

DOI: 10.19666/j.rlfd.202306101

基于数字孪生技术的热电联产 机组建模研究

王印松¹, 姜灵斌¹, 王莺歌²

(1.华北电力大学自动化系, 河北 保定 071003; 2.华能营口热电有限责任公司, 辽宁 营口 115000)

[摘要] 为建立精确有效的热电联产机组动态模型, 利用机组运行数据, 提出一种基于数字孪生技术的热电联产机组建模方法。首先提取机组数据服务器内存储的历史数据, 采用改进遗传模拟退火的模糊 C 均值方法对其进行聚类, 建立历史数据聚类库; 然后在机组运行期间, 采集并传输实时运行数据, 利用多级相似度识别策略, 在历史数据聚类库内检索最接近实时运行数据的历史数据; 接着基于优化的极限学习机将搜寻到的历史数据用于机组建模; 最后以杭州某热电联产机组为实验对象, 建立该机组的孪生模型并进行对比实验。结果表明: 所建模型满足精确性要求, 能够跟踪机组实时状态响应; 并且可以通过灵活改变建模过程中参数的设定来进一步优化模型精度。

[关键词] 热电联产机组; 数字孪生; 遗传模拟退火算法; 相似度识别; 极限学习机

[引用本文格式] 王印松, 姜灵斌, 王莺歌. 基于数字孪生技术的热电联产机组建模研究[J]. 热力发电, 2023, 52(12): 106-114.
WANG Yinsong, JIANG Lingbin, WANG Yingge. Research on modeling of cogeneration units based on digital twin technology[J]. Thermal Power Generation, 2023, 52(12): 106-114.

Research on modeling of cogeneration units based on digital twin technology

WANG Yinsong¹, JIANG Lingbin¹, WANG Yingge²

(1.Department of Automation, North China Electric Power University, Baoding 071003, China;
2.Huaneng Yingkou Thermal Power Co., Ltd., Yingkou 115000, China)

Abstract: To establish an accurate and effective dynamic model of cogeneration units, a modeling method based on digital twin technology is proposed using unit operation data. Firstly, the historical data stored in the unit data server is extracted, it is then clustered using the improved genetic simulated annealing fuzzy C-means method to establish a historical data clustering library. Then, during the operation of the unit, real-time operational data is collected and transmitted, and a multi-level similarity recognition strategy is used to retrieve the historical data closest to real-time operational data in the historical data clustering library. Then, based on the optimization, the extreme learning machine will use the searched historical data for unit modeling. Finally, a twin model of a cogeneration unit in Hangzhou is established and comparative experiments are conducted. The results show that, the built model meets the accuracy requirements and can track the real-time state response of the unit. The model accuracy can be further optimized by flexibly changing the parameter settings during the modeling process.

Key words: cogeneration units; digital twin; genetic simulated annealing algorithm; similarity recognition; ELM

热电联产是发电厂能够同时利用高温蒸汽生产电能和对用户供热的生产方式^[1-2]。目前, 热电联产机组建模基本原理比较成熟, 研究的较为完善。文献[3-7]均围绕机组内部发生的能量平衡、质量平衡建立了一系列机理模型, 然而由于整个建模过程

都是在离线状态下进行的, 模型并不能实时反映机组真实的运行状态。

目前数据存储技术在工业领域有着广泛应用。热电联产机组大量历史数据和实时数据存储在数据库内, 并且数据量仍然不断增多^[8]。庞大的历史

收稿日期: 2023-06-07 网络首发日期: 2023-09-15

第一作者简介: 王印松 (1967), 男, 博士, 教授, 主要研究方向为智能发电系统分析与优化, wys@ncepu.edu.cn。

通信作者简介: 姜灵斌 (1999), 男, 硕士研究生, 主要研究方向为热力系统建模与优化, 3159701452@qq.com。

数据是机组在各种工况下最真实的系统响应,为机组对象的建模提供了基础。

数字孪生技术是利用物理实体、实时采集数据以及相关历史数据等信息构建的物理实体与虚拟化模型之间实时精准的映射关系^[9]。文献[10-11]将数字孪生技术应用于航空航天领域,构建了航天器的孪生模型,解决了飞机全生命周期各个行为监测的问题;文献[12]建立了工程级别风力发电机组数字孪生系统,完成大规模并网风电智能化转型。文献[13-15]建立了多种工艺产品加工生产线孪生模型,实现了生产线物理空间与信息空间的实时交互;文献[16-18]聚焦于制造业,建立了机床的数字孪生模型,优化了机床的工作效率。

数字孪生技术发展的关键是:1)工业仿真软件平台功能的日益成熟,使得虚拟空间能够快速、准确地反映出物理系统的状态并指导系统的运行;2)物联网、大数据等通信信息技术保证了孪生系统实时感知物理实体的状态性能;3)云计算等计算机技术为处理海量数据信息提供了支持;4)神经网络、机器学习等智能算法为复杂系统建模与优化打下了基础^[19-20]。

国内外学者从数字孪生的理念、应用以及关键技术等多个层面展开了一系列研究,但在热电联产机组领域研究较少。

本文基于数字孪生技术,利用机组历史运行数据和数据挖掘理论,寻找最能反映当前运行工况规律数据集,通过数据驱动建模方法建立更具在线自适应能力热电联产机组孪生模型。

1 热电联产机组数字孪生模型

热电联产机组机理模型虽然能够较高还原生产过程中的本质规律,可靠性强,易于解释,但是这种离线模型并不能跟踪机组的实时状态响应。为此,本文提出一种热电联产机组数字孪生模型构建策略,让机组历史数据参与模型的构建中,利用实时运行数据及时修正更新模型运行状态,构建机组的孪生动态模型。图 1 为热电联产机组数字孪生模型构建流程。

