

基于 AHC-GP 混合模型的火电机组报警数据过滤方法

吴 锋^{1,2}, 张 悅^{1,2}, 董 泽^{1,2}, 封官斌³

(1. 华北电力大学 控制与计算机工程学院, 北京 102206; 2. 河北省发电过程仿真与优化控制技术创新中心, 河北 保定 071003; 3. 北京国华电力有限责任公司, 北京 100025)

摘要:针对火电机组设备工况复杂、报警系统效率低下及“报警泛滥”现象频发等问题,提出了基于 AHC-GP 混合模型的火电机组报警数据过滤方法,消除冗余性报警。首先,采用近邻传播算法(AP)结合类间类内划分指标(BWP)确定最佳聚类数目,再使用凝聚式层次聚类算法(AHC)进行聚类,区分各类复杂工况。其次,利用高斯过程模型(GP)结合后验报警概率估计值实现机组冗余性报警数据的准确过滤。最后,采用某电厂 1 000 MW 机组在 3 种典型故障下的实际主蒸汽温度、主蒸汽压力等报警数据作为实验数据集,验证所提方法有效性。结果表明:AHC-GP 混合模型相较于单一高斯过程模型,冗余性报警数据的过滤准确率提高了 10.7%,误判率降低了 50.1%,证明了模型的有效性;与支持向量机和梯度提升决策树等成熟算法相比,漏检率、误判率均较低,具有良好的报警数据过滤性能,可准确定位冗余性报警数据,减少“报警泛滥”问题的发生。

关键词:高斯过程;凝聚式层次聚类;近邻传播算法;报警泛滥;火电机组

中图分类号:TH81, TP181 文献标识码:A DOI:10.16146/j.cnki.rndlgc.2023.03.023

[引用本文格式]吴 锋, 张 悅, 董 泽, 等. 基于 AHC-GP 混合模型的火电机组报警数据过滤方法[J]. 热能动力工程, 2023, 38(3):174-182. WU Zheng, ZHANG Yue, DONG Ze, et al. Alarm data filtering method for thermal power units based on AHC-GP hybrid model[J]. Journal of Engineering for Thermal Energy and Power, 2023, 38(3):174-182.

Alarm Data Filtering Method for Thermal Power Units based on AHC-GP Hybrid Model

WU Zheng^{1,2}, ZHANG Yue^{1,2}, DONG Ze^{1,2}, FENG Guan-bin³

(1. School of Control and Computer Engineering, North China Electric Power University, Beijing, China, Post Code: 102206;
2. Hebei Technology Innovation Center of Simulation & Optimized Control for Power Generation, Baoding, China,
Post Code: 071003; 3. Beijing Guohua Electric Power Co., Ltd., Beijing, China, Post Code: 100025)

Abstract: To address the problems of complex equipment conditions, low efficiency of alarm system and frequent occurrence of "alarm flood" in thermal power units, an alarm data filtering method for thermal power units based on AHC-GP hybrid model was proposed to eliminate redundant alarms. Firstly, the optimal number of clusters was determined by using the nearest neighbor propagation algorithm (AP) combined with the inter-class and intra-class partitioning index (BWP), and then the agglomerative hierarchical clustering algorithm (AHC) was used for clustering to distinguish various complex operating conditions. Secondly, a Gaussian process model (GP) combined with a posteriori alarm probability estimate was used to achieve accurate filtering of unit redundancy alarm data. Finally, the actual main steam tem-

收稿日期:2022-02-24; 修订日期:2022-05-22

基金项目:河北省省级科技计划资助(22567643H);中央高校基本科研业务费专项资金(2018QN096);河北省自然科学基金(E2018502111)

Fund-supported Project: S & T Program of Hebei (22567643H); Fundamental Research Funds for the Central Universities (2018QN096); Natural Science Foundation of Hebei Province (E2018502111)

作者简介:吴 锋(1996-),男,华北电力大学博士研究生。

通讯作者:董 泽(1970-),男,华北电力大学教授。

perature and main steam pressure alarm data of a 1 000 MW unit in a power plant under three typical faults were used as the experimental data set to verify the effectiveness of the proposed method. The results show that the AHC-GP hybrid model improves the filtering accuracy of redundancy alarm data by 10.7% and reduces the misclassification rate by 50.1% compared with the single Gaussian process model, which proves the effectiveness of the model. Meanwhile, compared with the mature algorithms such as support vector machine and gradient boosting decision tree, it has a lower omission rate and misclassification rate as well as a better performance in filtering alarm data, which can accurately locate redundant alarm data and reduce the occurrence of "alarm flood" problem.

