Master thesis], Purdue University, USA, 2017.

17 De Castro-Cros M, Rosso S, Bahilo E, Velasco M, Angulo C. Condition assessment of industrial gas turbine compressor using a drift soft sensor based in autoencoder. *Sensors*, 2021, **21**(8): Article No. 2708

18 Lutz M A, Vogt S, Berkhout V, Faulstich S, Dienst S, Steinmetz U, et al. Evaluation of anomaly detection of an autoencoder based on maintenace information and scada-data. *Energies*, 2020, **13**(5): Article No. 1063

19 Guo Y F, Liao W X, Wang Q L, Yu L X, Ji T X, Li P. Multidimensional time series anomaly detection: A GRU-based Gaussian mixture variational autoencoder approach. In: Proceedings of the 10th Asian Conference on Machine Learning. Cambridge MA, USA: JMLR, 2018. 97−112

20 Yu W K, Zhao C H. Robust monitoring and fault isolation of nonlinear industrial processes using denoising autoencoder and elastic net. *IEEE Transactions on Control Systems Technology*, 2020, **28**(3): 1083−1091

21 Hu Y Y, Wang Y, Zhao C H. A sparse fault degradation oriented fisher discriminant analysis (FDFDA) algorithm for faulty variable isolation and its industrial application. *Control Engineering Practice*, 2019, **90**: 311−320

22 Zhao Chun-Hui, Yu Wan-Ke, Gao Fu-Rong. Data analytics and condition monitoring methods for nonstationary batch processes——Current status and future. *Acta Automatica Sinica*, 2020, **46**(10): 2072−2091
(赵春晖, 余万科, 高福荣. 非平稳间歇过程数据解析与状态监控——回顾与展望. 自动化学报, 2020, **46**(10): 2072−2091)

23 Gross K C, Singer R M, Wegerich S W, Herzog J P, VanAlstine R, Bockhorst F. Application of a model-based fault detection system to nuclear plant signals. In: Proceedings of the 9th International Conference on Intelligent Systems Applications to Power Systems. Seoul, Korea: Argonne National Lab., 1997.

24 Zavaljevski N, Gross K C. Sensor fault detection in nuclear power plants using multivariate state estimation technique and support vector machines. In: Proceedings of the 3rd International Conference of the Yugoslav Nuclear Society. Belgrade, Yugoslavia: Argonne National Lab., 2020.

25 Cheng S F, Pecht M. Multivariate state estimation technique for remaining useful life prediction of electronic products. In: Proceedings of the 2007 AAAI Fall Symposium on Artificial Intelligence for Prognostics. Arlington, USA: AAAI, 2007.

26 Wang Z Q, Liu C L. Wind turbine condition monitoring based on a novel multivariate state estimation technique. *Measurement*, 2021, **168**: Article No. 108388

27 Bockhorst F K, Gross K C, Herzog J P, Wegerich S W. MSET modeling of crystal river-3 venturi flow meters. In: Proceedings of the 6th International Conference on Nuclear Engineering. San Diego, USA: Argonne National Lab., 1998.

28 Fan Y J, Tao B, Zheng Y, Jang S S. A data-driven soft sensor based on multilayer perceptron neural network with a double LASSO approach. *IEEE Transactions on Instrumentation and Measurement*, 2020, **69**(7): 3972−3979

29 Zhang M, Liu X G, Zhang Z Y. A soft sensor for industrial melt index prediction based on evolutionary extreme learning machine. *Chinese Journal of Chemical Engineering*, 2016, **24**(8): 1013−1019

30 Ke W S, Huang D X, Yang F, Jiang Y H. Soft sensor development and applications based on LSTM in deep neural networks. In: Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI). Honolulu, USA: IEEE, 2017. 1−6

31 Yuan X F, Qi S B, Wang Y L, Xia H B. A dynamic CNN for nonlinear dynamic feature learning in soft sensor modeling of industrial process data. *Control Engineering Practice*, 2020, **104**: Article No. 104614

32 Zhu W B, Ma Y, Zhou Y Z, Benton M, Romagnoli J. Deep learning based soft sensor and its application on a pyrolysis reactor for compositions predictions of gas phase components. *Computer Aided Chemical Engineering*, 2018, **44**: 2245−2250

