

基于 TICC 聚类的自监督学习核电设备运行工况划分

张大志¹, 郑胜², 崔文浩^{2,✉}

(1. 中核武汉核电运行技术股份有限公司, 湖北 武汉 443074;
2. 三峡大学理学院, 湖北 宜昌 443302)

摘要: [目的]随着核电的数字化发展,越来越多的核电设备数据得以被采集,运维人员通过数据分析即可获得各设备的运行工况。准确的核电设备的运行工况划分是实现核电设备健康评估、核电设备异常发现的基础。但是由于核电设备内部的传感器种类繁多,导致需要分析的数据量过于庞大,为人工划分核电设备的运行工况带来了巨大的挑战。为了能够实现核电设备运行工况准确快速地自动划分,提出了一种基于 TICC 聚类的自监督学习核电运行工况划分算法。[方法]首先,对核电设备的历史运行数据进行归一化处理,并利用手肘法确定最佳聚类数。再利用 TICC 聚类算法为核电历史运行数据进行工况分类,通过分类结果为各个工况的数据片段打上工况标签。最终,利用带标签的工况数据训练卷积神经网络获得工况划分模型。最终通过真实的核电设备运行数据进行验证。[结果]实验结果表明,所提出的算法的划分准确率达到 96.6%,划分速度仅需耗费 3.2 s。[结论]相较于 K-means 算法与 TICC 算法,研究提出的算法在准确率上和划分速度上均有较大提升,该算法可以有效地帮助核电运维人员完成核电设备的运行工况划分。

关键词: 运行工况; TICC 聚类; 卷积神经网络; 工况划分模型

中图分类号: TL4; TL38+2

文献标志码: A

文章编号: 2095-8676(2024)

OA: <https://www.energychina.press/>

Operating Condition Classification of Self-supervised Learning Nuclear Power Equipment Based on TICC Clustering

ZHANG Dazhi¹, ZHENG Sheng², CUI Wenhao^{2,✉}

(1. China Nuclear Power Operation Technology Corporation, Ltd., Wuhan 443074, Hubei, China;
2. College of Science, China Three Gorges University, Yichang 443302, Hubei, China)

Abstract: [Introduction] With the digital development of nuclear power, more and more nuclear power equipment data can be collected, and the operation and maintenance personnel can obtain the operation conditions of each equipment through data analysis. Accurate operation condition classification of nuclear power equipment is the basis for realizing the health assessment and anomaly discovery of nuclear power equipment. However, due to the wide variety of sensors inside the nuclear power equipment, the amount of data to be analyzed is too large, which brings great challenges to the manual classification of the operation conditions of nuclear power equipment. To achieve accurate and rapid automatic classification of nuclear power equipment operation conditions, this paper proposes a self-supervised learning algorithm for nuclear power operation condition classification based on TICC clustering. [Method] Firstly, the historical operation data of nuclear power equipment was normalized, and the elbow method was used to determine the optimal cluster number. Then the TICC clustering algorithm was used to classify the historical operation data of nuclear power equipment, and the data fragments of each condition were labeled by the classification results. Finally, the labeled condition data was used to train the convolutional neural network to obtain the condition classification model. Ultimately, the real operation data of nuclear power equipment was used for verification. [Result] The experimental results show that the classification accuracy of the proposed algorithm is 96.6%, and

收稿日期: 2023-12-27 修回日期: 2024-03-25

基金项目: 中核集团核工业仿真技术重点实验室对外开放基金项目“基于半监督/自监督的深度学习异常检测方法研究”(B220631)

the classification needs only 3.2 seconds. [Conclusion] Compared with the K -means algorithm and TICC algorithm, the algorithm proposed in this paper has a great improvement in accuracy and classification speed, and the algorithm can effectively help nuclear power operation and maintenance personnel complete the classification of operating conditions of nuclear power equipment.

Key words: operating conditions; TICC clustering; convolutional neural network; working condition classification model

0 引言

随着时代的发展,工业数字化的大幕已经拉开。作为在我国新能源发展过程中占据重要地位的核电行业,对数字化的需求也日益增强^[1-2]。核电作为一种绿色能源,具有污染低、发电效率高的特点。2020年,我国明确提出了2030年“碳达峰”和2060年“碳中和”^[3-4]的发展目标,而核电刚好可以满足这一发展需求。因此核电的发展在我国未来的能源布局中将占有重要地位。由于其使用原料的特殊性,核电的安全性^[5-6]又受到了极大的关注。因此,如何保证核电站的安全运行便成了核电运维人员的研究重点。而核电设备的工况划分可以为核电设备的健康评估^[7-9]、核电设备的异常发现^[10-14]、在线监测^[15-16]打下基础。因此,核电设备的运行工况划分是一件具有重要意义的工作。

近些年,在工况的自动划分算法研究上已经取得了一些进展。聚类算法^[17-18]是一种可以将具有一定相似度的数据聚为一类的无监督算法,这种算法从原理上较为适用于工况划分。 K -means算法^[19]是一种快速有效的经典聚类算法,该算法基于距离完成聚类,该算法原理简单,收敛速度快,在处理大规模的数据集时具有良好的效果。王仲等^[20]基于火电机组的历史运行数据,利用 K -means算法完成了火电机组的稳态工况划分。但是 K -means算法对聚类初值和初始的聚类中心选择较为敏感,且不适用于非球形的数据集^[21-22]。赵帅等^[23]采用高斯混合模型(Gaussian Mixture Model, GMM)算法模型获得各工况下能效状态的基准区间,并利用BP算法建立机组多元工况下的基准值工况库。但是该方法在面对大体量数据时会出现计算困难的问题。陈梅等^[24]采用融合聚类算法得到篦冷机不同运行工况下的聚类标签。最后,利用RBF-SVM算法对得到的聚类标签进行训练得到工况识别模型,从而实现了篦冷机的运行工况自动划分。但该方法在聚类过程中需要一定的先验知识,使得该方法的普适性不强。