Key words: Gaussian process, agglomerative hierarchical clustering, nearest neighbor propagation algorithm, alarm flood, thermal power unit

引言

我国现阶段的能源结构不断调整,火力发电面临低排改造、新能源发电并入和偏离煤种运行等一系列新的困难和挑战,导致了机组在运行过程中常常出现波动,不可避免地产生了大量报警数据。在电厂集控中心,系统操作员通常需要处理大量实时报警信息,并对发电系统的运行做出重要决策。这些报警与设备故障、保护装置误操作等相关。由于火电机组的控制系统十分复杂,生产设备和参数数量众多,变量之间联系密切,同时现有的报警系统效率低下,“报警泛滥”问题十分常见^[1]。基于数据挖掘的报警数据过滤方法能够使现场操作人员及时消除冗余性报警,准确地定位重要报警数据,抑制报警溢流,减少异常传播引发的“报警泛滥”问题。现阶段的工业报警数据过滤技术已经被广泛应用于电力系统、铁路、航空航天、医学和化工等诸多领域。

目前,报警数据过滤技术主要分为两类,第1类是基于先验知识的报警数据过滤方法^[2-3],第2类是基于数据信息挖掘的报警数据过滤方法^[4-5]。文献[6]对工业报警系统的研究现状和导致报警泛滥问题的主要原因进行综述。文献[7]基于系统先验知识进行分析,提出了一种时间约束网络(TCN)的方法来分析电力系统的报警信息。文献[8]提出了一种分析二进制值报警变量之间因果关系的方法。归一化传递熵(NTE)和归一化直接传递熵(NDTE)被制定为基本统计指标,考虑随机延迟和警报发生之间的相互独立性,开发了一种改进的统计测试来确定 NTE 和 NDTE 的显著性阈值工业报警方法。

但上述报警数据过滤算法大多未考虑到工业过程数据中含有复杂工况的影响,不同工况数据间易发生误判,从而导致算法泛化性能较低,过滤精度不高。

凝聚型层次聚类算法(AHC)是一种常用的聚类算法。作为一种自底向上的层次聚类方法,算法将数据样本划分为初始簇,依据簇间度量距离对簇进行合并,直至满足聚类要求^[9]。凝聚型层次聚类算法以整个样本数据集为聚类中心,无需提前确定聚类中心,从而其聚类结果不会被局部最优解影响,对火电机组数据的工况聚类具有显著优势^[10]。高斯过程模型作为机器学习领域一个新的研究热点,具有强大的学习能力^[11]。高斯过程模型兼具贝叶斯推理学习和核机器学习的优点,模型先验知识直观,具有优良的模型泛化能力,可以拟合任意形式的数据。其中的协方差函数可大大降低模型中参数的计算难度,同时也可对概率信息进行输出,对于火电机组数据等复杂小样本数据具有良好的建模效果^[12]。

由于火电机组报警数据具有非线性、强耦合、高维度和时变性等特性,简单模型的报警数据过滤精度通常较差。同时,常规分级式报警策略和大多数报警过滤模型未考虑到机组复杂工况的影响,从而导致了消除冗余性报警性能不佳等问题。本文提出了凝聚式层次聚类与高斯过程相结合的混合模型(AHC-GP)对火电机组冗余性报警数据进行过滤,使用凝聚式层次聚类算法对报警数据进行聚类分析,区分不同复杂工况下数据,并针对不同工况下的数据,选用 RBF 作为高斯过程模型的核函数,结合后验报警概率估计值进行报警数据过滤。

1 凝聚式层次聚类描述

凝聚式层次聚类算法(AHC)将各类不规则数据看作一个独立的簇,对两个相近的簇进行合并,当满足所需的聚类结果后,合并终止。凝聚式层次聚类为自底向上的聚类方式,对不同复杂工况下的火电机组运行数据具有良好的聚类效果。凝聚式层次聚类算法流程如图1所示。

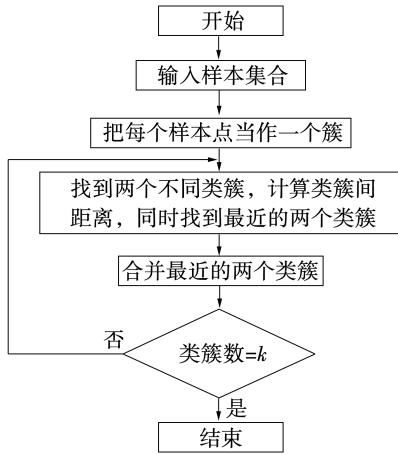


图1 凝聚层次聚类流程图

Fig.1 Flow chart of agglomerated hierarchical clustering

层次聚类算法的重点在于选定合并或分裂点,簇间的距离相似度为合并或分裂点选择的最重要依据。一般对于簇间距离相似度的度量方法分为最小距离、最大距离、均值距离和平均距离等^[13]。各类距离度量方法如图2所示。

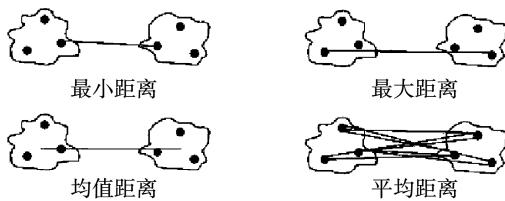


图2 各种距离度量方法示意图

Fig.2 Schematic diagram of various distance measurement methods

2 高斯过程模型描述

2.1 高斯过程模型

高斯过程模型(GPM)是基于贝叶斯概率框架下的无参数核方法,兼具贝叶斯推理学习和核机器学习的优点。高斯过程可通过一个随机过程 $f(x)$

的均值函数 $m(x)$ 和协方差函数(核函数) $K(x, x')$ 来表示:

$$f \sim GP(m(x), K(x, x')) \quad (1)$$

$$m(x) = E[f(x)] \quad (2)$$

$$K(x, x') = E[(f(x) - m(x))(f(x') - m(x'))] \quad (3)$$

式中: x —随机变量; $m(x)$ —均值函数; $K(x, x')$ —协方差函数。

在进行模型建立时,设置均值先验为零时高斯联合分布 $p(f | X) = N(0, K)$ 的一维高斯过程为:

$$p(f(x)) \sim GP(m(x)) = 0,$$

$$K(x, x') = \exp(-\frac{1}{2}(x - x')^2) \quad (4)$$

2.2 变分推断法

变分推断是近似推断中的一种常见方法,变分推断将后验分布推断求解问题转化为参数优化问题,使其具有良好的收敛和延展效果。本文基于研究对象及数据集特征,使用变分推断法对高斯过程模型参数进行高斯近似。

假设在监督学习中给定一组服从分布 p^* 的样本数据 (x_n, y_n) ,其中 $x_n \in R^p, y_n \in R, (n = 1, \dots, N)$ 。变分推断可以计算出与实际分布 p^* 相接近的近似分布 P_θ 。根据贝叶斯公式可知,当归一化系数 $p(x, y)$ 为常数,先验 $p(y)$ 为均匀分布时,似然项 $p(x | y, \theta)$ 与后验概率 $p(y | x, \theta)$ 成正比。因此,求解后验分布的问题转化为求解极大似然函数:

$$\theta^* = D_{KL}(p^* \| p_\theta) = \arg \max_\theta p_\theta(X | Y) \quad (5)$$

式中: $D_{KL}(p^* \| p_\theta)$ 为实际分布 p^* 与 p 的 KL 散度。参考引入 y_n ,引入隐变量 z_n ,则:

$$p_\theta(x) = \int_z p_\theta(x, z) dz \quad (6)$$

$$L(x) = \log p_\theta(x) = \log \int_z p_\theta(x, z) dz \quad (7)$$

使用期望最大算法对带有隐变量的最大似然函数进行求解:

$$L(x) \geq L(x) - D_{KL}(q_\varphi(z) \| p_\theta(z | x)) = E_{q_\varphi(x)} [\log p_\theta(x | z)] - D_{KL}(q_\varphi(z) \| p_\theta(z)) \quad (8)$$

$$F(\varphi, \theta) = E_{q_\varphi(x)} [\log p_\theta(x | z)] - D_{KL}(q_\varphi(z) \| p_\theta(z)) \quad (9)$$

由于式中含有包括似然函数的期望,因此在最大化给定隐变量 z_n 时,可以生成似然项的估计。同时,由于公式后半部分的 KL 散度恒不小于零,所以

最大化 $F(\varphi, \theta)$ 的问题可转化为 KL 散度取值最小值, 实现使 $q_\varphi(z)$ 与先验分布 $p_\theta(z)$ 接近, 从而将求解后验分布转化为一个优化问题。

3 基于 AHC-GP 混合模型的报警数据过滤方法

3.1 相关定义

3.1.1 近邻传播算法及 BWP 指标

近邻传播算法(AP)^[14]对高维、多类别数据的快速聚类具有良好的效果, 也可用于确定模型最佳聚类数目。

本文采用近邻传播算法(AP)进行最佳聚类数目的确定。令 $K = \{X, R\}$ 为聚类空间, 其中 $X = \{x_1, \dots, x_n\}$, 假设 n 个数据样本对象被聚类为 c 类, 类间类内划分指标(BWP)定义为第 j 类的第 i 个样本的聚类离差距离和聚类距离的比值, 表示为 $BWP_d(j, i)$:

$$\begin{aligned} BWP_d(j, i) &= \frac{\text{bswd}(j, i)}{\text{bawd}(j, i)} = \frac{\text{bd}(j, i) - \text{wd}(j, i)}{\text{bd}(j, i) + \text{wd}(j, i)} \\ &= \frac{\min_{1 \leq k \leq c, k \neq j} \left(\frac{1}{n_k} \sum_{p=1}^{n_k} \|x_p^{(k)} - x_i^{(j)}\| \right) -}{\min_{1 \leq k \leq c, k \neq j} \left(\frac{1}{n_k} \sum_{p=1}^{n_k} \|x_p^{(k)} - x_i^{(j)}\| \right) +} \rightarrow \\ &\leftarrow \frac{\frac{1}{n_j - 1} \sum_{q=1, q \neq i}^{n_j} \|x_q^{(j)} - x_i^{(j)}\|}{\frac{1}{n_j - 1} \sum_{q=1, q \neq i}^{n_j} \|x_q^{(j)} - x_i^{(j)}\|} \quad (10) \end{aligned}$$