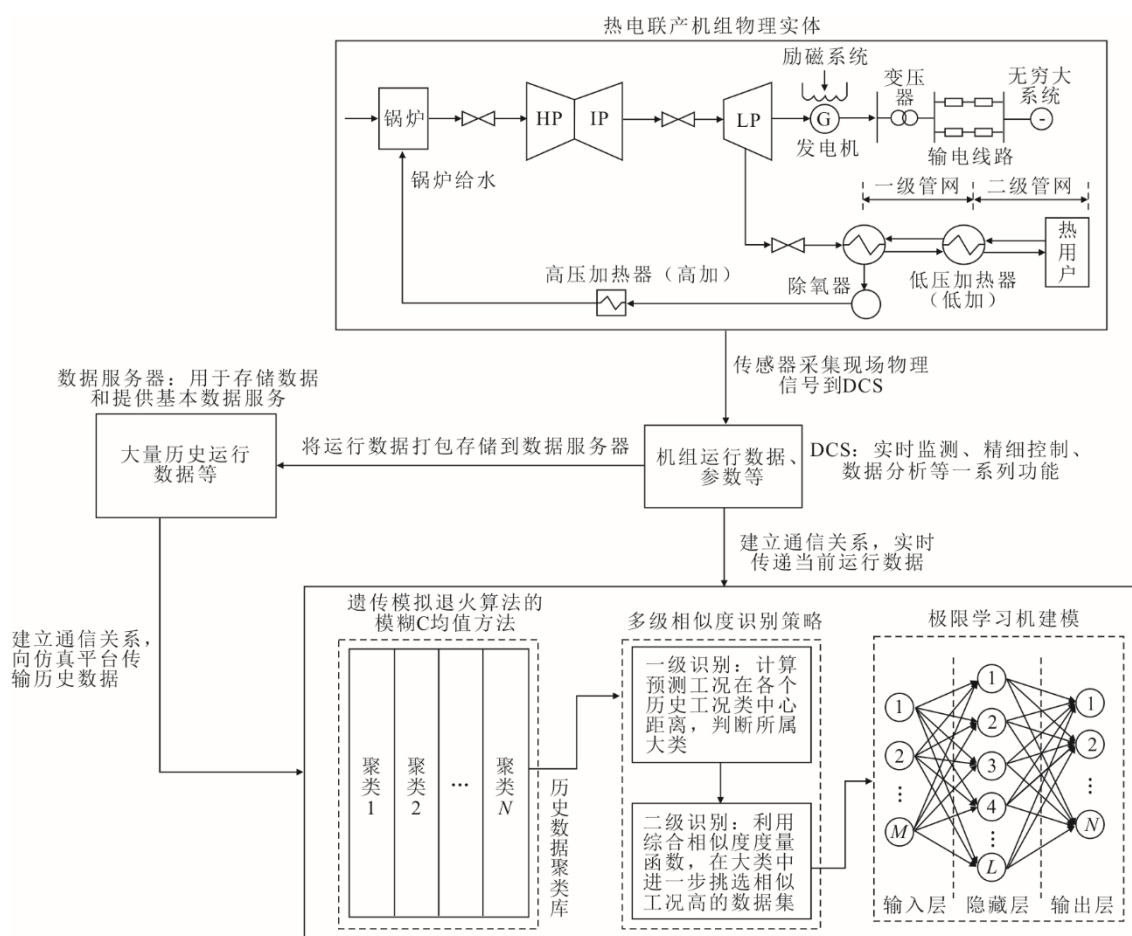


图 1 热电联产机组孪生模型构建流程

Fig.1 Construction process of twin models for cogeneration units

由图 1 可见, 热电联产机组孪生模型构建策略主要由 3 个部分组成: 1) 传感器采集机组现场物理信号到分散控制系统 (DCS), DCS 会将机组日常运行数据传输到数据服务器进行存储; 2) 建立数据服务器与仿真平台的通信关系并传输历史数据, 基于遗传模拟退火的模糊 C 均值方法进行聚类, 建立历史数据聚类库; 3) 建立 DCS 与仿真平台的通信关系, 利用多级相似度识别策略在历史数据聚类库中检索与实时运行数据最接近的历史数据, 并基于极限学习机建立当前运行状态下的机组模型。

随着热电联产机组的运行, 实时数据也在不断变化, 一方面越来越多的运行数据存储到数据服务器, 历史数据聚类库得到补充, 另一方面实时运行数据会不断通过 DCS 传递到仿真平台并寻找最为相似的历史数据集建立机组的孪生模型。所建孪生模型能实时反映当前机组的运行状态, 随着机组的不断运行, 孪生模型也随之不断修正, 具有很好的在线自适应能力。

2 改进遗传模拟退火的模糊 C 均值聚类算法

面对复杂多变工况的热电联产热力系统, 由于其多参数、强耦合的特点, 使得难以将含有相似工况信息的历史数据聚类。因此, 在建模之前采用一种改进遗传模拟退火的模糊 C 均值方法, 将机组数据服务器内相似数据聚为一类, 建立可靠的历史数据聚类库。

2.1 模糊 C 均值聚类算法

模糊 C 均值聚类 (FCM) 算法是基于“类内加权误差平方和最小化”准则^[21], 使用欧几里得空间确定数据点的几何贴适度概念, 目的是让非常相似的对象聚为一类^[22]。

给定数据集 $\mathbf{X}=\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, 其中 $\mathbf{x}_j \in \mathbb{R}^p$ 是第 j 个样本, 存在 P 个特征。其 FCM 算法目标函数为:

$$\min J(U, V) = \sum_{i=1}^c \sum_{j=1}^n r_{ij}^m \|\mathbf{x}_j - \mathbf{v}_i\|^2 \quad (1)$$

式中: c 为聚类个数, 要求其尽可能接近实际机组的运行工况数量, 过大会导致同一工况数据被划分到不同聚类当中, 造成相似工况信息不能有效利用, 过小则会导致不同工况数据聚为一类, 给相似度识别过程带来偏差以及增大计算量; m 为模糊权重系数; r_{ij} 为第 j 个样本 $\mathbf{x}_j$ 对第 i 类的隶属度, 并且有

$\sum_{i=1}^c r_{ij} = 1$, 隶属度矩阵 $\mathbf{R}=[r_{ij}] \in \mathbb{R}^{c \times n}$; $\mathbf{v}_i$ 为第 i 个聚类中心。