目前,核电设备的运行工况依然是通过运维人员的经验来进行划分,但是随着数字化的发展,传感器所记录的数据变得越来越多,核电运维人员难以综合各传感器因素进行准确的工况划分。因此,寻找到一种可以实现自动快速划分核电设备运行工况的方法变得尤为迫切。针对该问题,本研究提出了一种基于TICC聚类^[25-26]的自监督学习核电设备运行工况划分算法,该算法通过TICC聚类算法(Toeplitz Inverse Covariance-based Clustering)获得核电设备在不同运行工况下的聚类标签,通过已获得的聚类标签训练卷积神经网络^[27],从而获得核电设备的运行工况划分模型。最后,利用核电设备的历史运行数据展开实验,验证了该算法对核电设备运行工况划分的准确性,并与其他常见的聚类算法做比较,证实了该方法的有效性。

1 算法设计

1.1 基于TICC聚类的自监督学习

基于TICC聚类的自监督学习核电设备运行工况划分算法的具体步骤如下所示:

1) 导入核电设备的运行数据。

2) 由于核电设备中的传感器种类不同,导致所获得的数据存在着量纲不同的问题。为了消除量纲的影响,便于后续处理,需要对核电运行数据进行归一化处理。归一化公式如公式(1)所示。

$$x^* = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \quad (1)$$

式中:

x^* ——归一化数据;

x ——原始数据;

x_{\max} 和 x_{\min} ——原始数据中的最大值和最小值。

3) 由于聚类初值的选择,会对最终的聚类结果造成影响。因此,为了提高TICC聚类算法的聚类结果准确性,需要选择一个合适的聚类初值。对此,可利用手肘法确定最佳聚类数。手肘法的核心指标是SSE(Sum of the Squared Errors, 误差平方和),

SSE 表达所有的聚类样本的误差。通过该方法寻找最佳聚类数时, 由于在到达最佳聚类数前, SSE 会大幅下降, 而到达最佳聚类数以后, 再增加聚类数 SSE 的下降会趋于缓慢, 因此形成了类似于手肘的曲线图。

4) 由于核电设备运行数据是时间序列数据, 而 TICC 聚类算法对时间序列数据有很好的效果。因此, 利用 TICC 聚类算法, 对归一化矩阵进行聚类, 获得不同运行工况下的聚类标签。

5) 利用具有标签的数据训练卷积神经网络, 获得工况划分模型。

6) 导入真实的核电设备运行数据, 利用工况划分模型完成工况划分。

该算法的流程图如图 1 所示。

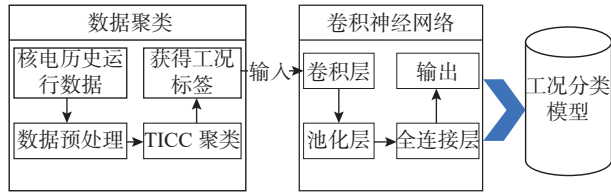


图 1 算法流程图

Fig. 1 Algorithm flow chart

1.2 TICC 聚类算法

TICC 聚类算法是由 HALLAC 于 2017 年提出的一种基于托普利兹逆协方差矩阵聚类算法。该算法通过考察各维度之间的相关性, 以确定信号间的相似性。而高维核电数据的每一行对应着一个时间戳, 每一列对应着一个传感器参数, 因此可以将核电运行数据集描述为一个时间序列矩阵 \mathbf{X} 。其中 \mathbf{X} 可以定义为公式(2)。

$$\mathbf{X} = [x_1, x_2, \dots, x_n] \quad (2)$$

式中:

x_n ——第 n 列的传感器数据。

TICC 算法的关键就在于将时序数据的各个维度按照时间戳切割成具有一定宽度的时间段, 然后把每个时间段内的数据进行聚类, 最终得到整体的聚类结果。核电数据是一个 n 维的数据向量, 那么每一个时间段内的数据可以以窗口的形式来定义, 假设窗口的大小为 w , 那么时间段的数据 \mathbf{X}_i 则可以定义为公式(3)的形式。

$$\mathbf{X}_i = [x_{i-w+1}, \dots, x_{i-1}, x_i] \quad (3)$$

此处, \mathbf{X}_i 为 $n \times w$ 维的数据矩阵。可将每一个数

据矩阵聚为 K 类, 且认为每一个数据段内的数据均服从 0 均值高斯分布, 即可获得其协方差逆矩阵 $\boldsymbol{\theta}_j$, $j=1, 2, \dots, K$ 。该矩阵由 $w \times w$ 个子矩阵构成, 每个子矩阵的维度为 $n \times n$, 最终构成的矩阵 $\boldsymbol{\theta}_i$ 的维度为 $nw \times nw$ 。任意两个位置上的子矩阵, 表示这两个时刻间各维度的协方差逆矩阵。由此可知, 矩阵 $\boldsymbol{\theta}_i$ 实质上是一个分块托普利兹矩阵。TICC 算法希望找到不同变量间的相关性, 由此 TICC 聚类的最终可以转化为数学问题, 即对 K 个逆协方差进行求解, 将数据段分配至 K 个分类之中, 并保每个数据点只有一种分类。其表达式如公式(4)。

$$\operatorname{argmin}_{\boldsymbol{\theta} \in \Phi, P} \left[\sum_{i=1}^K \left[\|\lambda^\circ \boldsymbol{\theta}_i\|_1 + \sum_{X_i \in P_i} (-\ell\ell(X_i, \boldsymbol{\theta}_i) + \beta l\{X_{i-1} \notin P_i\}) \right] \right] \quad (4)$$

式中:

P —— K 个分配集群的集合;

λ ——正则化参数;

$\|\lambda^\circ \boldsymbol{\theta}_i\|_1$ ——哈达玛乘积的 1 范数惩罚;

$\ell\ell(X_i, \boldsymbol{\theta}_i)$ —— X_i 来自聚类 i 的对数似然值;

β ——强制时间一致性的参数;

$l\{X_{i-1} \notin P_i\}$ ——检查相邻点是否被分配给同一聚类的指示函数;

