

详细设计说明书

名称	运行事件预警算法开发详细设计说明书
----	-------------------

共 24 页

编写单位	华中科技大学
编写人	
编写日期	
校对人	
校对日期	
批准人	
批准日期	
审核人	
审核日期	

目录

1.概述	3
2.引用文件	3
3.运行事件预警算法开发需求概述	4
4.运行事件预警算法开发总体设计	5
5.基于多模态 DPCA 的运行事件状态检测详细设计	6
5.1 针对正常工况单一模态的 DPCA 建模原理	6
5.2 基于多模态 DPCA 的异常检测方法	7
5.3 算法验证	8
5.3.1 数据集	8
5.3.2 结果展示	10
6.关键工况参数趋势预测详细设计	11
6.1 基于经验模态分解的趋势提取	12
6.2 基于自适应指数模型的关键工况参数趋势预测	15
6.2.1 自适应指数模型的算法原理	15
6.2.2 算法验证	17
6.3 基于自适应线性模型的关键工况参数趋势预测	19
6.3.1 自适应线性模型的算法原理	19
6.3.2 算法验证	20
6.4 基于自回归多步预测模型的关键参数趋势预测	21
6.4.1 自回归多步预测模型	22
6.4.2 算法验证	23

1.概述

运行事件预警算法是“运行事件分析应用”的模块应用功能之一，“运行事件分析应用”是项目“人工智能在核反应堆方面的应用(核反应堆智能运行支持关键技术研究)”的研究子课题之一，依托项目内的运行数据分析技术、核电运行数字孪生系统及其样机平台，开展针对“运行事件分析”的智能化技术研究，以期为核电厂的运行事件分析活动提供数字化技术支持。

核电站在正常运行过程中，可能会由于设备失效、仪控系统故障、电网故障、火灾、人员技能不足、人因失误等原因导致非预期的保护动作，例如甩负荷、停机、停堆等。当前机组运行时的异常情况发现主要依赖于 DCS 系统的报警信息，只有在到达阈值时才会触发相应报警，缺乏对趋势的分析；此外，虽然机组的典型异常事件在导致事故前大多数都有不同的先兆，但是 DCS 报警不能完全涵盖这些事件，也无法根据参数的关联性波动及时提示操纵员，所以在异常事件前期难以引起注意。

传统的监控方法基于单变量的控制图，已经无法反映复杂的工业过程的变化。现代过程监控是基于数据驱动的方法，数据驱动的量度是直接从过程数据中导出的。现代工业系统的大数据，虽然带来丰富的信息，但是工程师和操作人员无法根据这海量的数据从经验上得出过程运行的评估。借助数据驱动的过程监控方式，将高维数据变换为低维数据，并从中获取信息，然后计算出一些有意义的统计数字和生成直观的图表，捕捉过程变量之间相关关系的异常，从而告知操作人员及时应对。

本项目的研究目的是开发事件预警算法，使用不同算法监控电厂运行状态，建立早期预警及诊断模型。除停堆事件外，还根据多个核电厂事件调研情况，从电厂的事件报告和福清核电的机组“病例库”文件中，选取发生频率较高且对机组状态影响较大的若干个设备故障或失效事件作为研究对象：包括主冷却剂泵跳闸、主给水泵跳闸、GCT-c 旁排阀故障、CRF 循环水丧失 4 个场景进行预警分析。

2.引用文件

本详细设计说明书的编制参考下列资料：

- 核反应堆系统运行数字孪生应用系统需求分析说明书；
- 运行数字孪生专用人工智能引擎需求分析说明书；
- 核反应堆智能运行支持关键技术研究（人工智能在核反应堆方面的应用子课题一）运行事件预警算法需求分析说明书；

3.运行事件预警算法开发需求概述

本项目的研究目的是开发运行事件预警算法，对运行过程的异常进行监测、分析、预警，分别解决包括检测当前运行状态正常与否以及预测该异常事件发生时间在内的问题，为保障电厂平稳安全运行提供辅助支持。

预警算法将监测电厂实时运行状态，实现对停堆及若干个小事件的提前预警，具体事件包括：

- 1) 主冷却剂泵跳闸：在机组稳态运行过程中，如果一台主冷却剂泵故障跳闸，在30%FP 功率水平以上将直接触发反应堆跳闸，且一回路其中一个环路失去强迫循环，热量导出也会受到一定影响；对机组控制要求比较高；
- 2) 主给水泵跳闸：在正常运行过程中，两台主给水泵跳闸将触发机组快速甩负荷，引起一二回路巨大扰动；若三台主给水泵失效，将引发主给水丧失，对一回路热量导出带来巨大风险；
- 3) GCT-c 旁排阀故障：GCT-c 旁排阀门在跳机或者甩负荷以及启停工况下投入，分担一部分汽轮机负荷，排出多余的蒸汽，以保证一回路热量的正常导出，在多个核电机组中，因该阀门长期处于高温高压运行工况，且动作频繁，较容易出现故障，且该阀门故障后将直接影响蒸汽排放流量，从而引起一回路过冷或过热，影响较大；
- 4) CRF 循环水泵跳闸：循环水泵用于为冷凝器提供冷却海水，承担最终热阱功能，当一台循环水泵跳闸后，冷凝器失去一列冷却，真空度将恶化，若不手动快速降负荷，将引发汽轮机跳闸，若两台循环水泵跳闸，二回路将失去最终冷却，最终导致汽轮机跳闸和反应堆停运。

预警算法针对核电厂重大运行事件研发，将达到识别停堆及四种场景相关设备参数的异常变化，并能正确预测可能发生的事件及预计发生的时间。首先对事件相关参数进行异常检测，根据不同事件划分输入参数范围，采用数据处理、特征提取及多元统计分析、动态主成分分析等算法开展停堆事件及四个场景异常检测的研究，然后引入趋势预测，用于确定异常事件将在多长时间后发生，它依据历史电厂运行数据及参数间关联性进行回归建模，对特定事件的相关参数未来一段时间内的运行趋势进行预测，最后将信息合并输出为预警信息。