根据 Lagrange 乘子算法, 可以求得隶属度 r_{ij} 和聚类中心 $\mathbf{v}_i$ 的迭代公式:

$$r_{ij}^{(t+1)} = \left[\sum_{k=1}^c \left(\frac{\|\mathbf{x}_j - \mathbf{v}_i^{(t+1)}\|}{\|\mathbf{x}_j - \mathbf{v}_k^{(t+1)}\|} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (2)$$

$$\mathbf{v}_i^{(t+1)} = \frac{\sum_{j=1}^n (r_{ij}^{(t)})^m \mathbf{x}_j}{\sum_{j=1}^n (r_{ij}^{(t)})^m} \quad (3)$$

设定阈值 ε , 给定初始隶属度矩阵 $\mathbf{R}^{(0)}$, 利用式(2)和式(3)交替更新隶属度矩阵和聚类中心, 直至 $\|\mathbf{v}^{(t+1)} - \mathbf{v}^{(t)}\| \geq \varepsilon$ 。

2.2 改进的遗传模拟退火算法流程

FCM 算法初始聚类中心由随机生成的隶属矩阵函数确定, 若初始值的设置不合适, 容易收敛到局部最小值。为了克服该缺点, 将遗传算法用于 FCM 的优化计算。首先由遗传算法得到最优初始聚类中心; 然而快速收敛到后期时, 遗传种群内染色体都非常接近, 通过交叉操作都难以产生新的染色体, 容易陷入局部最优, 从 2 方面对此进行改进:

1) 模拟退火算法作为一种全局最优算法, 总能够产生一定的概率 P 接受非局部最优解, 从而保证算法不陷入局部最优中^[23]。

将其混合到遗传算法中, 采用一种改进的自适应接受概率来调整遗传种群, 有利于对 FCM 算法初始聚类中心的寻优。定义温度的参考值 T_0 为:

$$T_0 = \frac{f_{\text{average}} - f_{\text{max}}}{\ln \alpha} \quad (4)$$

式中: f_{average} 为初始种群平均适应度函数值; f_{max} 为初始种群最优适应度函数值; α 为 0~e 的常数。

则接受概率 P 为:

$$P = \begin{cases} 1 & , f_m^* < f_m \\ \exp\left(-\frac{\alpha_1(f_m^* - f_m)}{TT_0}\right) & , f_m^* > f_m \\ \exp\left(-\frac{\alpha_2}{TT_0}\right) & , f_m^* = f_m \end{cases} \quad (5)$$

式中: f_m 为种群选择交叉变异前种群中第 m 个个体适应度值; f_m^* 为选择交叉变异之后新个体的适应度

值; T 为当前温度; α_1 和 α_2 分别为 2 个概率常系数。

前期温度 T 较高, $f_m^* \geq f_m$ 的情况下, 接受概率也较大, 有利于跳出局部最优; 退火到后期, 温度较小, 接受概率较小, 种群更倾向于往最优靠近。

2) 传统遗传算法中交叉、变异概率为定值, 对此进行改进, 避免适应度较大的个体大量繁殖, 算法陷入局部最优, 维持种群多样性。

交叉、变异自适应概率:

$$P_c = \begin{cases} P_{cmax}, & f^* < \bar{f} \\ P_{cmax} - \frac{(P_{cmax} - P_{cmin})(f^* - \bar{f})}{f_{max} - \bar{f}}, & f^* \geq \bar{f} \end{cases} \quad (6)$$

$$P_m = \begin{cases} P_{mmax}, & f < \bar{f} \\ P_{mmax} - \frac{(P_{mmax} - P_{mmin})(f - \bar{f})}{f_{max} - \bar{f}}, & f \geq \bar{f} \end{cases} \quad (7)$$

式中: f^* 为交叉过程中较大个体适应度; $\bar{f}$ 为种群平均适应度; f_{max} 为种群中最大适应度; f 为个体适应度; P_c 、 P_{cmax} 、 P_{cmin} 分别为交叉概率、最大交叉概率、最小交叉概率; P_m 、 P_{mmax} 、 P_{mmin} 分别为变异概率、最大变异概率、最小变异概率。

可以看出, 在进化过程中, 当个体适应度高于平均适应度时, 较低的交叉变异概率使得该优良个体得到保护顺利进入下一代; 而当个体适应度低于平均适应度时, 较高的交叉、变异概率使得该个体朝着新的方向进化, 维持了种群的多样性。

至此可以得到基于改进遗传模拟退火的模糊 C 均值方法, 基本步骤如下:

- 1) 初始化 c 个聚类中心; 初始化控制参数有最大迭代次数 $g_{en,max}$ 、种群个体数目、交叉变异概率、退火初始温度 T_i , 终止温度 T_{end} , 温度冷却系数 k ;
- 2) 计算各个体的隶属度和适应度;
- 3) 设置循环次数 $g_{en}=0$;
- 4) 选择、交叉、变异获取新种群; 计算 c 个聚类中心、各样本隶属度和各个体适应度, 并用改进后的模拟退火算法接受概率 P 替换旧个体;
- 5) 若 $g_{en} < g_{en,max}$, 则 $g_{en}=g_{en}+1$, 转至操作 4); 否则判断 $T_i < T_{end}$, 若成立, 算法结束, 返回最优解, 否则 $T_{i+1}=kT_i$, 转至步骤 3)。

图 2 为基于改进遗传模拟退火的模糊 C 均值算法流程。

3 多级相似度识别策略的工况识别

在实际热电联产机组中, 各个属性对预测结果所做出的贡献值是不一样的, 仅仅依靠欧氏距离作

为相似度度量方法显然不够。并且在数据集维度过高的时候, 考虑到提高搜索和建模速度的需求, 需要对属性维度进行一定的简化, 突出贡献值大的属性。此外, 随着机组历史数据的逐渐增多, 在工况识别过程会出现搜索范围扩大、相似度计算量增多、识别效率降低等一系列问题。本节重点探究一种考虑属性权重的综合相似度度量函数作为相似工况度量指标, 用于历史数据聚类库内当前运行状态下相似工况数据的快速识别。