Φ —— nw 相邻点的对称分块托普利兹矩阵的集合。

在此, 对似然函数 $\ell\ell(X_i, \boldsymbol{\theta}_i)$ 可以展开为公式(5)。

$$\ell\ell(X_i, \boldsymbol{\theta}_i) = -\frac{1}{2}(X_i - \mu_i)^\top \boldsymbol{\theta}_i (X_i - \mu_i) + \frac{1}{2} \lg[\det(\boldsymbol{\theta}_i)] - \frac{n}{2} \lg(2\pi) \quad (5)$$

μ_i 为第 i 类的经验平均值。对于整个 TICC 问题可分解为两个需要轮流求解的问题: (1) 给定 $\boldsymbol{\theta}_j$, 求解数据段的分类 P_j ; (2) 给定分类 P_j , 求解逆协方差矩阵 $\boldsymbol{\theta}_j$ 。 P_j 指的是数据段分类为 K 类时第 j 分类的集合。当给定 $\boldsymbol{\theta}_j$ 时, 求解分类集合, 即将数据点分配到聚类当中时。可通过公式(6)求解。

$$\operatorname{minimize} \sum_{i=1}^K \sum_{X_i \in P_i} -\ell\ell(X_i, \boldsymbol{\theta}_i) + \beta l\{X_{i-1} \notin P_i\} \quad (6)$$

该公式即解决将 T 个时间子序列分配至 K 个聚类之中。在给定了分类集合 P_j 时, 需要求解各个逆协方差矩阵 $\boldsymbol{\theta}_j$, 该问题可通过公式(7)求解。

$$\sum_{X_i \in P_i} -\ell(X_i, \theta_i) = -|P_i| \{ \lg[\det(\theta_i)] + \text{tr}(S_i \theta_i) \} + C \quad (7)$$

式中:

$|P_i|$ ——处于第 i 个聚类中的数据点数量;

S_i ——这些点的经验协方差;

C ——常数。

利用 EM 算法可对该问题进行迭代优化,其 M 步的表达式可以为公式(8)所示。

$$\begin{aligned} & \text{minimize } \lg[\det(\theta_i)] + \text{tr}(S_i \theta_i) + \frac{1}{|P_i|} \|\lambda \circ \theta_i\|_1 \\ & \text{subject to } \theta_i \in \Phi \end{aligned} \quad (8)$$

TICC 算法的关键就在于解决上述两个问题,将每个数据点分配至聚类之中和更新聚类参数。因此,便有了 ADMM 算法来求解上述问题。引入一致变量 Z ,便可得到公式(9)所示。

$$\begin{aligned} & \text{minimize } \lg[\det(\theta)] + \text{tr}(S\theta) + \|\lambda \circ Z\|_1 \\ & \text{subject to } \theta \in Z, Z \in \Phi \end{aligned} \quad (9)$$

该式的带拉格朗日项如公式(10)所示。

$$\begin{aligned} \zeta_\rho(\theta, Z, U) = & -\lg[\det(\theta)] + \text{tr}(S\theta) + \\ & \|\lambda \circ Z\|_1 + \frac{\rho}{2} \|\theta - Z + U\|_F^2 \end{aligned} \quad (10)$$

$\rho > 0$ 为 ADMM 惩罚系数, $U \in R^{m \times m \times w}$ 为缩放对偶变量, $Z \in \Phi$ 。ADMM 的迭代过程可以为公式(11-13)所示。

$$\theta^{k+1} = \underset{\theta}{\text{argmin}} \zeta_\rho(\theta, Z^k, U^k) \quad (11)$$

$$Z^{k+1} = \underset{Z \in \Phi}{\text{argmin}} \zeta_\rho(\theta^{k+1}, Z, U^k) \quad (12)$$

$$U^{k+1} = U^k + (\theta^{k+1} - Z^{k+1}) \quad (13)$$

其中, k 为迭代次数,三个步骤迭代优化直至收敛

为止。

1.3 工况划分模型

卷积神经网络^[28-29]在采用固定尺度的卷积核提取目标特征时只能提取到局部信息,而核电设备在发生工况切换的过程中,不同传感器的切换过程的持续时间会不同,如果只使用固定尺度的卷积核去提取核电运行数据的特征,会造成对部分变化信息的提取遗漏,无法准确地还原整个工况变化信息。为了能够更为准确地捕捉到核电运行数据的整个数据特征,可采用不同尺度的卷积核对整个核电数据进行特征提取^[30],获取设备的工况信息。卷积神经网络的工作过程如图2所示。

输入为经过 TICC 聚类分析获得了工况标签的数据。采用不同大小的卷积核对输入数据进行特征提取,其中三个输入均会经过不同尺寸的卷积核,以保证可以充分提取到工况信息的特征。经过卷积操作以后,利用 Relu 函数作为激活函数加速模型训练速度。其中 Relu 激活函数的表达式如公式(14)所示。

$$\text{Relu}(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (14)$$

之后再利用池化层 Max pooling 对卷积层对提取到的特征进行降维压缩,以加快运算速度,并降低模型过拟合的风险。完成上述工作以后,将经过处理的特征数据输入到全连接层,全连接层将最终的拼接特征进行整合,作为分类器的输入。由于 Dropout 层可以使全连接层中的一部分神经元不工作,从而提高了网络的泛化能力。所以可以在此过程中添加 Dropout 层,以达到降低模型过拟合风险的目的。最终通过了 Softmax 层输出分类结果。

在该网络中,训练集和测试集以 8:2 的比例进行划分。输入数据的尺寸为 $(n, 17)$,输出数据的尺寸为 $(n, 4)$ 。其中, n 为样本数。17 为输入数据的维度