4.运行事件预警算法开发总体设计

该算法总体流程图如下：

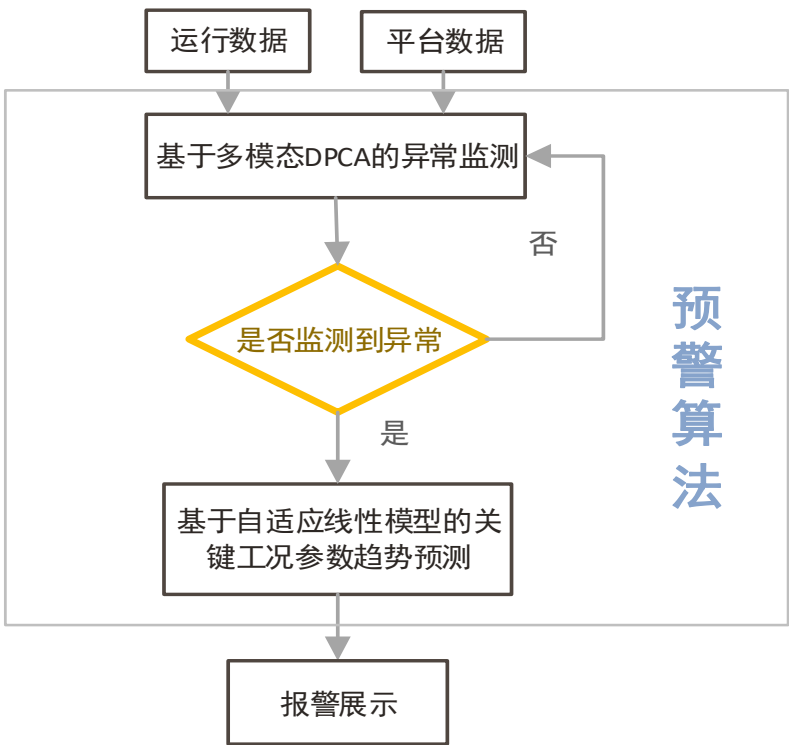


图 1 停堆及典型事件预警功能框架图

预警算法从“智能运行支持系统的数据库平台”获取机组实时运行数据或平台孪生系统数据作为输入，经过预警计算的结果进行显示。

预警算法包含两部分，首先利用统计过程监控中基于多模态 DPCA 的方法确定电厂系统整体状态，如果发现异常，再利用基于自适应线性模型的关键参数预测发生问题的时间，两种方法共同运行，最终得出结论将结果送系统最终显示。算法开发完成后，将运行在核电运行数字孪生系统中，实时预警当前状态是否异常、以及可能的发生时间。

在接下来的部分里将详细介绍电厂设备的异常状态监测算法模块和关键工况参数趋势预测算法模块，其中，在关键工况参数趋势预测算法的研究中，提出并测试了两种算法方案，对比之下，基于自适应线性模型的关键工况参数趋势预测算法对停堆事件发生时间能给更为准确的估计，故在参数趋势预测模块中采用此算法。

5.基于多模态 DPCA 的运行事件状态检测详细设计

当核电厂出现设备故障或失效引发的异常工况时，在线监测数据会偏离正常工况下的行为模式。数据驱动的运行事件状态检测能够通过多维数据之间的关联关系实现更高灵敏度的异常工况报警，为核电厂的安全运行和经济运行提供支持。

其基本思想是：在离线建模阶段，利用正常工况历史数据建立正常工况的监测模型，描述正常工况下多维监测变量之间的时空关联关系；在在线监测阶段，利用上述监测模型产生在线数据对应的残差信号，以指示在线数据偏离正常工况的程度，由此确定异常事件报警。由于核反应堆在正常工况下存在多种工作模态，每一工作模态下多维监测变量之间的时空关联关系不同，因此本项目采用基于多模态数据驱动建模的异常检测方法。

5.1 针对正常工况单一模态的 DPCA 建模原理

本项目采用 DPCA (Dynamic Principal Component Analysis) 模型描述单一模态下多维监测变量之间的时空关联特性，其建模的具体方法如下。

获取反应堆中样本数据矩阵 $X = [X(k)^T X(k-1)^T \cdots X(k-l+1)^T]^T \in R^{md \times l}$ ，其中， l 为时间窗口长度， $X(k) = [x(k) x(k-1) \cdots x(k-d+1)]$ 是带有时延的数据向量，其中， d 是时延， $x(k) \in R^{1 \times m}$ ， m 为传感器变量个数。表示通过奇异值分解(SVD)算法，可以将矩阵 X 分解为分数矩阵 T 和加载矩阵 P ，即

$$\begin{aligned} X &= TP^T + \tilde{X} = TP^T + \tilde{T}\tilde{P}^T \\ &= [T \quad \tilde{T}][P \quad \tilde{P}]^T \equiv \bar{T}\bar{P}^T \end{aligned} \quad (1)$$

其中， $\tilde{X} = \tilde{T}\tilde{P}^T$ 为残差矩阵， $\bar{T} = [T \quad \tilde{T}]$ ， $\bar{P} = [P \quad \tilde{P}]^T$ 。由于 \bar{T} 的列是正交的，所以协方差矩阵是

$$S \approx \frac{1}{N-1} X^T X = \bar{P} \bar{\Lambda} \bar{P}^T$$

其中，

$$\bar{\Lambda} = \frac{1}{N-1} \bar{T}^T \bar{T} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m\}$$

当 N 很大时， λ_i 是协方差矩阵的按降序排列的特征值。主成分子空间 (Principal Component Subspace, 简称 PCS) 为 $S_p = \text{span}\{P\}$ ，残差子空间 (Residual Subspace, 简称 RS) 为 $S_r = \text{span}\{\tilde{P}\}$ 。

一个样本向量 x 可以分别通过以下式子投影到 PCS 和 RS 上：

$$\begin{aligned} \hat{x} &= PP^T x \equiv Cx \in S_p \\ \tilde{x} &= \tilde{P}\tilde{P}^T x = (I - PP^T)x = (I - C)x \in S_r \end{aligned} \quad (2)$$

由于 S_p 和 S_r 是正交的， $\hat{x}^T \tilde{x} = 0$ 并且 $x = \hat{x} + \tilde{x}$ 。

5.2 基于多模态 DPCA 的异常检测方法

由于正常工况包含多个工作模态多模态过程中，多维监测变量在各个模态下具有不同的时空关联特性，无法满足传统多元统计监测方法要求数据来自单一工况的假设。为此，本项目在第 4.1 节的基础上采用基于多模态 DPCA 的异常检测方法。