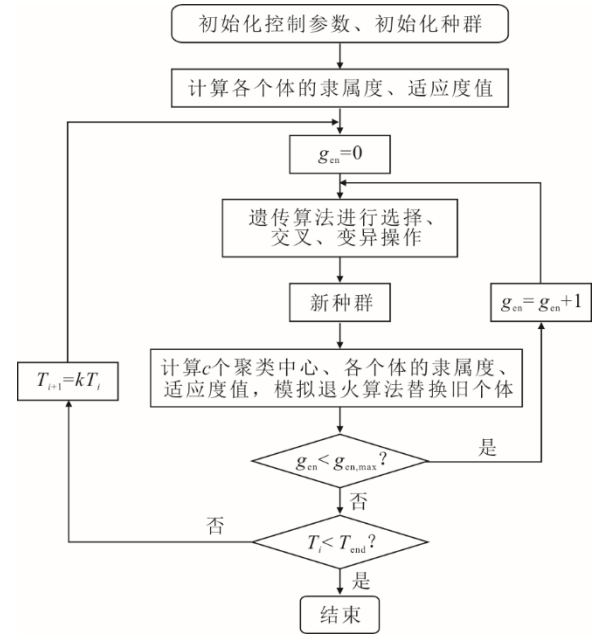


图 2 基于改进遗传模拟退火的模糊 C 均值算法流程
Fig.2 Fuzzy C-means algorithm flow based on improved genetic simulated annealing

多级相似度识别策略主要组成为: 1) 对比实时运行数据到各个聚类中心距离, 确定该运行数据所属的类别; 2) 在所确定的聚类中, 利用核主元分析方法 (KPCA) 计算核矩阵, 求出特征属性的贡献值矩阵, 将属性权重融入实时运行数据和历史数据当中^[24]。利用改进的相似度度量函数得到实时数据与历史数据的相似度, 挑选出相似度最高的 F 组数据作为相似工况数据集。这里 F 的取值过大会使相似程度不高, 过小会导致不能充分利用历史相似工况信息。

3.1 核主元分析方法

KPCA 算法中, 存在 N 组 m 维特征的历史数据, 利用数据集矩阵 $X_{m \times N}$ 求出该矩阵的核矩阵 C :

$$C = \phi(X_{m \times N})(\phi(X_{m \times N}))^T \quad (8)$$

式中核函数一般采用径向基函数 (高斯核函数)。

$$Cu = \lambda u \quad (9)$$

式中: λ 为核矩阵特征值; $\mathbf{u}$ 为相对应的特征向量。

取前 m 个较大的特征值 λ_d , $d=1, 2, \dots, m$, 从大到小排列, 则每个特征属性的权重系数可以用对应特征值的贡献率表示:

$$\mu_d = \frac{\lambda_d}{\sum_{i=1}^m \lambda_d} \quad (10)$$

式中: μ_d 为第 d 个特征的权重; λ_d 为核矩阵第 d 个特征值。

将属性权重融入实时运行数据和历史数据当中, 重新构建:

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_m)^T \quad (11)$$

$$\mathbf{X}_i^W = \mathbf{X}_i \text{diag}(\boldsymbol{\mu}), i=1, 2, \dots, N \quad (12)$$

$$\mathbf{X}_t^W = \mathbf{X}_t \text{diag}(\boldsymbol{\mu}) \quad (13)$$

式中: $\boldsymbol{\mu}$ 为特征的权重向量; $\text{diag}(\boldsymbol{\mu})$ 为 $\boldsymbol{\mu}$ 对角矩阵; $\mathbf{X}_i$ 和 $\mathbf{X}_t$ 分别为历史数据和运行数据; $\mathbf{X}_i^W$ 和 $\mathbf{X}_t^W$ 分别为融合属性权重后的历史数据和运行数据。

3.2 相似度度量函数

考虑到高维数据向量大小和方向价值尺度的差异, 这里采用综合考虑欧氏距离 $d(\mathbf{X}_i^W, \mathbf{X}_t^W)$ 和角度距离 $\theta(\mathbf{X}_i^W, \mathbf{X}_t^W)$ 的相似度度量函数:

$$S_w(\mathbf{X}_i^W, \mathbf{X}_t^W) = \gamma e^{-d(\mathbf{X}_i^W, \mathbf{X}_t^W)} + (1-\gamma) \cos[\theta(\mathbf{X}_i^W, \mathbf{X}_t^W)] \quad (14)$$

其中, 欧氏距离和角度距离定义为:

$$d(\mathbf{X}_i^W, \mathbf{X}_t^W) = \|\mathbf{X}_i^W - \mathbf{X}_t^W\| \quad (15)$$

$$\theta(\mathbf{X}_i^W, \mathbf{X}_t^W) = \arccos\left(\frac{\mathbf{X}_i^{WT} \mathbf{X}_t^W}{\|\mathbf{X}_i^W\| \|\mathbf{X}_t^W\|}\right) \quad (16)$$

式中: γ 为欧氏距离权重占比, 取值 0~1。需要使得同一工况数据集之间的相似度度量函数值尽可能大, 而不同工况数据集之间的相似度度量函数值尽可能小。对于特定的机组系统, 可以事先根据专家知识, 经过多次仿真实验得到合适的 γ [25]。

4 基于优化的极限学习机的辨识

在确定当前运行状态的相似工况数据集之后, 挑选合适的建模手段直接关系到最终机组动态模型的效果。极限学习机 (ELM) 非参数回归建模没有迭代过程, 输入层权重和隐藏层神经阈值随机给定, 并通过广义逆矩阵理论求解输出层权重 [26]。ELM 虽然没有传统神经网络的精度那么高, 但是计算过程大大简化, 缩短了建模时间, 在特征参数繁多的热电联产机组中应用效果更加明显。图 3 为 ELM 结构。

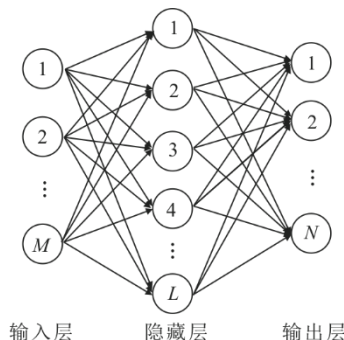


图 3 ELM 结构

Fig.3 ELM structure

图 3 中有 M 个输入, L 个隐含层神经元, N 个输出。给定 P 组训练样本 $(\mathbf{x}_j, \mathbf{t}_j)$, $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jM})$ 为 M 维输入样本数据, $\mathbf{t}_j = (t_{j1}, t_{j2}, \dots, t_{jN})$ 为 N 维输出样本数据, 激活函数为 $g(x)$, ELM 网络满足:

$$\sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = t_j \quad (17)$$

式中: $\mathbf{w}_i = (w_{i1}, w_{i2}, \dots, w_{iM})$ 为第 i 个隐含层神经元和输入节点之间的连接权值; $\beta_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{iN})$ 为第 i 个隐含层神经元与输出节点之间的连接权值; b_i 为第 i 个隐含层神经元阈值; $\mathbf{w}_i$ 和 b_i 随机生成。