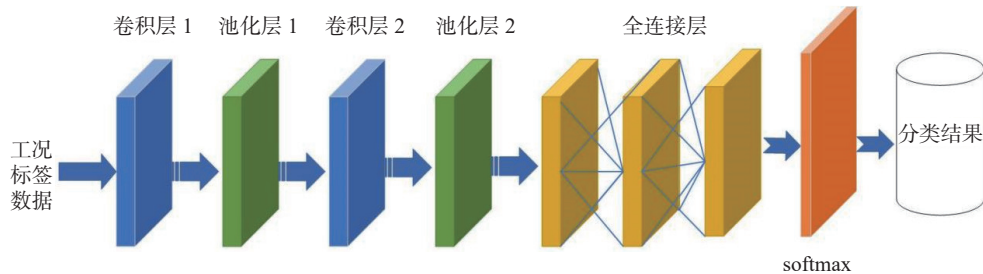


图2 卷积神经网络结构

Fig. 2 Convolutional neural network structure

(传感器个数), 4 为分类个数。

2 实验过程及结果分析

2.1 RCV002PO 数据

为了验证本研究提出算法的有效性, 将利用核电厂设备 RCV002PO 的运行数据对方法进行验证。该设备为 RCV(化学和容积控制系统)系统内部的 002 号上充泵, 该泵的作用是将容控箱中的来水升压送入一回路。该设备的数据集记录了 RCV002PO 自 2019 年 7 月 29 日至 2021 年 12 月 31 日的运行数据, 共含有 17 个相关参数, 其中参数类型包括了电流、振动和温度。其具体参数如表 1 所示。

2.2 实验过程及结果分析

2.2.1 实验过程

对核电运行数据先进行归一化, 消除量纲的影响。针对核电运行数据难以确定工况个数的问题, 利用手肘法确定最佳聚类数 K 。由图 3 可以看到, SSE(误差平方和)在聚类数由 2 至 3 时出现了大幅的下降, 由 3 至 4 时也有较大的降幅, 之后再增加聚类数时, 降幅趋于缓慢, 因此在聚类时可以将聚类数

确定为 3 或者 4, 但是为了使更加多的散点可以被归为一类, 可将最佳聚类数取为 4, 从而使得散点可以不被错误地划分到其他工况。

对于窗口 w 的大小取值, 不能取得过大, 因为过大的窗口会导致在窗口边界上的点难以被正确分类。因此, 根据运维人员的经验将窗口大小定为 180。利用 TICC 聚类算法对数据进行聚类, 为核电设备的运行数据提供了标签。将获得标签的数据输入至多尺度特征提取网络当中对网络进行训练, 如图 4 展现了整个网络的训练过程, 可以看到训练集的准确率随着迭代进行的变化过程。在训练开始以后, 训练集的整个网络的损失迅速下降, 而准确率迅速上升。在经过 10 个轮次的训练以后, 整个网络的损失和准确率开始趋于稳定, 整个模型的参数得到了充分的学习, 在进一步的学习当中, 损失和准确率均未发生较大的波动, 此时可认为工况分类模型已经训练完成。

2.2.2 结果分析

在获得了训练好的工况分类模型以后, 可将核电设备的运行数据导入, 开始进行工况分类测试, 如

表 1 参数详情表

Tab. 1 Parameter details table

传感器名称	类型	2019-07-29T 14: 13: 19	2019-07-29T 14: 13: 20	2021-12-31T 23: 51: 27	2021-12-31T 23: 51: 28
RCV002MI	电流	44.41	44.39	-0.08	-0.08
RCV200MT	温度	37.96	38.02	26.37	26.37
RCV201MT	温度	42.29	42.57	26.11	26.11
RCV202MT	温度	60.64	60.86	30.47	30.47
RCV203MT	温度	54.91	55.13	30.38	30.38
RCV204MT	温度	63.69	64.06	30.81	30.81
RCV205MT	温度	46.43	46.39	24.39	24.39
RCV206MT	温度	45.51	45.51	24.57	24.57
RCV210MT	温度	71.37	71.37	24.04	24.04
RCV222MT	温度	66.73	66.76	24.18	24.18
RCV223MT	温度	61.49	61.49	24.44	24.44
RCV224MT	温度	47.77	47.77	24.35	24.35
RCV231MV	振动	0.56	0.56	0.06	0.06
RCV232MV	振动	0.51	0.51	0.05	0.05
RCV233MV	振动	0.67	0.67	0.15	0.15
RCV234MV	振动	1.20	1.21	0.06	0.06
RCV254MT	温度	75.56	75.56	23.91	23.91

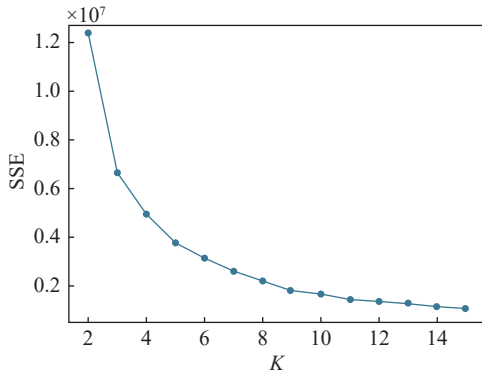
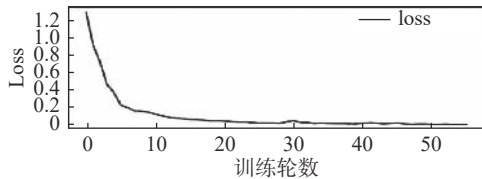
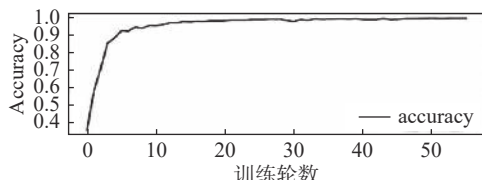