BWP 指标反映了单个样本集的聚类精确程度。通过求取报警数据集中所有样本的 BWP 指标平均值来代表其聚类效果, 其平均值越大, 说明该数据集的聚类效果越好, 其最大值对应的聚类数即为最佳聚类数目。

$$\text{avgBWP}(k) = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} BWP(j, i) \quad (11)$$

$$k_{\text{opt}} = \arg \max_{2 \leq k < n} \{ \text{avgBWP}(k) \} \quad (12)$$

式中: $\text{avgBWP}(k)$ — 数据集聚成 k 类时的平均 BWP 指标值; k_{opt} — 最佳聚类数目。

3.1.2 评价指标

火电机组报警数据过滤可看作一类特殊的分类问题, 即在大量报警数据中过滤出冗余性报警数据。高斯过程二分类的原理为: 建立一个预测函数模型, 对于任意输入数据, 有效地找到其类别标签 y , 其中

$y = \{-1, 1\}$ 。本方法定义后验报警概率估计概念, 通过概率估计值进行两类报警数据的区分过滤:

$$p(y_* = \{-1, 1\} | X, y, x_*) \quad (13)$$

式中: X — 训练集样本; y, y_* — 一类别标签; $1, -1$ — 冗余报警数据类别和有效报警数据类别; x_* — 无标签测试集样本; p — 报警概率估计。

同时, 采用准确率、误判率和漏检率对模型的冗余性报警数据过滤效果进行描述, 具体评价指标及定义为:

$$\text{准确率} = \frac{\text{正确过滤个数}}{\text{真实冗余报警数据个数}} \quad (14)$$

$$\text{漏检率} = \frac{\text{漏检个数}}{\text{真实冗余报警数据个数}} \quad (15)$$

$$\text{误判率} = \frac{\text{误判个数}}{\text{真实冗余报警数据个数}} \quad (16)$$

3.2 算法描述

输入: 训练数据集 $x = \{x^1, \dots, x^N\}$ 、训练数据集类标签 t 、时间数据 h 、搜索范围 $[n_{\min}, n_{\max}]$ 、测试集、核函数和超参数。

输出: 最佳聚类数、准确率、误检率、漏检率、后验报警概率估计和冗余报警数据位置。

训练过程:

步骤 1: 输入聚类数搜索范围 $[n_{\min}, n_{\max}]$, 采用近邻传播聚类算法, 计算各点的 BWP 指标值及数据集整体平均 BWP 指标值, 确定最佳聚类数目 n 。

步骤 2: 添加时间标签约束。将时间数据 h 加入数据集中 $x = \{x^1, \dots, x^N, x^{N+1}\}$, 再利用 PCA 进行降维处理, 得出新数据集 $x' = \{x'^1, \dots, x'^N\}$ 。

步骤 3: 采用最佳聚类数目, 重复步骤 4 ~ 步骤 7 对数据集聚类。

步骤 4: 计算每个类别和总的 ESS 值。

$$\text{ESS} = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \quad (17)$$

式中: ESS — 误差平方和。

步骤 5: 枚举所有二项聚类, 计算合并后总 ESS 值。

步骤 6: 选择总 ESS 值最小的两个聚类进行合并。

$$TD_{c_1 \cup c_2} = \sum_{x \in c_1 \cup c_2} D(x, \mu_{c_1 \cup c_2})^2 \quad (18)$$

式中: $c_i (i=1, 2)$ — 聚类集合; $TD_{c_1 \cup c_2}$ — 合并后聚类。

步骤 7: 循环步骤 3 ~ 步骤 5, 直至类别满足要求。

步骤 8: 对每一聚类 $n \in N$, 重复步骤 9~步骤 10。

步骤 9: 选用 RBF 核函数, 设定先验均值为 0, 采用高斯过程分类算法训练。

$$K(x, x') = \exp\left\{-\frac{|x - x'|^2}{2^2}\right\} \quad (19)$$

步骤 10: 采用变分推断法获得高斯近似, 使用 logistic sigmoid 函数进行离散变换, 并计算后验报警概率估计:

$$p(t_{N+1} = 1 | x_{N+1}, t_N) = \int p[t_{N+1} = 1 | f(x_{N+1})] \quad (20)$$

$$p[(x_{N+1}) | X_{N+1}, t_N] df(x_{N+1}) \quad (21)$$

式中: x_{N+1} — 数据样本点; t_{N+1} — 类别变量; $P(t_{N+1} | x, t)$ — 后验概率分布。

步骤 11: 输入测试集。输出准确率、误检率、漏检率、后验报警概率值和冗余性报警数据位置。

AHC-GP 算法整体流程如图 3 所示。

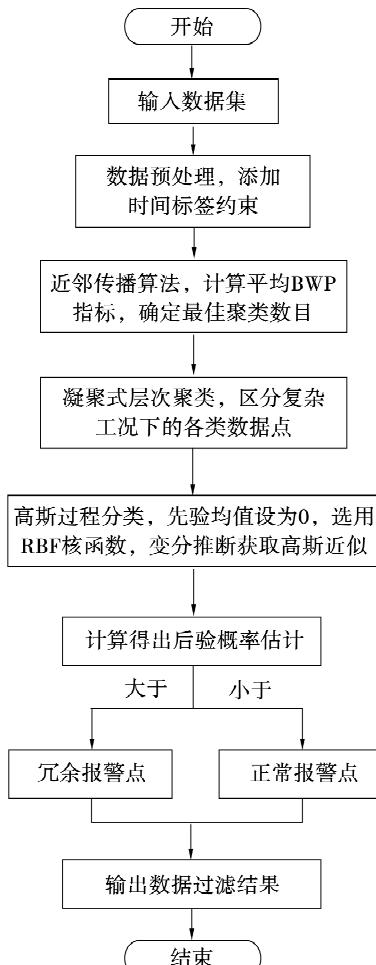


图 3 AHC-GP 模型流程图

Fig. 3 Flow chart of AHC-GP model

4 仿真实验分析

4.1 实验数据集

采用某电厂 1 000 MW 机组在某实际运行时间段中的火电机组报警数据组成仿真数据集, 包含有热偏差、减温水异常和烟道二次燃烧等典型故障下的正常报警数据和其他冗余性报警数据。仿真数据集由主汽温、主汽压等共计 5 000 组数据组成。训练数据集总数据量为 3 000 组, 测试数据集总数据量为 2 000 组。仿真数据集如图 4 所示, 数据集中各类故障报警和冗余性报警数据点数量如表 1 所示。

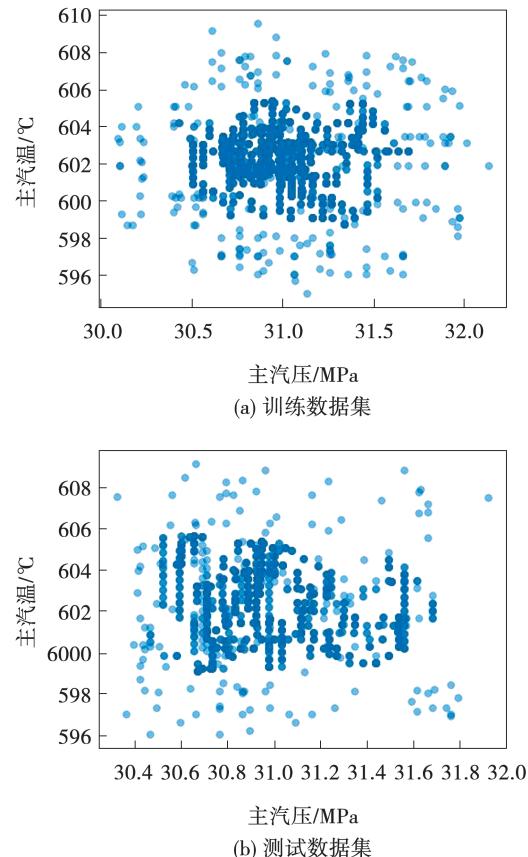


图 4 仿真数据集图像

Fig. 4 Simulation data set image

表 1 各类故障报警点与冗余性报警点数量表

Tab. 1 Table of the number of fault alarm points and redundancy alarm points of each type

数据集	热偏差 异常	减温水 异常	烟道二 次燃烧	其他冗余 性报警
训练数据集	921	1 079	787	213
测试数据集	599	739	562	100

4.2 实验结果

AHC-GP 模型通过近邻传播算法(AP)计算各点的 BWP 指标值,进一步得出各聚类数目下所有样本点平均指标值,从而搜索出最佳聚类数目。各聚类数目下的 BWP 指标值分布如图 5 所示。由图 5 可知,当聚类数目为 4 时平均 BWP 指标值最大,表明其聚类精确度最高,为最佳聚类数目。

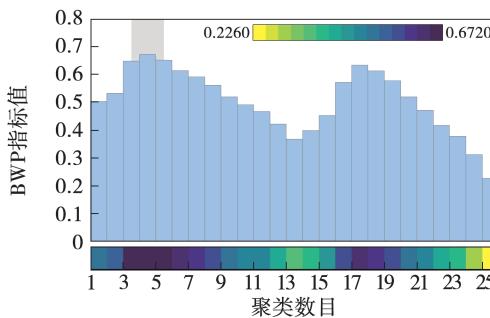


图 5 各聚类数目 BWP 指标值图像

Fig. 5 Image of BWP index values for the number of clustering

考虑到火电机组运行数据为具有强时间关联性的时间序列数据,故对数据集添加时间维度约束,增强模型聚类准确性,再采用凝聚式层次聚类算法进行聚类,区分各类工况,防止不同工况下数据相互掺杂影响报警数据过滤结果。AHC-GP 模型的聚类结果和类别变换图像如图 6 所示。

由图 6(a)、图 6(c)可知,如直接使用凝聚式层次聚类算法进行聚类,模型仅考虑到各点数值及聚类距离,未考虑机组运行数据的工程特点,即时间维度上的连续性。数据聚类结果在各类别间反复跳动,不满足现场实际火电机组运行情况。由图 6(b)、图 6(d)可知,对数据添加时间标签约束后,数据不再随时间在各工况间反复跳动,过程数据在时间维度上的连续性清晰地体现了出来,符合实际机组数据运行特点,极大地提高了模型聚类的准确性。使用 RBF(高斯核函数)作为建立高斯过程分类模型,通过网格搜索方法确定最佳模型参数,使其达到最佳检测效果。各工况下测试数据集的决策边界如图 7 所示。其中蓝色点为冗余报警数据点,橙色点为正常报警数据点,红蓝色背景交接处即为决策边界。图中,红色标记点为漏检点,绿色标记点为误检点。由上图所见,图 7(a)存在 1 个漏检点、无误检点。