写成矩阵形式:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \quad (18)$$

式中: $\mathbf{H}$ 为隐含层输出矩阵; $\boldsymbol{\beta}$ 为隐含层和输出层之间连接的权重矩阵; $\mathbf{T}$ 为输出矩阵。

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \dots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_p + b_1) & \dots & g(\mathbf{w}_L \cdot \mathbf{x}_p + b_L) \end{bmatrix}_{p \times L} \quad (19)$$

$$\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_p)^T_{p \times N} \quad (20)$$

利用矩阵运算, 基于最小二乘法准则求得极限学习机的输出权值:

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^+ \mathbf{T} \quad (21)$$

式中: $\mathbf{H}^+$ 为隐含层输出矩阵 $\mathbf{H}$ 的广义逆矩阵。

由于 ELM 的隐含层权值和阈值均随机给定, 可能出现训练过拟合等问题, 一定程度上影响着最终 ELM 预测性能 [27], 这里同样采取 2.2 节改进的遗传模拟退火算法对 ELM 初始权值和阈值寻优, 利用优化后的 ELM 建模辨识。

5 实例验证

5.1 模型参数设定

以杭州某热电联产机组系统为例, 该热工程 2×300 MW 的 1、2 号机组于 2011 年 6 月和 10 月先后投入生产, 汽轮机型号为 C280/N300-16.7/538/

538, 型式为亚临界、一次中间再热、单抽、两缸、两排汽、抽汽凝汽式汽轮机; 锅炉型号为 HG-1025/17.5-YM28, 型式为亚临界、自然循环、单炉膛、一次中间再热、露天布置、全钢构架、平衡通风、直流摆动燃烧器、固态排渣燃煤汽包炉; 发电机型号为 QFSN-300-2, 型式为三相、二极、隐极式转子同步发电机。

该机组数据服务器存储着包含燃料量、抽汽压力、LV 阀开度、给水流量等 21 个关键参数的历史数据。为了更加直观的对机组热力转换情况进行监测, 将热力媒介温度压力作为模型输出, 其余参数作为输入, 具体见表 1, 分别有 11 个输入、10 个输出。

表 1 输入、输出参数设定
Tab.1 Input and output parameter settings

输入	输出
燃料量	一抽至 1 号高加蒸汽压力
汽轮机主控指令	一抽至 1 号高加蒸汽温度
EV 阀开度	一抽至 2 号高加蒸汽压力
LV 阀开度	一抽至 2 号高加蒸汽温度
给水流量	一抽至 3 号高加蒸汽压力
除氧器进口凝结水流量	一抽至 3 号高加蒸汽温度
除氧器进口温度	5 号低加出水温度
抽汽压力	6 号低加出水温度
低压供热流量	7 号低加出水温度
除氧器温度	8 号低加出水温度
除氧器压力	

5.2 模型预测输出

基于提出的热电联产机组数字孪生模型构建策略, 建立数据服务器、DCS、仿真平台之间的通信。在机组运行时在仿真平台进行实验, 步骤如下:

1) 根据第 2 节提出的聚类算法, 将不同工况下的历史数据聚类划分, 建立历史数据聚类库。聚类数量的设定需要满足每一类内部的差别小, 各类之间差别大的要求, 使得同一工况下的数据分到一类, 这里拟定聚类数量为 8;

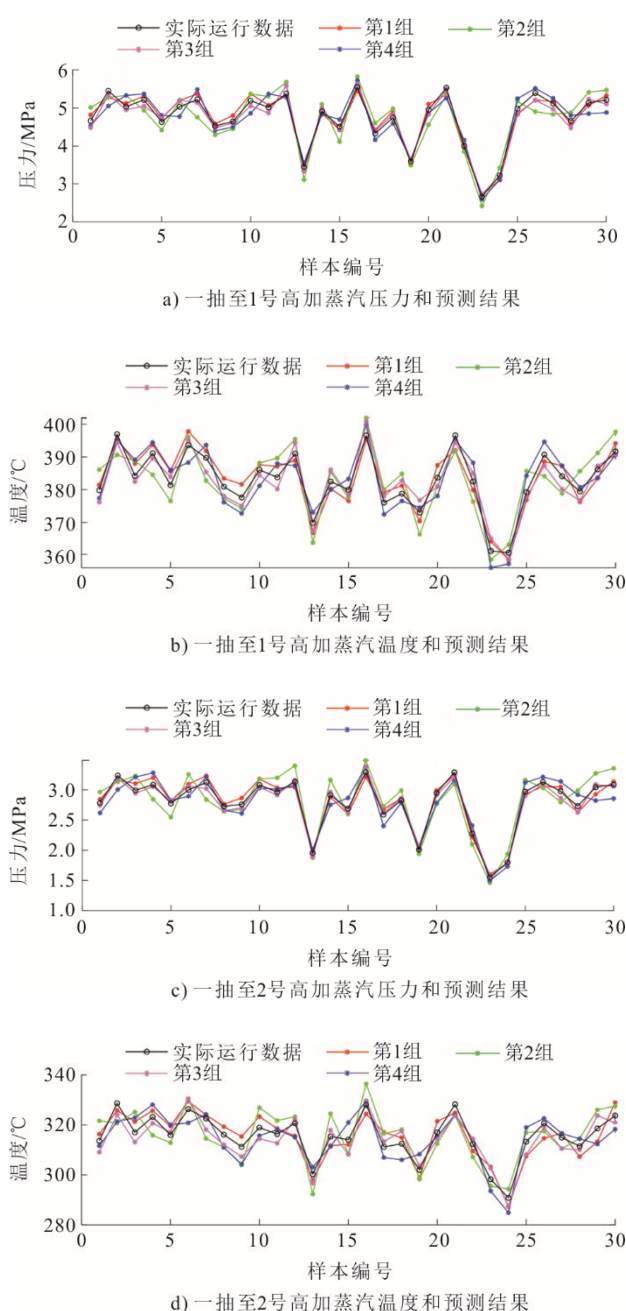
2) 根据第 3 节提出的多级相似度识别策略, 使得实时运行数据能够在历史数据聚类库内检索到最相近的历史数据用于后续的建模。这里拟将相似度度量函数中欧氏距离占比设为 0.6, 角度距离占比设为 0.4, 挑选出相似度最高的 20 组数据;