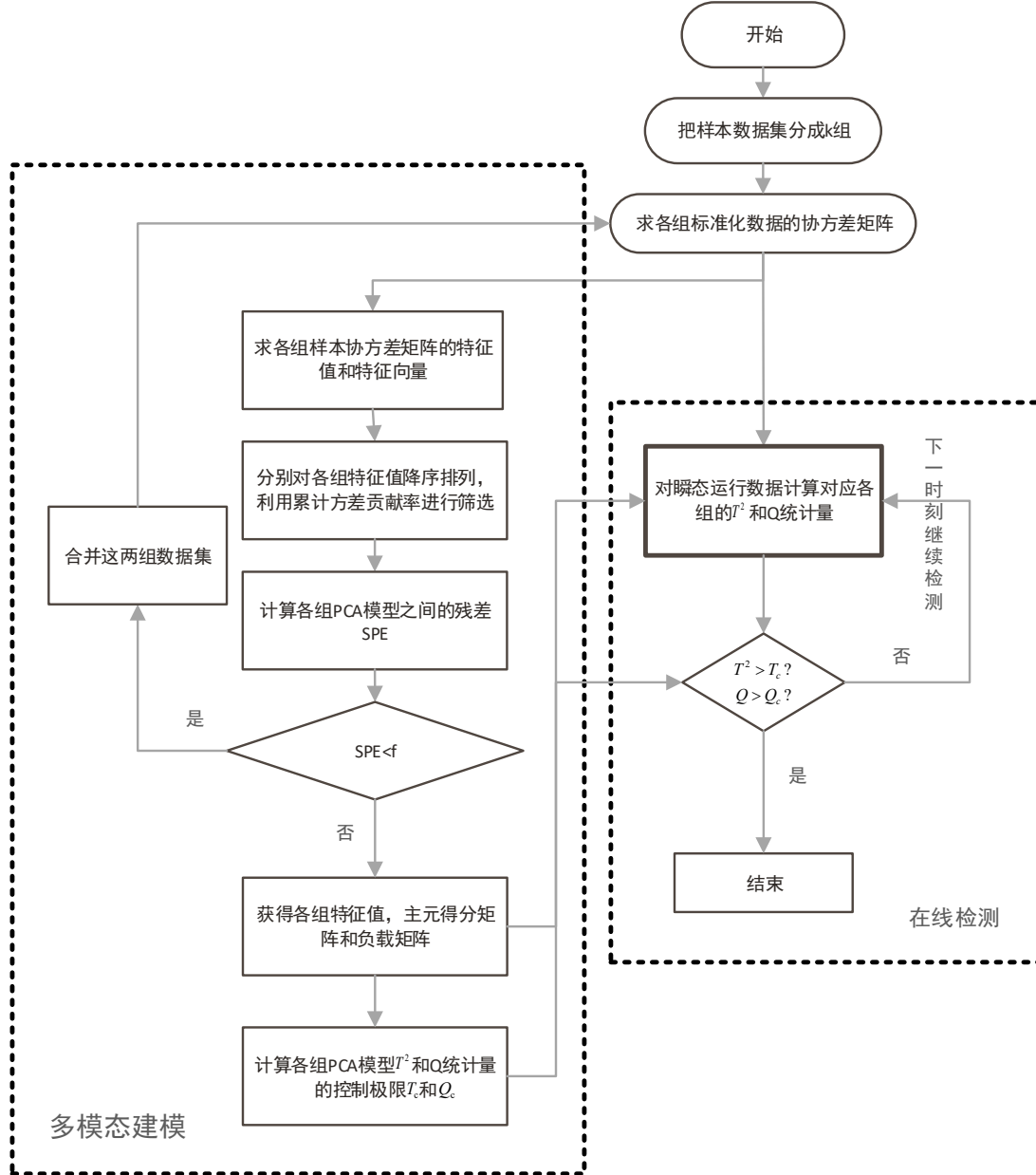


图 5-1 DPCA 故障检测流程

多模态 DPCA 建模不仅包括第 5.1 节所述的传统 DPCA 建模, 还包括模态的划分, 其具体方法如下:

- (1) 将 $X_1^{(i)}, \dots, X_{n_i}^{(i)}, i=0$ 表示为初始数据组, 其中上标 i 表示第 i 次分组, 下标表 n_i

示分组组数，通过以下迭代过程构建 DPCA 模型。

(2) 对于每一组 $X_k^{(i)}$ ，PCA 模型的 $P_k^{(i)}$ 通过传统的 DPCA 方法独立获得，保留的主成分的数量可以通过交叉验证来确定，并将测试数据集作为 $X_k^{(i)}, k=1, \dots, n_i$ 的一部分。

(3) 两两 DPCA 模型进行相似度比较，以残差作为相似度量：

$$SPE = \tilde{P}\tilde{P}^T x = (I - PP^T)x$$

(4) 若残差大于阈值 f ，则合并相应的数据组，将这些新生成的数据组表示为 $X_1^{(i)}, \dots, X_{n_i}^{(i)}$ ，转到步骤 (2)，否则停止。

(5) 确定检测阈值：阈值的取值为对应模态训练集的一倍均值和六倍方差之和。

在上述多模态 DPCA 模型的基础上，基于多模态 DPCA 的异常检测方法如下：

- (1) 当接收到在线数据时，通过式 (2) 求其与各个组模态的残差 SPE。
- (2) 检测逻辑：若至少有一组残差小于阈值，则为正常样本，否则为故障样本。

5.3 算法验证

5.3.1 数据集

我们采用表 5-1 中的数据集来验证我们的算法。使用大量正常数据集进行训练以达到误报率要求，并能够检测出在主冷却剂泵跳闸、主给水泵跳闸、GCT-c 旁排阀故障、CRF 循环水丧失、停堆事件等典型异常工况。其中故障工况样本为稳压器故障和下泄故障。选用表 5-2 所示的传感器变量。

表 5-1 验证数据集	
训练集数据	2018 年 6 月 3 日-2018 年 6 月 18 日
	2018 年 8 月 18 日-2018 年 10 月 4 日
	2019 年 1 月 14 日-2019 年 5 月 2 日
	2020 年 2 月 19 日-2020 年 4 月 1 日
测试集（正常工况样本）	2020 年 4 月 9 日-2020 年 6 月 15 日
测试集（故障工况样本）	仿真数据组 1
	仿真数据组 2

表 5-1 中数据包含了不同的稳态工况和过渡态工况下 12 个变量的正常工况下的数据和仿真故障数据。这些数据的采样时间为 1s。

表 5-2 稳压器液位关联参数表			
测点	描述	关系	目标参数

RCV018MD	上充流量	正相关	RCP007MN
RCV021MD	主泵 1 轴封注入流量	正相关	RCP008MN
RCV022MD	主泵 2 轴封注入流量	正相关	RCP011MN
RCV023MD	主泵 3 轴封注入流量	正相关	
RCV005MD	下泄流量	负相关	
RCP005MP	稳压器压力	负相关	
RCP006MP	稳压器压力	负相关	
RCP013MP	稳压器压力	负相关	
RCP611KM	1 环路平均温度	正相关	
RCP615KM	2 环路平均温度	正相关	
RCP619KM	3 环路平均温度	正相关	