图 3 手肘法
Fig. 3 Elbow method



(a) 训练损失曲线图



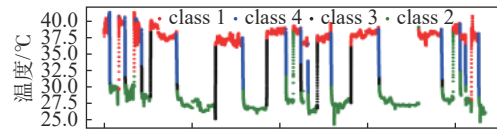
(b) 训练准确率曲线图

图 4 网络训练过程

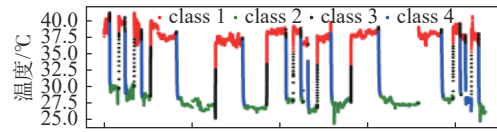
Fig. 4 Network training process

图 5 所示, 为本研究所提出方法与 TICC 聚类算法和 K-means 聚类算法的划分结果对比图。从图中可以清晰地看到两种方法的划分工况数均为 4 类, 划分方式也基本一致。两个平稳的运行工况分别划归为工况 1(红色)和工况 2(绿色), 上升段的运行状态划分为工况 3(黑色), 下降段的运行状态划分为工况 4(蓝色)。每个工况在划分准确的前提下, 应该是只有一种颜色而没有其他颜色掺杂其中的。例如在工况 3(上升段)这种工况中, 整段应该均为黑色, 如若混入其他颜色的点, 则可认为是划分错误的点。可以清晰的看到, 三种方法在工况 1 和工况 2 的分类准确率上基本一致, 三种方法的分类效果差异主要体现在对变化段工况的准确性上。从划分的结果图上看, 可以看到本研究所提出方法相比于 TICC 算法对上升段工况和下降段工况有较为准确的划分, 但是 TICC 算法在对工况 3 的划分上出现了较为明

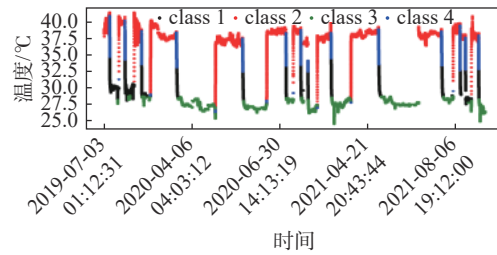
显的偏差。TICC 算法在这一类工况划分上出现了工况穿插的情况, 在本属于工况 3 的分类中混入了工况 1 的分类。造成这一现象出现的原因是因为 TICC 聚类算法是以窗口划分数数据段的方式进行聚类的, 在聚类的边界往往会出现这种工况穿插的情况。由于本研究所提出的算法是基于 TICC 算法所提取的标签训练来完成的, 因此在本研究算法中也出现了类似的现象。



(a) 文章算法



(b) TICC 算法



(c) K-means 算法

图 5 工况划分结果图

Fig. 5 Results of working condition classification

K-means 算法在精确度上较本研究算法和 TICC 聚类方法存在一定的差距, K-means 算法的划分结果在平稳段工况的表现较好, 其差距主要体现在了上升段工况和下降段工况上, K-means 算法的划分结果上出现了将整体的上升段工况划分至了平稳工况之中, 对于下降段工况的处理也不理想, 出现了非常明显的工况延伸情况, 总体的精确度比较差, 造成这种状况的原因在于 K-means 聚类算法的聚类方式, K-means 聚类算法是基于划分的聚类算法, 该算法通过欧式距离来完成计算, 该种方式会造成下降段工况中的部分点被错误地归纳到邻近的平稳段工况中, 当下降段中的点数较少时就造成了整体工况的划分错误。

文章算法在准确率上较 TICC 聚类算法和 K-means 聚类算法的划分结果均有较为明显的提升, 但

是也有工况延伸的情况出现,造成这种情况的原因是因为每次送入网络的数据的数量是固定的,数量过大或过小均会造成误差的出现,在部分工况段数据点数量较少时就会出现明显的工况延伸情况,但是这种情况相较于 TICC 和 K -means 算法已经得到了较好的抑制,因此准确率也得到了提升。

除了在精确率上有所提高以外,文章算法相较于 TICC 算法和 K -means 算法的最大优势在于本研究算法的划分速度得到了大大的提高。尤其是与 TICC 算法的对比中,文章算法由于第一次使用时需要利用 TICC 算法提取工况标签,因此在耗时上与 TICC 相差不大。但是在获得了工况分类模型以后,当新的数据被送入划分模型之后,文章算法可以在极短的时间内划分出工况,而 TICC 算法需要再次经过大量的计算才可获取工况分类,因此本研究所提出的算法在划分效率上得到了大大的提升。

在准确率对比上,可采用分工况计算方式,由于划分了四种工况,以人工划分结果为基准,利用算法划分结果与人工划分结果做对比,与人工划分结果一致的为正确。最终三种算法的性能对比如表 2 所示。

表 2 不同算法的工况划分准确率对比

Tab. 2 Comparison of accuracy of different algorithms for working condition classification

方法	工况1 准确率	工况2 准确率	工况3 准确率	工况4 准确率	总体时间/s	划分时间/s
文章算法	98.97%	98.78%	92.61%	91.93%	4221.7	3.4
TICC算法	98.89%	98.71%	87.33%	94.78%	4218.3	4218.3
K -means算法	98.56%	98.23%	32.12%	60.24%	45.4	45.4

3 结论

本研究提出了一种基于 TICC 聚类的自监督学习核电设备运行工况划分算法。该方法在应用于真实的核电设备运行数据以后展现了良好的性能。(1)在工况划分准确率上,本研究所提出的方法达到了 96.6%,而在获得了工况分类模型以后,整个划分过程所需的时间为 4221.7 s。但是,除去提取标签的时间以后,划分时间仅为 3.4 s,之后的每一次划分均不需要提取标签,大大提高了划分效率;(2)相较于 TICC 算法,本研究所提出的算法在具备了 TICC 算

法的精确性同时,在划分效率上有了巨大的提高,避免了 TICC 聚类算法耗时时间长的难题;(3)与 K -means 算法相比较而言,本研究所提出的算法不仅仅准确率在上升段和下降段的准确率体现出了巨大的优势。在划分速度上,在模型建立完成后,划分速度仅为 3.4 s,相较于 K -means 算法也有较大提升。本研究所提出的算法在同样的数据集上,其划分准确率更高,划分所需时间更短,展现了更为优秀的划分效果。其性能相较于 TICC 算法和 K -means 算法更加适用于现实核电运维人员的需求。可以实现核电设备运行工况的快速准确地自动划分。

参考文献:

- [1] 谭思超,李桐,刘永超,等.关于人工智能在核能领域应用的若干思考[J].核动力工程,2023,44(2):1-8. DOI: 10.13832/j.jnpe.2023.02.0001.
TAN S C, LI T, LIU Y C, et al. Thoughts on the application of artificial intelligence in nuclear energy field [J]. Nuclear power engineering, 2023, 44(2): 1-8. DOI: 10.13832/j.jnpe.2023.02.0001.
- [2] 叶奇蓁.我国核电及核能产业发展前景[J].南方能源建设,2015,2(4):18-21. DOI: 10.16516/j.gedi.issn2095-8676.2015.04.001.
YE Q Z. Development and prospect of nuclear power and nuclear energy industry in China [J]. Southern energy construction, 2015, 2(4): 18-21. DOI: 10.16516/j.gedi.issn2095-8676.2015.04.001.
- [3] 胡道明,李蛟,杜晓东,等.碳中和背景下含氢综合能源系统碳排放和经济性分析[J].热能动力工程,2023,38(4):111-120. DOI: 10.16146/j.cnki.rndlgc.2023.04.015.
HU D M, LI J, DU X D, et al. Carbon emissions and economic analysis of integrated energy system containing hydrogen using for carbon neutrality [J]. Journal of engineering for thermal energy and power, 2023, 38(4): 111-120. DOI: 10.16146/j.cnki.rndlgc.2023.04.015.
- [4] 蔡绍宽.双碳目标的挑战与电力结构调整趋势展望[J].南方能源建设,2021,8(3):8-17. DOI: 10.16516/j.gedi.issn2095-8676.2021.03.002.
CAI S K. Challenges and prospects for the trends of power structure adjustment under the goal of carbon peak and neutrality [J]. Southern energy construction, 2021, 8(3): 8-17. DOI: 10.16516/j.gedi.issn2095-8676.2021.03.002.
- [5] 柴国早,杨志义,肖军,等.福岛核事故后核安全改进行动及安全要求研究[J].原子能科学技术,2022,56(3):399-409. DOI: 10.7538/yzk.2022.youxian.0053.
CHAI G H, YANG Z Y, XIAO J, et al. Nuclear safety improvement actions and safety requirements in post-fukushima era [J]. Atomic energy science and technology, 2022, 56(3): 399-409. DOI: 10.7538/yzk.2022.youxian.0053.