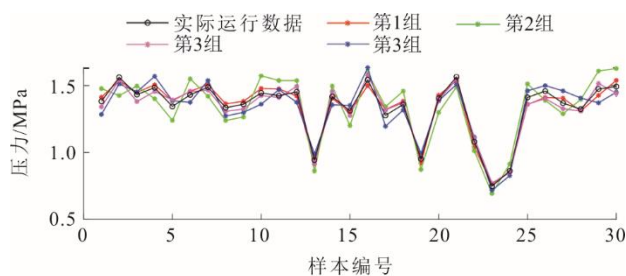
3) 将得到的 20 组数据按照 5.1 节参数设定, 依据第 4 节完成 ELM 的构建, 并利用实时运行数据最终获取到当前运行状态下模型预测的 10 个输出 $\bar{y}$ 。

5.3 结果分析

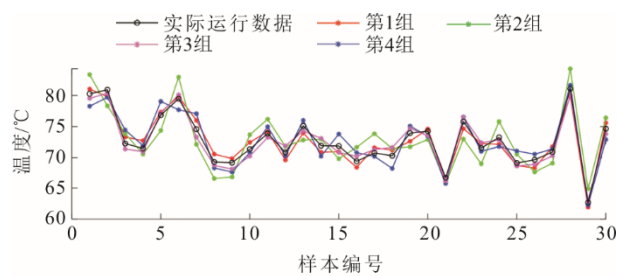
为了探究其他一些可能对实验结果产生影响的因素, 更改建模过程中的参数再进行 3 组实验。加上 5.2 节实验, 一共 4 组, 各进行 30 次实验, 将实际运行数据 y 和所得到的 10 个机组模型输出 $\bar{y}$ 进行对比。

图 4 为特征对比结果。其中, 第 1 组为 5.2 节实验结果, 用红色线表示; 第 2 组为聚类数量改为 4 的实验结果, 用绿色线表示; 第 3 组为相似度距离占比为 0.75 和 0.25 的实验结果, 用粉色线表示; 第 4 组为挑选 40 组相似度最高的历史数据的实验结果, 用蓝色线表示; 实际运行数据用黑色线表示。

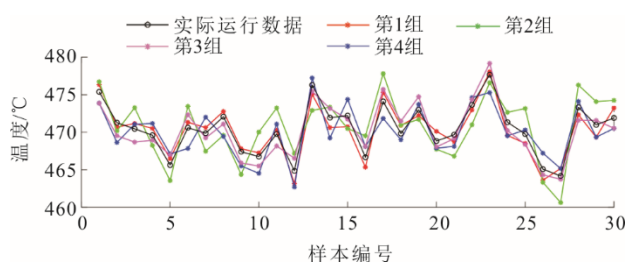




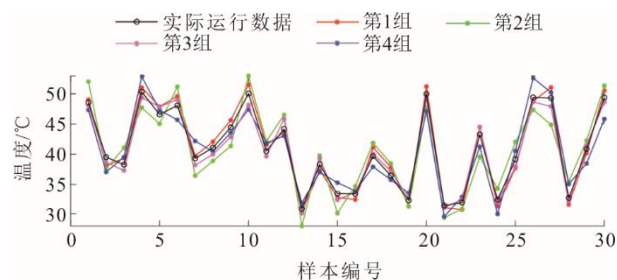
e) 一抽至3号高加蒸汽压力和预测结果



i) 7号低加出水温度和预测结果



f) 一抽至3号高加蒸汽温度和预测结果



j) 8号低加出水温度和预测结果

图4 特征对比结果

Fig.4 Feature comparison results

本文采用平均绝对百分比误差 (mean absolute percentage error, δ_{MAPE})、均方根误差 (root mean squared error, δ_{RMSE}) 作为模型精度评价指标。平均绝对百分比误差反映模型输出值偏离实际值的程度; 均方根误差反映模型输出值与实际值的样本标准差, 体现样本的离散程度。这 2 种评价指标越小代表模型精度越高, 其计算公式如下:

$$\delta_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \bar{y}_i}{y_i} \right| \times 100\% \quad (22)$$

$$\delta_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (23)$$

式中: y_i 为实际值; $\bar{y}_i$ 为模型输出值; n 为样本数量。

上述实验结果误差分析见表 2。

表2 不同特征参数实验结果误差分析

Tab.2 Error analysis of experimental results with different characteristic parameters

项目	$\delta_{MAPE}/\%$				δ_{RMSE}			
	第1组	第2组	第3组	第4组	第1组	第2组	第3组	第4组
一抽至1号高加蒸汽压力	2.32	5.60	2.23	3.89	0.119	0.276	0.117	0.209
一抽至1号高加蒸汽温度	0.73	1.26	0.70	0.93	2.920	5.093	2.861	3.786
一抽至2号高加蒸汽压力	2.42	6.27	1.97	4.43	0.075	0.189	0.059	0.141
一抽至2号高加蒸汽温度	1.16	1.76	1.09	1.36	3.878	5.912	3.691	4.665
一抽至3号高加蒸汽压力	2.27	6.78	2.29	4.28	0.032	0.093	0.034	0.062
一抽至3号高加蒸汽温度	0.22	0.50	0.26	0.35	1.082	2.487	1.291	1.769
5号低加出水温度	1.87	3.56	2.09	2.17	2.330	4.434	2.604	2.791
6号低加出水温度	2.23	3.19	2.08	2.32	2.161	2.966	2.033	2.187
7号低加出水温度	1.26	2.90	1.06	1.70	0.968	2.236	0.815	1.387
8号低加出水温度	2.55	5.84	2.15	4.30	1.094	2.500	0.980	1.939