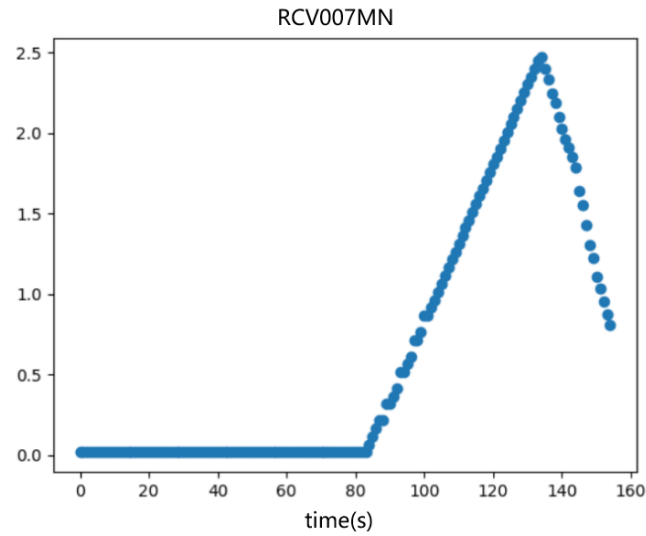


图 5-2 故障情况下目标变量的状态曲线

图 5-2 是仿真故障数据集 1 中目标变量 RCV007MN 的变化曲线，我们可以看到液位变量随着时间逐渐增长，在第 130 秒左右触发停堆条件。

正常工况下的数据具有多种模态，其中包括多种稳态和过渡态，如 RCP007MN 液位会在 -0.75m 和 0m 的不同模态下工作。针对不同模态的特点，我们采用能够识别不同模态的故障监测算法。

5.3.2 结果展示

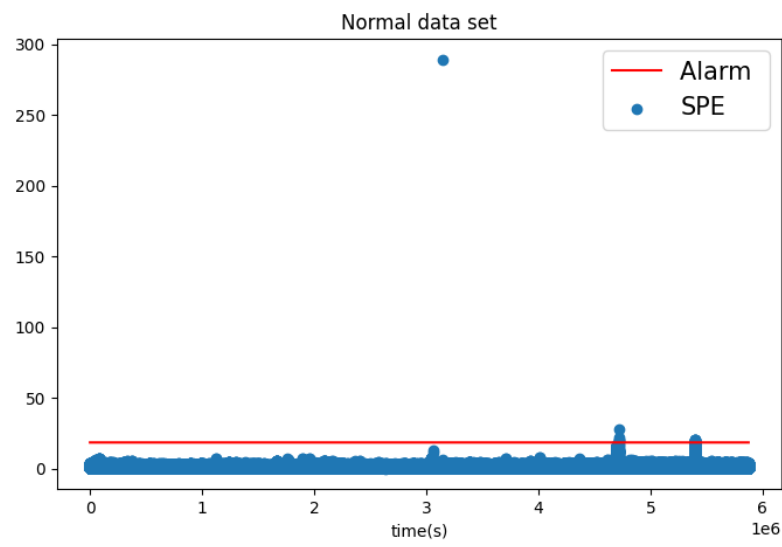
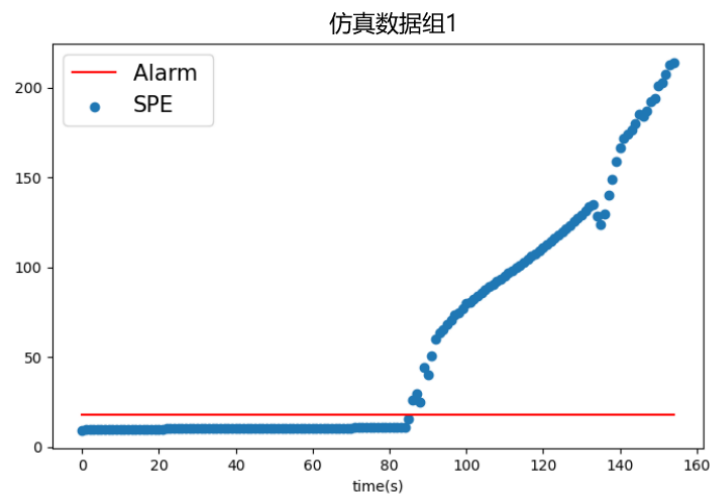


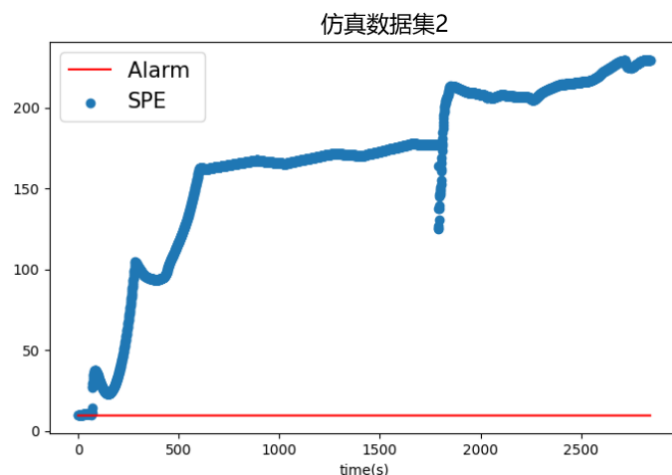
图 5-3 正常样本的故障监测

表 5-3 正常样本的误报率

样本名称	总样本数	误报样本数	误报率
Normal data set（2020 年 4 月 9 日-6 月 15 日）	5875201	25	小于 3%



(a)



(b)

图 5-4 故障样本的故障监测

表 5-4 故障样本的检测率

样本名称	总样本数	故障样本数	检测故障样本数	检测率
仿真数据组 1	155	68	68	100%
仿真数据组 2	2845	2845	2845	100%

图 5-3 和图 5-4 中，红色直线表示自适应报警阈值，蓝色的点表示每一时刻的残差值。其中正常工况样本的误报率小于 3%，仿真故障工况样本在出现停堆趋势时就被检测出来。

6.关键工况参数趋势预测详细设计

在检测到异常工况之后，对关键工况参数的变化趋势进行预测，进而预测自动停堆、停泵等事件的发生时间，对于异常工况报警后的及时维护具有重要的参考意义。异常工况趋势预测的难点在于历史数据中异常事件极为少见，也难以验证模拟机仿真生成的异常工况数据能否准确反映实际情况。为此，本项目在没有异常工况历史数据的条件下，利用异常工况报警后的在线数据实现关键工况参数趋势变化的在线自适应建模与预测。如图 6-1 所示，主要步骤包括趋势提取、在线建模与预测剩余未停堆时间。