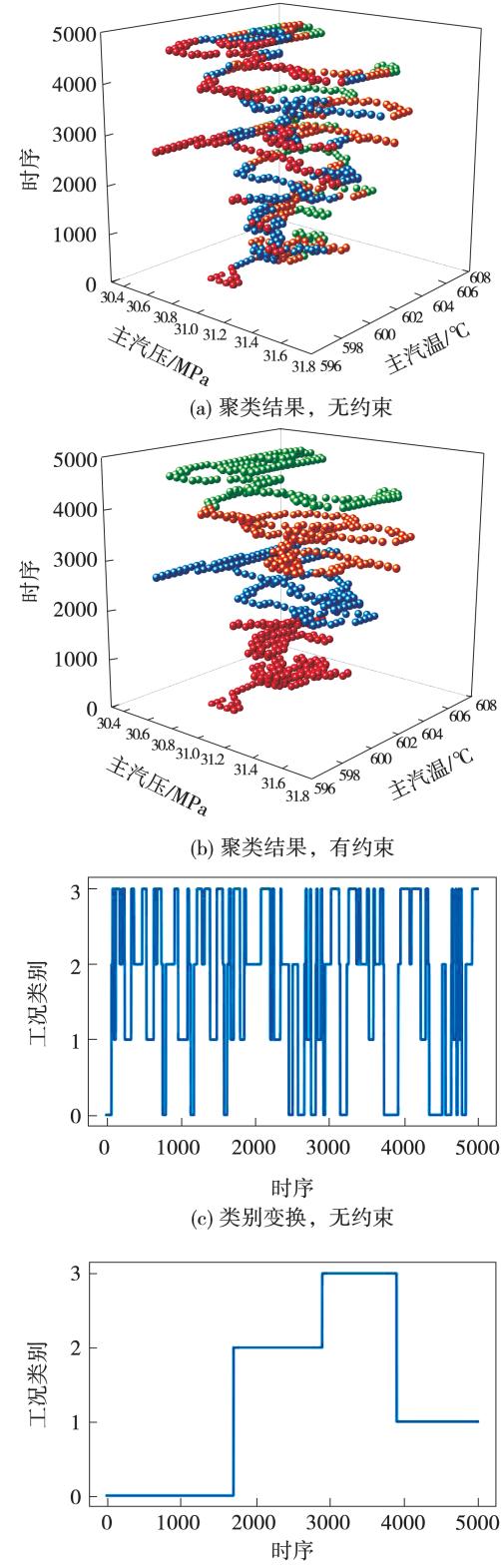


图 6 模型聚类结果及类别变换图像

Fig. 6 Model clustering results and class transformation image

图中,红色标记点为漏检点,绿色标记点为误检点。由上图所见,图 7(a)存在 1 个漏检点、无误检点。

点,图 7(b)存在 2 个误检点、无漏检点,图 7(c)存在 2 个漏检点、无误检点,图 7(d)无漏检点和误检点。经过模型参数的调整与优化,4 类测试数据集上的决策边界整齐,无过拟合和欠拟合状况。