统计每组实验平均消耗时间 (average time consumption, δ_{ATC}), δ_{ATC} 反映模型输出预测结果的速度, 数值越小模型收敛速度越快, 计算公式如下:

$$\delta_{ATC} = \frac{1}{n} \sum_{i=1}^n t_i \quad (24)$$

式中: t_i 为模型输出预测结果的耗时; n 为样本数量。上述实验结果耗时分析见表 3。

表 3 实验结果耗时分析 单位: s
Tab.3 Time consuming analysis of experimental results

项目	第 1 组	第 2 组	第 3 组	第 4 组
δ_{ATC}	1.442	2.056	1.488	1.783

由表 2、表 3 可以看出: 10 个特征误差中, 无论是哪种精度评价指标, 第 1 组和第 3 组的误差结果均为最小, 平均绝对百分比误差最小为 0.22%、均方根误差最小为 0.032, 并且这 2 组的耗时也是最少的, 第 1 组平均 1.442 s 就能输出预测结果, 说明 5.2 节实验和更改相似度距离占比所建立的机组模型精度最高, 模型计算收敛速度最快; 而第 2 组误差最大, 平均绝对百分比误差达到 6.78%, 均方根误差达到 5.912, 聚类数量减少使得相似度识别计算量增多, 平均消耗时间达到 2.056 s, 说明参数的设定确实会较大程度影响模型。

6 结 论

本文从热电联产机组建模的角度进行研究, 基于数字孪生相关技术, 建立通信关系并利用存储的大量运行数据构建了机组孪生模型, 为将来热力系统的智能优化提供了可靠的分析工具。以杭州某热电联产机组为研究对象进行验证, 结果表明:

1) 所建模型平均绝对百分比误差最小仅 0.22%, 均方根误差最小仅 0.032, 平均消耗时间为 1.442 s, 说明该建模方法的精度较高, 收敛速度较快, 并且在线条件下有效弥补了传统机理建模难以跟踪机组实时状态响应的不足。

2) 聚类数量改为 4 时, 平均绝对百分比误差为 6.78%, 均方根误差为 5.912, 平均消耗时间为 2.056 s, 模型精度降低、收敛速度变慢。因此建模过程中参数不同, 模型效果也会受到较大影响, 在实际应用的时候, 应该根据具体情况灵活选取参数来建立更高效的机组模型。

【参 考 文 献】

[1] JI H. Analysis of the operation principle of typical

cogeneration units[C]//Journal of Physics Conference Series. JPhCS, 2022.

- [2] 郭喜燕, 刘嘉康, 白雪, 等. 基于碳排放特性及碳交易规则的热电联产机组经济性分析[J]. 热力发电, 2023, 52(4): 14-23.
GUO Xiyang, LIU Jiakang, BAI Xue, et al. Economic analysis of cogeneration units based on carbon emission characteristics and carbon trading rules[J]. Thermal Power Generation, 2023, 52(4): 14-23.
- [3] DE MELLO F P. Boiler models for system dynamic performance studies[J]. IEEE Transactions on Power Systems, 1991, 6(1): 66-74.
- [4] 王琪. 风电规模化并网条件下供热机组优化控制研究[D]. 保定: 华北电力大学, 2013: 1.
WANG Qi. Research on optimization control for heating units under the condition of large-scale integration of wind power[D]. Baoding: North China Electric Power University, 2013: 1.
- [5] 邓拓宇, 田亮, 刘吉臻. 供热机组热电耦合特性分析与解耦协调控制[J]. 系统仿真学报, 2017, 29(10): 2593-2599.
DENG Tuoyu, TIAN Liang, LIU Jizhen. Thermal and electric load coupling analysis and decoupling coordinated control in heat supply units[J]. Journal of System Simulation, 2017, 29(10): 2593-2599.
- [6] 席嫣娜, 王印松. 一种基于自适应反演算法的火电单元机组协调控制策略[J]. 中国电机工程学报, 2018, 38(7): 2070-2081.
XI Yanna, WANG Yinsong. A coordinated control strategy for thermal power unit based on adaptive inversion algorithm[J]. Proceedings of the CSEE, 2018, 38(7): 2070-2081.
- [7] 田亮, 冯荣荣. 多热电联产机组建模及热动态耦合特性分析[J]. 动力工程学报, 2022, 42(1): 75-82.
TIAN Liang, FENG Rongrong. Modeling of multi-cogeneration unit and analysis of thermoelectric dynamic coupling characteristics[J]. Journal of Chinese Society of Power Engineering, 2022, 42(1): 75-82.
- [8] 尹二新. 大型火电机组控制系统数据驱动建模方法研究与应用[D]. 北京: 华北电力大学, 2018: 1.
YIN Erxin. Research and application of data-driven modeling method for large thermal power units control system[D]. Beijing: North China Electric Power University, 2018: 1.
- [9] 胡阳, 王蔚然, 房方, 等. 风电机组运行动态数字孪生建模及半物理仿真[J/OL]. 系统仿真学报: 1-12[2023-11-22]. <https://doi.org/10.16182/j.issn1004731x.joss.22-1307>.
HU Yang, WANG Weiran, FANG Fang, et al. Dynamic digital twin modeling and semi-physical simulation of wind turbine operation[J/OL]. Journal of System Simulation: 1-12[2023-11-22]. <https://doi.org/10.16182/j.issn1004731x.joss.22-1307>.
- [10] CERRONE A, HOCHHALTER J, HEBER G, et al. On the effects of modeling as-manufactured geometry: toward digital twin[J]. International Journal of Aerospace Engineering, 2014, 2014: 439278.1- 439278.10.
- [11] 董雷霆, 周轩, 赵福斌, 等. 飞机结构数字孪生关键建模仿真技术[J]. 航空学报, 2021, 42(3): 113-141.
DONG Leiting, ZHOU Xuan, ZHAO Fubin, et al. Key technologies for modeling and simulation of airframedigital twin[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(3): 113-141.
- [12] 房方, 姚贵山, 胡阳, 等. 风力发电机组数字孪生系统