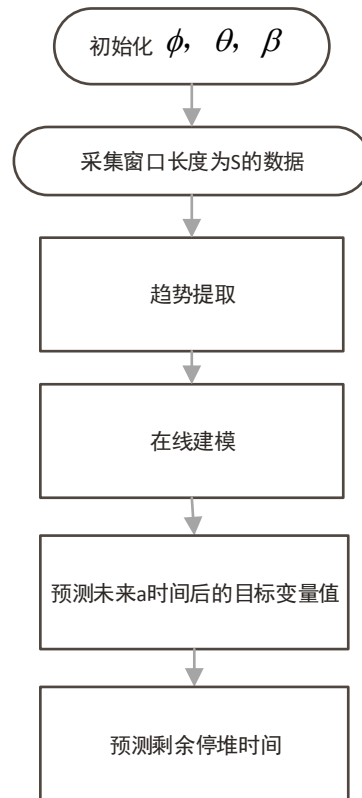


图 6-1 关键工况参数趋势预测的总体思路

6.1 基于经验模态分解的趋势提取

监测数据本身包含的噪声和局部振荡会干扰趋势预测。为此，本项目采用经验模态分解（Empirical mode decomposition，EMD）方法来对数据进行去噪和趋势提取。

基于 EMD 的希尔伯特黄变换（Hilbert-Huang transform，HHT）是分析非线性、非平稳信号的方法。HHT 包含两部分，即 EMD 和希尔伯特谱分析（Hilbert spectrum analysis，HAS）。它的特点是通过 EMD 对非线性、非平稳的数据进行线性和平稳化处理，得到固有模态函数（IMF-Intrinsic mode function）分量，并在分解的过程中保留了数据本身的特性。该分解方法从信号本身的尺度特征出发对信号进行分解，因而能尽快的分析出信号的特征，具有良好的局部适应性。

用 EMD 方法从原信号中提取的 IMF，突出了原信号的局部特征信息。IMF 具有如下特点

- 1）整个时间历程内，波形过零点的次数等于极值点数，或两者至多相差 1。
- 2）信号上任意一点，由局部极大值定义的上包络线和由局部极小值定义的下包络线的均值为 0。

在上述两个限制条件中，其中第一个条件其实和常用的平稳高斯过程的窄带要求十

分相似。第二个条件是一个新的限制条件，这里将过去常用的全局信号条件变成了部分信号条件，该条件使得信号可以进行分解和求取相关结果，同时分解出的多层 IMF 也是具有物理意义的结果。

但是并不是所有的信号都满足上述提到的两个 IMF 的要求，因为在任意时刻一般信号不止包括一种模态。

EMD 方法分解信号基于以下三条假定：

- 1) 信号具有至少一个极大值点和一个极小值点；
- 2) 特征时间尺度设置为相邻极值点之间的时间间隔；
- 3) 如果遇到信号没有极值点只有拐点的情况，那么在对其分解之前首先将原信号微分一次或者多次来获得极值点，然后对所得结果进行积分就可以得到相应的分量。

EMD 具体是通过一种被称为“筛分”处理的过程的分解。其流程详见图 6-2，具体的操作过程如下：

- 1) 处理的第一步，需要求出原信号 $s(t)$ 的所有极大值点和极小值点，然后将所有极大值点用三次样条曲线连接形成 $s(t)$ 的上包络线，用同样的方法连接所有极小值点得到 $s(t)$ 的下包络线。根据信号 $s(t)$ 的局部极大值和局部极小值求出上包络 $v_1(t)$ 以及下包络 $v_2(t)$ 的平均值

$$m_{11} = \frac{1}{2}[v_1(t) + v_2(t)] \quad (3)$$

然后求出 $s(t)$ 与 m_{11} 的差 h_{11} ：

$$h_{11} = s(t) - m_{11} \quad (4)$$

在理想情况下 m_{11} 应该满足 IMF 的定义，因为 h_{11} 的构造应该使其满足对称性。经过第一轮抽取，这个值有概率会成为局部极大值。在式 (4) 中产生的新的局部极值点说明了在初始步骤中遗失的符合要求的正确模态，经过反复的分解和抽取，抽取过程能够恢复原来的信号，将其表示成低振幅驾驭波。

抽取过程有两个目的：剔除驾驭波，并使波形更对称。围绕这两个目的，抽取过程必须重复所需要的次数，使减去抽取信号后成为一个 IMF。在随后的抽取过程中， h_{11} 仅仅能被认为是 IMF 原型，如果若 h_{11} 不是 IMF，将知视为新的 $s(t)$ ，重复式(3) 2 k 次。