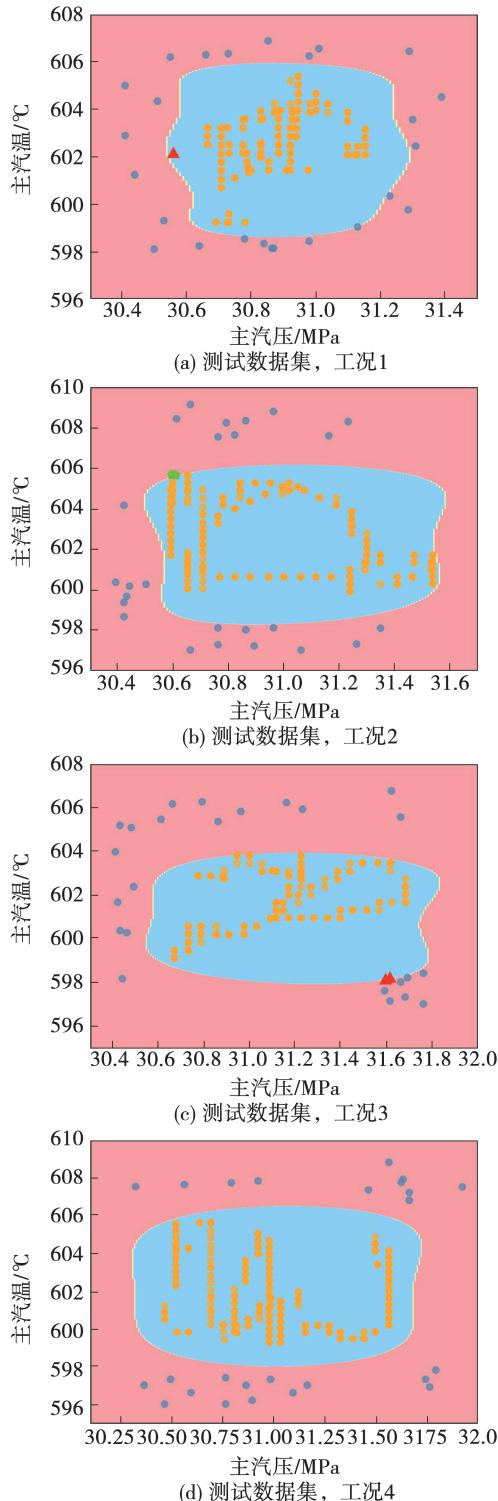


图 7 各工况下决策边界图像

Fig. 7 Decision boundary image under various working conditions

各工况下的后验报警概率估计值如图 8 所示。

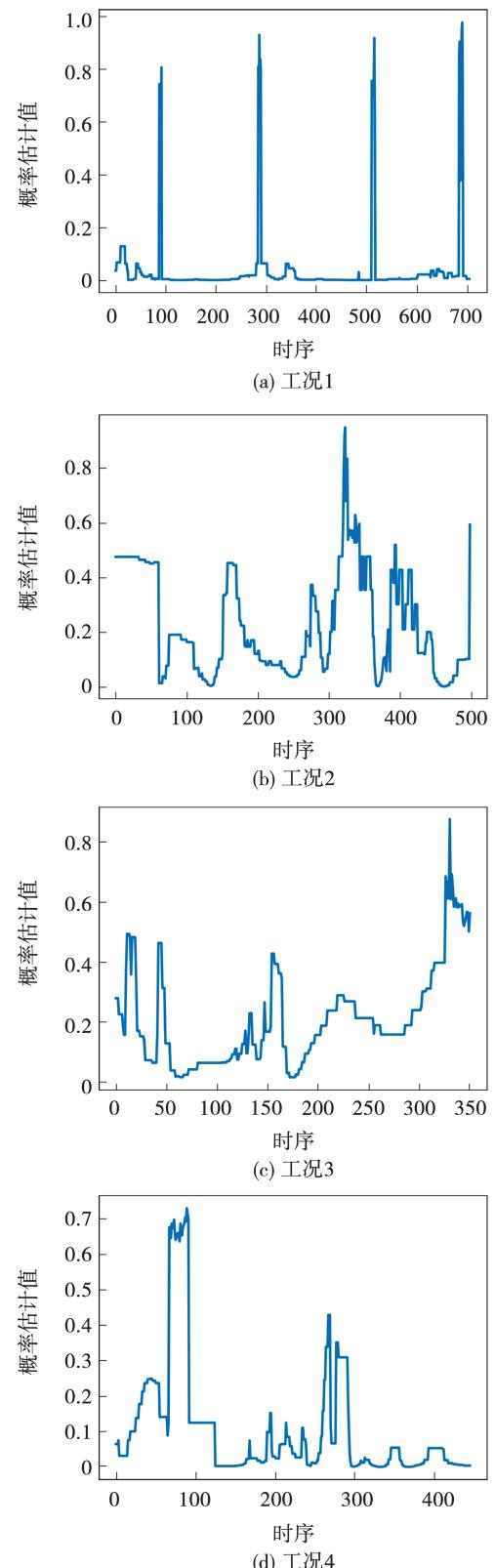


图 8 各工况下报警概率估计值图像

Fig. 8 Image of estimated alarm probability under various working conditions

由图 8 可知,各故障下正常报警数据点的概率估计值均较小,但当存在冗余报警数据时,该点的概率估计值较大。模型可结合各点后验报警概率估计值的大小对冗余性报警数据点进行筛选过滤。

分别对各工况下的冗余性报警数据过滤准确性进行统计分析,各工况下的数据误检率和漏检率如表 2 所示。通过表 2 可知,模型的报警数据过滤性能良好,可对冗余性报警数据准确定位,从而减少“报警泛滥”现象的发生。

表 2 各工况检测结果

Tab. 2 Test result of each working condition

类别	工况 1	工况 2	工况 3	工况 4
漏检率	0.04	0	0.08	0
误判率	0	0.08	0	0

4.3 对比实验

为进一步验证方法的有效性,选用国电泰州二次再热机组 800 MW 工况下的主蒸汽温度、主蒸汽压力等 2 000 组运行数据,并选取现阶段数据过滤效果较好的成熟算法支持向量机分类(SVC)和梯度提升决策树(GBDT)等进行冗余报警数据过滤的准确率、漏检率、误检率和正确检出数目等各方面性能的对比实验。各算法的具体性能如图 9 所示。

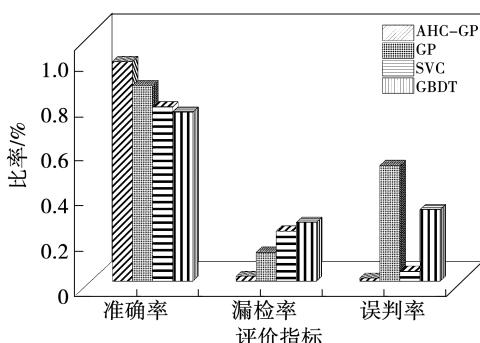


图 9 不同算法冗余报警数据过滤性能对比

Fig. 9 Comparison of redundant alarm data filtering performance of different algorithms