33 Chang Shu-Chao, Zhao Chun-Hui. A spatio-temporal synergistic graph convolution long short-term memory network and its application for industrial soft sensors. *Control and Decision*, 2022, **37**(1): 77−86
(常树超, 赵春晖. 一种时空协同的图卷积长短期记忆网络及其工业软测量应用. 控制与决策, 2022, **37**(1): 77−86)

34 Kipf T N, Welling M. Semi-supervised classification with graph-convolutional networks. In: Proceedings of the 5th International Conference on Learning Representations. Toulon, France: arXiv.org, 2017.

35 Feng L J, Zhao C H, Li Y L, Zhou M, Qiao H L, Fu C. Multichannel diffusion graph convolutional network for the prediction of endpoint composition in the converter steelmaking process. *IEEE Transactions on Instrumentation and Measurement*, 2021, **70**: 1−13

36 Wu Z H, Pan S R, Long G D, Jiang J, Chang X J, Zhang C Q. Connecting the dots: Multivariate time series forecasting with graph neural networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: Association for Computing Machinery, 2020. 753−763

37 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, **9**(8): 1735−1780

38 Gers F A, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. *Neural Computation*, 2000, **12**(10): 2451−2471

39 Feng L J, Zhao C H, Sun Y X. Dual attention-based encoder-decoder: A customized sequence-to-sequence learning for soft sensor development. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, **32**(8): 3306−3317

40 Feng L J, Zhao C H, Huang B. Adversarial smoothing tri-regression for robust semi-supervised industrial soft sensor. *Journal of Process Control*, 2021, **108**: 86−97

41 Schuster M, Paliwal K K. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 1997, **45**(11): 2673−2681

42 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, **60**(6): 84−90

43 Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: ACM, 2017. 6000−6010

44 Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. Sardinia, Italy: PMLR, 2010. 249−256

45 Li Q M, Han Z C, Wu X M. Deeper insights into graph convolutional networks for semi-supervised learning. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence and 30th Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. New Orleans, USA: AAAI, 2018. 3538−3545

46 Chiang W L, Liu X Q, Si S, Li Y, Bengio S, Hsieh C J. Cluster-GCN: An efficient algorithm for training deep and large graph convolutional networks. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Anchorage, USA: Association for Computing Machinery, 2019. 257−266

47 Terrell G R, Scott D W. Variable kernel density estimation. *The Annals of Statistics*, 1992, **20**(3): 1236−1265

48 Gilbertson D D, Kent M, Pyatt F B. Data analysis and interpretation III: Correlation and regression using spearman′s rank correlation coefficient and semi-averages regression. *Practical Ecology for Geography and Biology*. New York, USA: Springer, 1985. 218−236

49 Geladi P, Kowalski B R. Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, 1986, **185**: 1−17

50 Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications. *Neurocomputing*, 2006, **70**(1−3): 489−501

51 Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman D J. 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 2021, **151**: Article No. 107398

赵健程 浙江大学控制科学与工程学院博士研究生. 2021 年获得浙江大学学士学位. 主要研究方向为大数据分析, 深度学习和零样本学习.
E-mail: zhaojiancheng@zju.edu.cn
(**ZHAO Jian-Cheng** Ph.D. candidate at the College of Control Science and Engineering, Zhejiang University. He received his bachelor degree from Zhejiang University in 2021. His research interest covers big data analysis, deep learning, and zero-shot learning.)

赵春晖 浙江大学控制科学与工程学院教授. 2003 年, 2006 年, 2009 年分别获得东北大学学士、硕士和博士学位. 主要研究方向为机器学习, 工业大数据解析与应用, 包括化工, 能源以及医疗领域. 本文通信作者.
E-mail: chhzhao@zju.edu.cn
(**ZHAO Chun-Hui** Professor at the College of Control Science and Engineering, Zhejiang University. She received her bachelor, master, and Ph.D. degrees from Northeastern University, in 2003, 2006, and 2009, respectively. Her research interest covers machine learning, analytics of industrial big data, and their applications in chemical, energy, and medical fields. Corresponding author of this paper.)