- [J]. 中国科学: 技术科学, 2022, 52(10): 1582-1594.
FANG Fang, YAO Guishan, HU Yang, et al. Digital twin system of a wind turbine[J]. Scientia Sinica: Technologica, 2022, 52(10): 1582-1594.
- [13] 官邦, 汪东红, 马洪波, 等. 熔模铸件尺寸控制的数字孪生建模关键技术与应用[J/OL]. 金属学报: 1-16[2023-11-22]. <http://kns.cnki.net/kcms/detail/21.1139.TG.20230327.1010.002.html>.
GUAN Bang, WANG Donghong, MA Hongbo, et al. Key technology and application of digital twin modeling for deformation control of investment castings[J/OL]. Acta Metallurgica Sinica: 1-16[2023-11-22]. <http://kns.cnki.net/kcms/detail/21.1139.TG.20230327.1010.002.html>.
- [14] 仇晓黎, 朱睿, 幸研, 等. 螺线管装配生产线数字孪生建模技术[J]. 计算机集成制造系统, 2022, 28(6): 1696-1703.
QIU Xiaoli, ZHU Rui, XING Yan, et al. Digital twin modeling technology for solenoid assembly production line[J]. Computer Integrated Manufacturing Systems, 2022, 28(6): 1696-1703.
- [15] 姜珊, 王仲奇, 夏松, 等. 飞机柔性工装数字孪生几何模型构建方法[J]. 航空制造技术, 2022, 65(12): 86-91.
JIANG Shan, WANG Zhongqi, XIA Song, et al. Construction method of digital twin geometry model for aircraft flexible tooling[J]. Aeronautical Manufacturing Technology, 2022, 65(12): 86-91.
- [16] 肖通, 江海凡, 丁国富, 等. 五轴磨床数字孪生建模与监控研究[J]. 系统仿真学报, 2021, 33(12): 2880-2890.
XIAO Tong, JIANG Haifan, DING Guofu, et al. Research on digital twin-based modeling and monitoring of five-axis grinder[J]. Journal of System Simulation, 2021, 33(12): 2880-2890.
- [17] 郭安, 于东, 胡毅, 等. 基于 FFSM 的数控机床加工状态建模方法[J]. 计算机集成制造系统, 2019, 25(1): 71-80.
GUO An, YU Dong, HU Yi, et al. Fuzzy finite state machine based processing state modeling method for computer numerical control machine tools[J]. Computer Integrated Manufacturing Systems, 2019, 25(1): 71-80.
- [18] 王春晓. 基于数字孪生的数控机床多领域建模与虚拟调试关键技术研究[D]. 济南: 山东大学, 2018: 1.
WANG Chunxiao. Multi-domain modeling and virtual debugging of CNC machine tool based on digital twin[D]. Jinan: Shandong University, 2018: 1.
- [19] 孟松鹤, 叶雨玫, 杨强, 等. 数字孪生及其在航空航天中的应用[J]. 航空学报, 2020, 41(9): 6-17.
MENG Songhe, YE Yumei, YANG Qiang, et al. Digital twin and its aerospace applications[J]. Acta Astronautica Sinica, 2020, 41(9): 6-17.
- [20] 叶少槟. 基于数字孪生的复杂驾驶场景仿真模型的研究和实现[D]. 广州: 广东工业大学, 2022: 1.
- YE Shaobin. Research and implementation of complex driving scenario simulation model based on digital twin[D]. Guangzhou: Guangdong University of Technology, 2022: 1.
- [21] PARK J, LECHEVALIER D, AK R, et al. Gaussian process regression (GPR) representation in predictive model markup language (PMML)[J]. Smart & Sustainable Manufacturing Systems, 2017, 1(1): 121.
- [22] 王锡辉, 陈厚涛, 朱晓星, 等. 基于模糊 C 均值聚类的电站锅炉燃烧在线诊断[J]. 热力发电, 2019, 48(9): 77-82.
WANG Xihui, CHEN Houtao, ZHU Xiaoxing, et al. Online diagnosis for combustion of power station boilers based on fuzzy c mean clustering[J]. Thermal Power Generation, 2019, 48(9): 77-82.
- [23] 刘忠慧, 陈建宇, 宋国杰, 等. 基于模拟退火法的概念集构造算法[J]. 模式识别与人工智能, 2021, 34(8): 723-732.
LIU Zhonghui, CHEN Jianyu, SONG Guojie, et al. Construction algorithm of concept set based on simulated annealing algorithm[J]. Pattern Recognition and Artificial Intelligence, 2021, 34(8): 723-732.
- [24] 黄泽丰, 杨莘, 邓慧萍, 等. 基于 MDLatLRR 和 KPCA 的光场图像全聚焦融合[J]. 光子学报, 2023, 52(4): 255-269.
HUANG Zefeng, YANG Shen, DENG Huiping, et al. Light field all-in-focus image fusion based on MDLatLRR and KPCA[J]. Acta Photonica Sinica, 2023, 52(4): 255-269.
- [25] 牛大鹏, 刘元清. 基于改进即时学习算法的湿法冶金浸出过程建模[J]. 化工学报, 2017, 68(7): 2873-2879.
NIU Dapeng, LIU Yuanqing. Modeling hydrometallurgical leaching process based on improved just-in-time learning algorithm[J]. CIESC Journal, 2017, 68(7): 2873-2879.
- [26] 冯磊华, 张杰, 詹毅. 基于改进麻雀搜索算法和核极限学习机的电站锅炉燃烧优化[J]. 热力发电, 2022, 51(9): 96-102.
FENG Leihua, ZHANG Jie, ZHAN Yi. Combustion optimization of utility boilers based on ISSA and KELM[J]. Thermal Power Generation, 2022, 51(9): 96-102.
- [27] 毛力, 王运涛, 刘兴阳, 等. 基于改进极限学习机的短期电力负荷预测方法[J]. 电力系统保护与控制, 2012, 40(20): 5.
MAO Li, WANG Yuntao, LIU Xingyang, et al. Short-term power load forecasting method based on improved extreme learning machine[J]. Power System Protection and Control, 2012, 40 (20): 5.

(责任编辑 杜亚勤)