根据对比结果,AHC-GP 模型与其他现阶段数据过滤效果较好的支持向量机分类(SVC)和梯度提升决策树(GBDT)等成熟算法相比,在冗余报警点

的检测准确率、漏检率、误检率和正确检出数目等各方面表现均优于对比算法。同时,使用单一 GP 模型会导致大量的数据误判,采用凝聚式层次聚类与高斯过程模型相结合的方式使冗余报警数据过滤的准确率提高了 10.7%,误判率降低了 50.1%,模型整体具有较高的报警数据过滤性能。

5 结 论

(1) 针对常规分级式报警策略和大多数报警过滤模型未考虑到火电机组各类复杂工况的影响、导致对冗余性报警过滤性能不佳等问题,提出了一种基于凝聚式层次聚类与高斯过程相结合的混合模型(AHC-GP),对冗余性报警数据进行过滤。

(2) AHC-GP 混合模型采用近邻传播算法(AP)结合类间类内划分(BWP)指标对最佳聚类数目进行搜索,使用凝聚式层次聚类算法(AHC)进行各类复杂工况区分,再结合高斯过程模型(GP)后验报警概率估计值实现火电机组冗余性报警数据的准确过滤。

(3) 相较于单一高斯过程模型,AHC-GP 混合模型准确率提高了 10.7%,误判率降低了 50.1%。同时,冗余报警点的检测准确率、漏检率和误检率等各方面表现均优于支持向量机分类(SVC)和梯度提升决策树(GBDT)等成熟算法,具有良好的报警数据过滤性能,能够有效减少“报警泛滥”问题的发生。

参考文献:

- [1] WANG Jian-dong, YANG Fan, CHEN Tong-wen, et al. An overview of industrial alarm systems: main causes for alarm overloading, research status, and open problems[J]. IEEE Transactions on Automation Science and Engineering, 2015, 13(2):1–17.
- [2] XU Xue, LIU Qiang, DING Jin-liang. A modified dynamic PLS for quality related monitoring of fractionation processes[J]. IFAC-PapersOnLine, 2018, 51(7):315–320.
- [3] LU Qiu-gang, JIANG Ben-ben, GOPALUNI R B, et al. Locality preserving discriminative canonical variate analysis for fault diagnosis[J]. Computers & Chemical Engineering, 2018, 117(2):309–319.

- [4] CHEN Guang-jie, GE Zhi-qiang. Hierarchical Bayesian network modeling framework for large-scale process monitoring and decision making [J]. IEEE Transactions on Control Systems Technology, 2018, 28(2):1–9.
- [5] LAI Shi-qi, CHEN Tong-wen. A method for pattern mining in multiple alarm flood sequences [J]. Chemical Engineering Research and Design, 2017, 117(1):831–839.
- [6] GUO Wen-xin, WEN Fu-shuan, LIAO Zhi-wei, et al. An analytic model-based approach for power system alarm processing employing temporal constraint network [J]. IEEE Transactions on Power Delivery, 2010, 25(4):2435–2447.
- [7] HU Wen-kai, WANG Jian-dong, CHEN Tong-wen, et al. Cause-effect analysis of industrial alarm variables using transfer entropies [J]. Control Engineering Practice, 2017, 64(1):205–214.
- [8] WANG Jian-dong, YANG Zi-jiang, SU Jian-jun, et al. Root-cause analysis of occurring alarms in thermal power plants based on Bayesian networks [J]. International Journal of Electrical Power & Energy Systems, 2018, 103(1):67–74.
- [9] 刘翠玲, 李天瑞, 于重重. 凝聚层次聚类算法在电厂设备故障预测上的研究 [J]. 电气时代, 2017(7):96–99.
LIU Cui-ling, LI Tian-rui, YU Chong-chong. Research on cohesive hierarchical clustering algorithm for power plant equipment fault prediction [J]. Electric Age, 2017(7):96–99.
- [10] PRAVEEN P, RANJITH K M, SHAIK M A, et al. The comparative study on agglomerative hierarchical clustering using numerical data [J]. IOP Conference Series: Materials Science and Engineering, 2020, 981(2):1–6.
- [11] 熊志化, 张卫庆, 赵瑜, 等. 基于混合高斯过程的多模型热力参数测量软仪表 [J]. 中国电机工程学报, 2005, 25(7):30–33, 40.
XIONG Zhi-hua, ZHANG Wei-qing, ZHAO Yu, et al. Thermal parameter soft sensor based on the mixture of Gaussian processes [J]. Proceedings of the CSEE, 2005, 25(7):30–33, 40.
- [12] 贾昊. 基于高斯过程回归的热工系统数据驱动建模研究 [D]. 北京: 华北电力大学, 2020.
JIA Hao. Research on data-driven modeling of thermal systems based on Gaussian process regression [D]. Beijing: North China Electric Power University, 2020.
- [13] SRINIVAS M, MOHAN C K. Efficient clustering approach using incremental and hierarchical clustering methods [C]//The 2010 International Joint Conference on Neural Networks. Glasgow, UK: IEEE, 2010:1–7.
- [14] FREY B J, DUECK D. Clustering by passing messages between data points [J]. Science, 2007, 315(2):972–976.

(丛敏 编辑)