- [6] 邢继, 吴宇翔, 刘倩雯, 等. 整体平衡的核安全观和能动与非能动先进核电厂 [J]. *核科学与工程*, 2023, 43(1): 141-150. DOI: [10.3969/j.issn.0258-0918.2023.01.022](https://doi.org/10.3969/j.issn.0258-0918.2023.01.022).
XING J, WU Y X, LIU Q W, et al. Overall & balanced nuclear safety concept and active & passive advanced nuclear power plant [J]. *Nuclear science and engineering*, 2023, 43(1): 141-150. DOI: [10.3969/j.issn.0258-0918.2023.01.022](https://doi.org/10.3969/j.issn.0258-0918.2023.01.022).
- [7] 沈江飞, 潘天成, 毛晓明, 等. 核电厂重大设备健康状态的模糊综合评价方法 [J]. *核动力工程*, 2018, 39(6): 104-110. DOI: [10.13832/j.jnpe.2018.06.0104](https://doi.org/10.13832/j.jnpe.2018.06.0104).
SHEN J F, PAN T C, MAO X M, et al. Study on condition detection of major equipment in nuclear power plants based on fuzzy synthetic assessment [J]. *Nuclear power engineering*, 2018, 39(6): 104-110. DOI: [10.13832/j.jnpe.2018.06.0104](https://doi.org/10.13832/j.jnpe.2018.06.0104).
- [8] 严超. 核电厂热阱维持系统健康状态评估技术研究 [D]. 哈尔滨: 哈尔滨工程大学, 2016. DOI: [10.7666/d.Y3176873](https://doi.org/10.7666/d.Y3176873).
YAN C. Study on a health state assessment technology for heat sink systems in nuclear power plant [D]. Harbin: Harbin Engineering University, 2016. DOI: [10.7666/d.Y3176873](https://doi.org/10.7666/d.Y3176873).
- [9] 赵性爽. 核电厂运行健康状态评估方法研究 [D]. 哈尔滨: 哈尔滨工程大学, 2016. DOI: [10.7666/d.Y3176867](https://doi.org/10.7666/d.Y3176867).
ZHAO X S. Research on operation health condition assessment for nuclear power plant [D]. Harbin: Harbin Engineering University, 2016. DOI: [10.7666/d.Y3176867](https://doi.org/10.7666/d.Y3176867).
- [10] 许勇, 蔡云泽, 宋林. 基于数据驱动的核电设备状态评估研究综述 [J]. *上海交通大学学报*, 2022, 56(3): 267-278. DOI: [10.16183/j.cnki.jsjtu.2021.502](https://doi.org/10.16183/j.cnki.jsjtu.2021.502).
XU Y, CAI Y Z, SONG L. Review of research on condition assessment of nuclear power plant equipment based on data-driven [J]. *Journal of Shanghai Jiaotong University*, 2022, 56(3): 267-278. DOI: [10.16183/j.cnki.jsjtu.2021.502](https://doi.org/10.16183/j.cnki.jsjtu.2021.502).
- [11] 侯修群, 蒋庆磊, 包彬彬, 等. 基于相关系数的核电主泵振动异常定位方法研究 [J]. *核科学与工程*, 2021, 41(5): 920-928. DOI: [10.3969/j.issn.0258-0918.2021.05.006](https://doi.org/10.3969/j.issn.0258-0918.2021.05.006).
HOU X Q, JIANG Q L, BAO B B, et al. Study on abnormal vibration localization method of nuclear main pump based on correlation coefficient [J]. *Nuclear science and engineering*, 2021, 41(5): 920-928. DOI: [10.3969/j.issn.0258-0918.2021.05.006](https://doi.org/10.3969/j.issn.0258-0918.2021.05.006).
- [12] 沈小成. 基于 LSTM-Bayes 模型的核电汽轮机监测信号异常识别方法 [D]. 广州: 华南理工大学, 2021. DOI: [10.27151/d.cnki.ghnlu.2021.004798](https://doi.org/10.27151/d.cnki.ghnlu.2021.004798).
SHEN X C. Monitoring signal abnormal identification method based on LSTM-Bayes model for nuclear power turbine [D]. Guangzhou: South China University of Technology, 2021. DOI: [10.27151/d.cnki.ghnlu.2021.004798](https://doi.org/10.27151/d.cnki.ghnlu.2021.004798).
- [13] 杨隽豪, 韦延方, 王鹏, 等. 基于 Swin Transformer 的柔性直流电网单端量故障诊断 [J]. *电力建设*, 2023, 44(5): 53-60. DOI: [10.12204/j.issn.1000-7229.2023.05.006](https://doi.org/10.12204/j.issn.1000-7229.2023.05.006).
YANG J H, WEI Y F, WANG P, et al. Single-ended fault diagnosis of flexible DC grid based on swin transformer [J]. *Electric power construction*, 2023, 44(5): 53-60. DOI: [10.12204/j.issn.1000-7229.2023.05.006](https://doi.org/10.12204/j.issn.1000-7229.2023.05.006).
- [14] 肖发龙, 吴岳忠, 沈雪豪, 等. 基于深度学习和知识图谱的变电站设备故障智能诊断 [J]. *电力建设*, 2022, 43(3): 66-74. DOI: [10.12204/j.issn.1000-7229.2022.03.008](https://doi.org/10.12204/j.issn.1000-7229.2022.03.008).
XIAO F L, WU Y Z, SHEN X H, et al. Intelligent fault diagnosis of substation equipment on the basis of deep learning and knowledge graph [J]. *Electric power construction*, 2022, 43(3): 66-74. DOI: [10.12204/j.issn.1000-7229.2022.03.008](https://doi.org/10.12204/j.issn.1000-7229.2022.03.008).
- [15] 刘文倩, 韩利峰, 黄丽, 等. 基于数字孪生的 TMSR-SF0 数据监控与可视化方案 [J]. *核技术*, 2022, 45(2): 020604. DOI: [10.11889/j.0253-3219.2022.hjs.45.020604](https://doi.org/10.11889/j.0253-3219.2022.hjs.45.020604).
LIU W Q, HAN L F, HUANG L, et al. TMSR-SF0 data monitoring and visualization scheme based on digital twin [J]. *Nuclear techniques*, 2022, 45(2): 020604. DOI: [10.11889/j.0253-3219.2022.hjs.45.020604](https://doi.org/10.11889/j.0253-3219.2022.hjs.45.020604).
- [16] 沈江飞, 王双飞, 黄立军, 等. 核电厂设备状态多参数同步监测技术研究 [J]. *核动力工程*, 2022, 43(4): 168-173. DOI: [10.13832/j.jnpe.2022.04.0168](https://doi.org/10.13832/j.jnpe.2022.04.0168).
SHEN J F, WANG S F, HUANG L J, et al. Research on multi-parameter synchronous monitoring technology of nuclear power plant equipment status [J]. *Nuclear power engineering*, 2022, 43(4): 168-173. DOI: [10.13832/j.jnpe.2022.04.0168](https://doi.org/10.13832/j.jnpe.2022.04.0168).
- [17] 姬强, 孙艳丰, 胡永利, 等. 深度聚类算法研究综述 [J]. *北京工业大学学报*, 2021, 47(8): 912-924. DOI: [10.11936/bjtxb2021010013](https://doi.org/10.11936/bjtxb2021010013).
JI Q, SUN Y F, HU Y L, et al. Review of clustering with deep learning [J]. *Journal of Beijing university of technology*, 2021, 47(8): 912-924. DOI: [10.11936/bjtxb2021010013](https://doi.org/10.11936/bjtxb2021010013).
- [18] 李君卫, 汤亚芳, 郝正航, 等. 聚类分析及其在电力系统中的应用综述 [J]. *现代电力*, 2019, 36(3): 1-10. DOI: [10.19725/j.cnki.1007-2322.20181130.001](https://doi.org/10.19725/j.cnki.1007-2322.20181130.001).
LI J W, TANG Y F, HAO Z H, et al. Survey of cluster analysis and its application in power system [J]. *Modern electric power*, 2019, 36(3): 1-10. DOI: [10.19725/j.cnki.1007-2322.20181130.001](https://doi.org/10.19725/j.cnki.1007-2322.20181130.001).
- [19] HARTIGAN J A, WONG M A. A K-Means clustering algorithm [J]. *Journal of the royal statistical society series C: applied statistics*, 1979, 28(1): 100-108. DOI: [10.2307/2346830](https://doi.org/10.2307/2346830).
- [20] 王仲, 顾煜炯, 韩旭东, 等. 基于历史数据聚类的火电机组工况划分 [J]. *仪器仪表学报*, 2019, 40(2): 90-95. DOI: [10.19650/j.cnki.cjsi.J1804232](https://doi.org/10.19650/j.cnki.cjsi.J1804232).
WANG Z, GU Y J, HAN X D, et al. Operating condition classification of thermal power unit based on historical data clustering [J]. *Chinese journal of scientific instrument*, 2019, 40(2): 90-95. DOI: [10.19650/j.cnki.cjsi.J1804232](https://doi.org/10.19650/j.cnki.cjsi.J1804232).
- [21] CHAUDHURI D, CHAUDHURI B B. A novel multiseed nonhierarchical data clustering technique [J]. *IEEE transactions on systems, man, and cybernetics, part B (cybernetics)*, 1997, 27(5): 871-876. DOI: [10.1109/3477.623240](https://doi.org/10.1109/3477.623240).