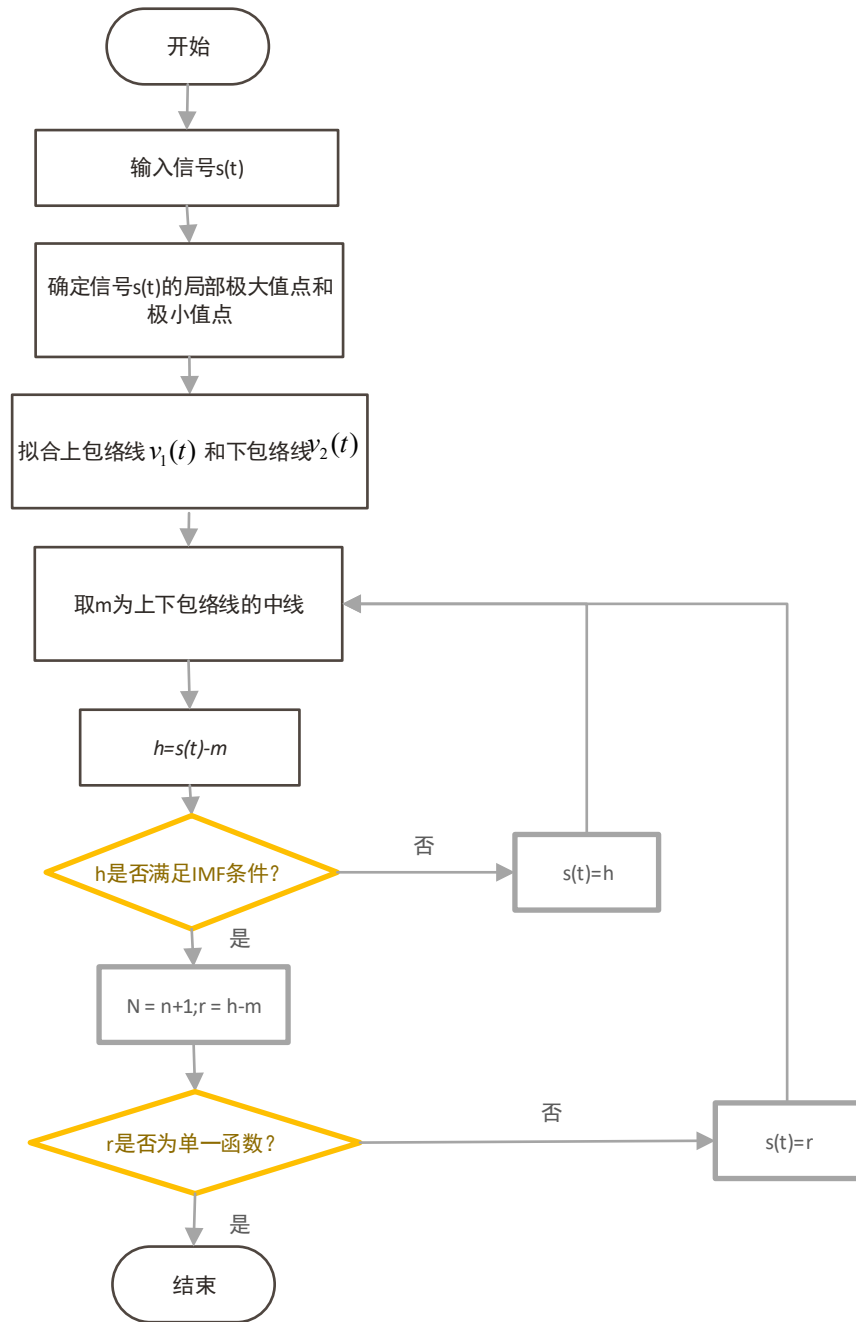


图 6-2 EMD 流程图

$$h_{1k} = h_{1(k-1)} - m_{1k} \quad (5)$$

式中 h_{1k} 和 $h_{1(k-1)}$ 分别为第 k 次和第 $k-1$ 次筛选所得的数据, m_{1k} 为 $h_{1(k-1)}$ 上下包络线的中线 (平均值), 式 (5) 计算出的 S_d 值可以用来看出每次“筛选”的结果是否为符合要求的 IMF 分量:

$$S_D = \sum_{t=0}^r \left| \frac{h_{1(k-1)}(t) - h_{1k}(t)}{h_{1(k-1)}^2(t)} \right|^2$$

这里需要一个关键性的停止准则：如果 S_D 值在 0.2-0.3 之间则令 $c_1 = h_{1k}$ 视为一个 IMF，作 $r = s(t) - c_1$ ，再视 r 为新的 $s(t)$ ，重复以上过程一次的得到第二个 $IMFc_2$ ，第三个 $IMFc_3 \dots$ 直到 $r(t)$ 成为单调函数或 $r(t)$ 的绝对值小于特定阈值时可视为测量误差即可停止，于是由各层 IMF 分量和余量叠加的信号 $s(t) = \sum_{i=1}^n c_i + r$ 表明了 EMD 分解的完备性。在上式中 r 为残量，代表信号中的平均趋势，各 IMF 分量 c_i 分别代表各个频率段信号的含量波形， i 值越小，频率越高，这种不同频率元素的部分时刻分布是随信号本身的变化而变化。

经过 EMD 处理后，提取 IMF 中代表总体趋势的分量作为预处理后的数据，输入到预测器中。

6.2 基于自适应指数模型的关键工况参数趋势预测

由于异常事件下的关键工况参数通常呈现出单调变化的特性，因此采用指数模型对其变化趋势进行建模和预测。随着异常工况报警之后在线数据的增加，自适应更新上述指数模型，以便获得更准确反映工况变化的趋势预测结果。

6.2.1 自适应指数模型的算法原理

在 t_i 时刻的目标变量趋势曲线可以用如下的指数模型表示

$$z_i = \phi + \theta e^{-t_i/\beta} = f(t_i, \phi, \theta, \beta)$$

上式中 ϕ, θ, β 为待定的模型参数， z_i 为 i 时刻目标变量的估计值。通过引入一个时域窗口，可将指数模型的目标函数描述为在线的优化问题：

$$\min_{\phi, \theta, \beta} \sum_{i=1}^N [y_i - f(t_i, \phi, \theta, \beta)]^2 \quad (6)$$

其中 y_i 为 i 时刻目标变量的真实值， N 为窗口长度， $f(t_i, \phi, \theta, \beta)$ 为目标变量的估计值。为最小化目标函数，采用迭代算法来求解。 k 时刻的最优解记为 $(\phi^*, \theta^*, \beta^*)_k$ ，代入指数模型，得到 k 时刻目标变量的估计值 z_k 。

在线预测的过程中，上述指数模型应采用过去一段时间上的数据不断自适应更新。基于自适应指数模型的关键工况参数趋势预测算法流程如下所示：

- (1) 采集 t_k 时刻，窗口长度为 s 的数据 z_k ，对其进行经验模态分解，提取包含整体趋势的 IMF 分量。

- (2) 为指数模型参数赋初值。
- (3) 在 k 时刻，基于 t_{k_s} 到 t_k 时刻的数据 $\{z_i\}_{i=k_s}^k$ 训练一个指数模型 $f(t_i, \phi, \theta, \beta)$ 。
- (4) 这个指数模型 $f(t_i, \phi, \theta, \beta)$ 能估计 k_s 到 k 时刻的预测值 $\{z_i\}_{i=k_s}^k$ ，即，