- [22] 张锐, 王义武, 朱啸龙, 等. 基于 UPGMA 的优化初始中心 K-means 算法研究 [J]. *计算机技术与发展*, 2018, 28(2): 50-53,58. DOI: [10.3969/j.issn.1673-629X.2018.02.012](https://doi.org/10.3969/j.issn.1673-629X.2018.02.012).
ZHANG R, WANG Y W, ZHU X L, et al. Research on K-means algorithm for optimizing initial center based on UPGMA [J]. *Computer technology and development*, 2018, 28(2): 50-53,58. DOI: [10.3969/j.issn.1673-629X.2018.02.012](https://doi.org/10.3969/j.issn.1673-629X.2018.02.012).
- [23] 赵帅, 顾煜炯, 李欣, 等. 供热机组多元工况下能效状态基准研究 [J]. *热力发电*, 2022, 51(4): 88-95 DOI: [10.19666/j.rfd.202108151](https://doi.org/10.19666/j.rfd.202108151).
ZHAO S, GU Y J, LI X, et al. Research on energy efficiency state benchmark of heating units under multiple operating conditions [J]. *Thermal power generation*, 2022, 51(4): 88-95. DOI: [10.19666/j.rfd.202108151](https://doi.org/10.19666/j.rfd.202108151).
- [24] 陈梅, 陈熙鹏, 陈薇, 等. 基于融合聚类的篦冷机工况划分与识别研究 [J]. *控制工程*, 2023, 30(2): 361-367. DOI: [10.14107/j.cnki.kzgc.20200861](https://doi.org/10.14107/j.cnki.kzgc.20200861).
CHEN M, CHEN X P, CHEN W, et al. Research on division and recognition of grate cooler operating conditions based on fusion clustering [J]. *Control engineering of China*, 2023, 30(2): 361-367. DOI: [10.14107/j.cnki.kzgc.20200861](https://doi.org/10.14107/j.cnki.kzgc.20200861).
- [25] HALLAC D, VARE S, BOYD S, et al. Toeplitz inverse covariance-based clustering of multivariate time series data [C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, Canada, August 04, 2017. Halifax: ACM, 2017: 215-223. DOI: [10.1145/3097983.3098060](https://doi.org/10.1145/3097983.3098060).
- [26] 肖钊, 邓杰文, 刘晓明, 等. 基于运行规律和 TICC 算法的风电 SCADA 高维时序数据聚类方法 [J]. *机械工程学报*, 2022, 58(23): 196-207. DOI: [10.3901/JME.2022.23.196](https://doi.org/10.3901/JME.2022.23.196).
XIAO Z, DENG J W, LIU X M, et al. Clustering method of high-dimensional time series SCADA data from wind turbines based on operational laws and TICC algorithm [J]. *Journal of mechanical engineering*, 2022, 58(23): 196-207. DOI: [10.3901/JME.2022.23.196](https://doi.org/10.3901/JME.2022.23.196).
- [27] 任荣荣, 周明全, 耿国华, 等. 基于深度神经网络的多尺度特征提取方法 [J]. *西北大学学报 (自然科学版)*, 2017, 47(2): 215-221. DOI: [10.16152/j.cnki.xdxbzr.2017-02-011](https://doi.org/10.16152/j.cnki.xdxbzr.2017-02-011).
REN R R, ZHOU M Q, GENG G H, et al. The multi-scale features extraction method based on deep neural network [J]. *Journal of Northwest University (natural science edition)*, 2017, 47(2): 215-221. DOI: [10.16152/j.cnki.xdxbzr.2017-02-011](https://doi.org/10.16152/j.cnki.xdxbzr.2017-02-011).
- [28] 李文华, 牛国波, 刘羽佳, 等. 基于多通道卷积神经网络的液压系统多故障诊断 [J]. *组合机床与自动化加工技术*, 2023(3): 91-94,98. DOI: [10.13462/j.cnki.mmtamt.2023.03.023](https://doi.org/10.13462/j.cnki.mmtamt.2023.03.023).
LI W H, NIU G B, LIU Y J, et al. Multi-fault diagnosis of hydraulic system based on multi-channel convolutional neural network [J]. *Modular machine tool & automatic manufacturing technique*, 2023(3): 91-94,98. DOI: [10.13462/j.cnki.mmtamt.2023.03.023](https://doi.org/10.13462/j.cnki.mmtamt.2023.03.023).
- [29] 惠子, 余立, 周欢, 等. 基于卷积神经网络算法的 X 射线晶体衍射实验数据筛选 [J]. *核技术*, 2023, 46(3): 030101. DOI: [10.11889/j.0253-3219.2023.hjs.46.030101](https://doi.org/10.11889/j.0253-3219.2023.hjs.46.030101).
HUI Z, YU L, ZHOU H, et al. X-ray crystallography experimental data screening based on convolutional neural network algorithms [J]. *Nuclear techniques*, 2023, 46(3): 030101. DOI: [10.11889/j.0253-3219.2023.hjs.46.030101](https://doi.org/10.11889/j.0253-3219.2023.hjs.46.030101).
- [30] 孙启涛, 罗智孙, 梁好, 等. 基于多尺度特征提取的风机音频信号故障诊断方法研究 [J]. *机电工程*, 2023, 40(1): 39-46. DOI: [10.3969/j.issn.1001-4551.2023.01.005](https://doi.org/10.3969/j.issn.1001-4551.2023.01.005).
SUN Q T, LUO Z S, LIANG H, et al. Fault diagnosis method of fan audio signal based on multi-scale feature extraction [J]. *Journal of mechanical & electrical engineering*, 2023, 40(1): 39-46. DOI: [10.3969/j.issn.1001-4551.2023.01.005](https://doi.org/10.3969/j.issn.1001-4551.2023.01.005).

 作者简介:


张大志

郑胜

1965, 男, 博士, 教授, 主要研究方向为人工智能技术, 核电异常检测。



崔文浩

张大志(第一作者)

1977-, 男, 博士, 总工程师, 主要研究方向核电仿真技术, 核电异常检测 (e-mail) zhangdz02@cnpn.com.cn。

崔文浩(通信作者)

1997, 男, 硕士, 主要研究方向为核电异常检测 (e-mail) 731096738@qq.com。

(编辑 徐嘉铨)