$$\hat{z}_j = \phi + \theta e^{-t_j/\beta}, j = k_s, k_s + 1, \dots, k$$

- (5) 等待下一时刻数据到来，跳转至步骤（1）。

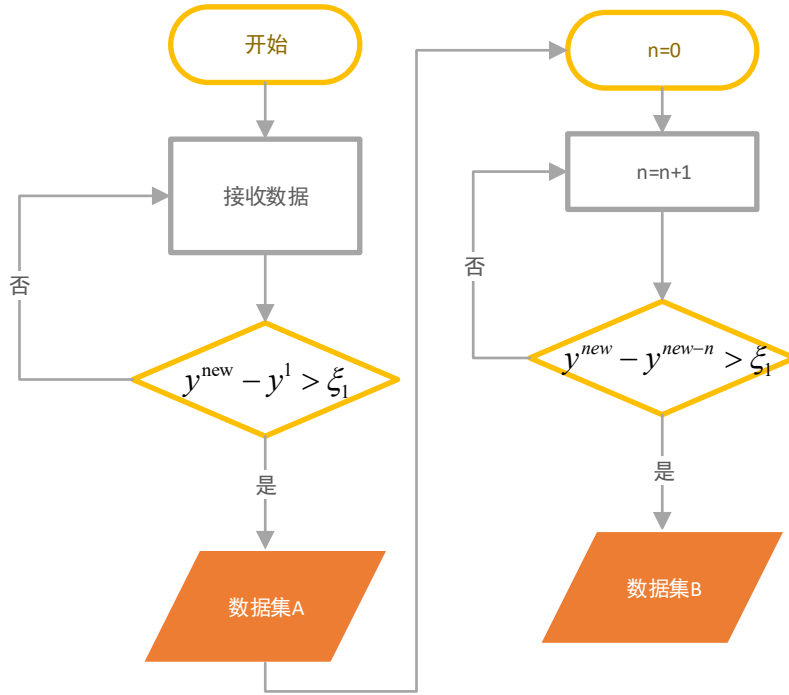


图 6-3 数据采集模块

在上述算法中，为使训练模型的数据窗口能够充分反映趋势信息而非局部变化，要求数据窗口所包含的信号具有充分大的变化。为此，窗口长度 s 的选取方式如下：设定阈值 ξ_1 ，随着时间不断采集新的数据，当采集的最新数据 y^{new} 与历史数据中的最初时刻数据 y^1 的差值大于阈值 ξ_1 时，确定数据窗口，再次索引历史数据中与最新数据 y^{new} 差值大于阈值 ξ_1 的最短窗口长度 S 。其中初始窗口长度设定为采集到的第一个数据与报警阈值的差值的六分之一，滑动窗口长度设定为初始窗口长度的2倍。

在 k 时刻，通过迭代法对式（6）求解，得最优解 $(\phi^*, \theta^*, \beta^*)_k$ 。可得 k 时刻目标变量估计值：

$$z_k = \phi^* + \theta^* e^{-t_k/\beta^*}$$

已知目标变量的报警阈值 ξ_2 ，满足：

$$\xi_2 = \phi^* + \theta^* e^{-(t_k + T_k)/\beta^*}$$

可求得剩余未停堆时间：

$$T_k = -\beta^* \ln \frac{\xi_2 - \phi^*}{\theta^*} - t_k$$

其中， T_k 表示在 k 时刻的剩余未停堆时间。

6.2.2 算法验证

由于没有实际故障数据，故选用正常工况下目标变量的过渡段数据来验证趋势预测算法。基于指数模型的预测算法验证数据集如表 6-1 所示,皆为堡垒机数据，停堆逻辑为 SG1 液位高与上 P7（P7=P13 或 P10）。

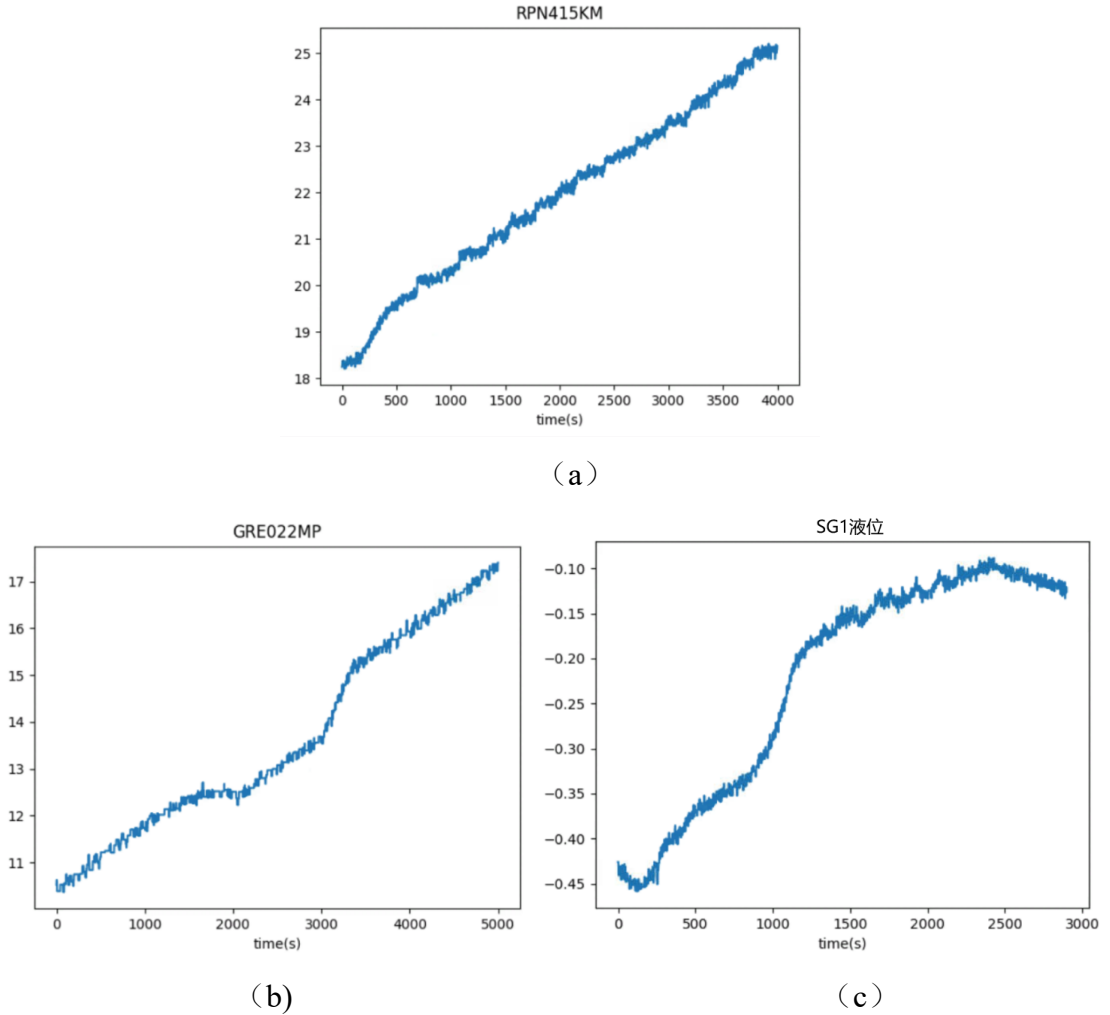


图 6-4 正常工况下多种目标变量的状态曲线

表 6-1 预测验证数据描述

目标变量	停堆阈值	验证数据时间段	目标参数
RPN415KM	24.5m	2019 年 4 月 15 日 16 时-17 时	P10
ARE052MN	-0.1m	2019 年 5 月 12 日 04 时-05 时	SG1
GRE022MP	17m	2019 年 4 月 15 日 15 时-17 时	P13

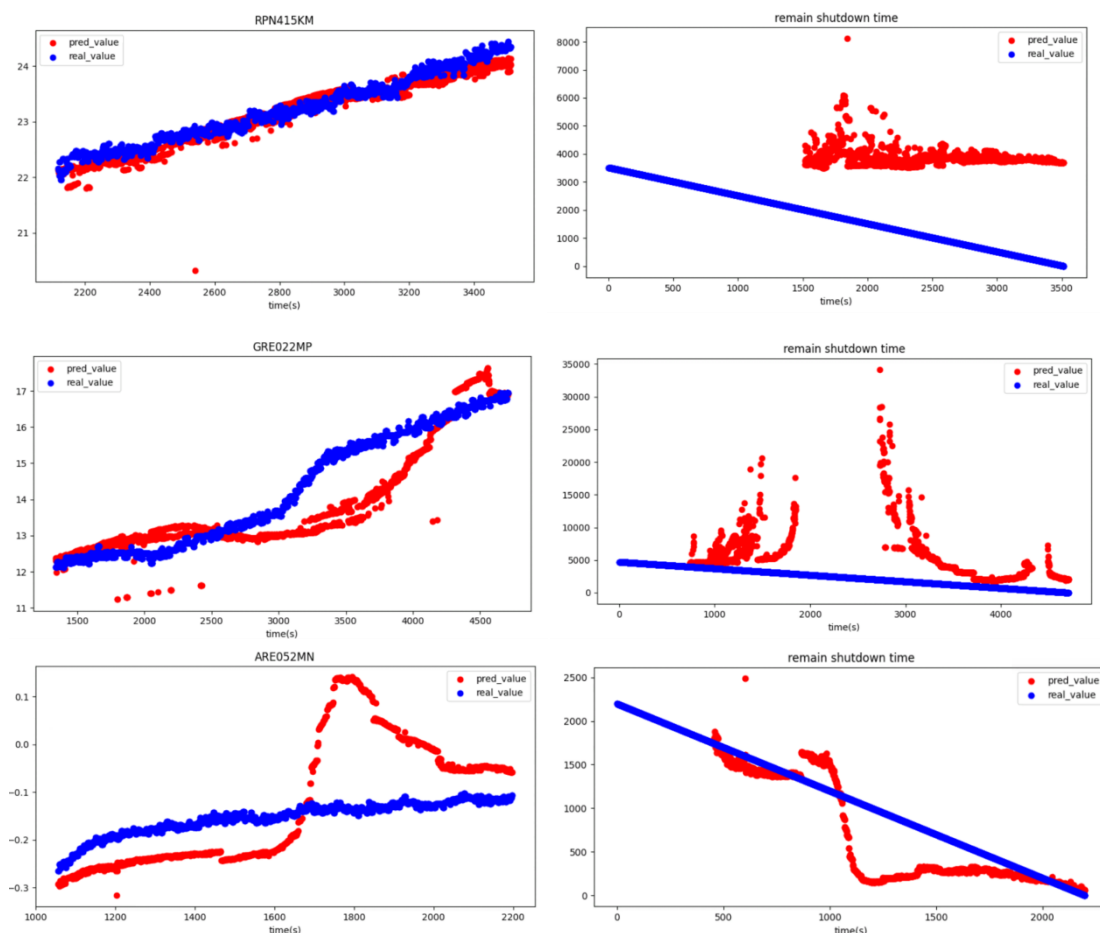


图 6-5 左图：趋势预测效果 右图：剩余未停堆时间

上述数据集上的算法验证效果如图 6-5 所示，其中左侧图中蓝色值表示目标变量的真实值，红色值表示目标变量十分钟的预测值。右侧图中曲线代表左侧目标变量在不同时刻所到达报警阈值的实际剩余停堆时间；红色点代表其所到达报警阈值的估计剩余停堆时间。

由图 6-5 可知，虽然指数模型可以对目标变量未来十分钟的数值进行较为准确的预测，但由于测试集数据是一段非常近似于线性模型的数据，故求解出来的参数 $(\phi^*, \theta^*, \beta^*)$ 使得指数模型中的指数部分 e^{-t_k/β^*} 接近于 1，在估计 k 时刻的剩余停堆时间时，所求得的 T_k 数值总是很大，这就出现了如图 6-6 中右图所示的情况，剩余停堆时间的估计值与真实值有很大偏离。

6.3 基于自适应线性模型的关键工况参数趋势预测

由于故障工况下以及正常工况的过渡情况下，目标参数的数据表现为近似线性，于是考虑采用线性模型，使其求解出来的参数更接近真实情况，从而使得剩余停堆时间的预测更加准确。

6.3.1 自适应线性模型的算法原理

在 t_i 时刻的目标变量趋势曲线可以用如下的线性模型表示：

$$z_i = \phi + \theta t_i = f(t_i, \phi, \theta)$$

上式中 ϕ, θ 为随机变量， z_i 为 i 时刻目标变量的估计值。通过引入一个时域窗口，可将线性模型的目标函数描述为在线的优化问题：

$$\min_{\phi, \theta} \sum_{i=1}^N [y_i - f(t_i, \phi, \theta)]^2 \quad (7)$$

其中 y_i 为 i 时刻目标变量的真实值， N 为窗口长度， $f(t_i, \phi, \theta)$ 为目标变量的估计值。为最小化目标函数，采用迭代算法来求解。 k 时刻的最优解记为 $(\phi^*, \theta^*)_k$ ，代入指数模型，得到 k 时刻目标变量的估计值 z_k 。

在线预测的过程中，上述线性模型应采用过去一段时间上的数据不断自适应更新。基于自适应指数模型的关键工况参数趋势预测算法流程如下所示：

- (6) 采集 t_k 时刻，窗口长度为 s 的数据 z_k ，对其进行经验模态分解，提取包含整体趋势的IMF分量。
- (7) 为线性模型参数赋初值。
- (8) 在 k 时刻，基于 t_{k_s} 到 t_k 时刻的数据 $\{z_i\}_{i=k_s}^k$ 训练一个线性模型 $f(t_i, \phi, \theta)$ 。
- (9) 这个线性模型 $f(t_i, \phi, \theta)$ 能估计 k_s 到 k 时刻的预测值 $\{z_i\}_{i=k_s}^k$ ，即，

$$\hat{z}_j = \phi + \theta t_j, j = k_s, k_s + 1, \dots, k$$

- (10) 等待下一时刻数据到来，跳转至步骤(6)。

在 k 时刻，通过迭代法对式(7)求解，得最优解 $(\phi^*, \theta^*)_k$ 。可得 k 时刻目标变量估计值：

$$z_k = \phi^* + \theta^* t_k$$

已知目标变量的报警阈值 ξ_2 ，满足：

$$\xi_2 = \phi^* + \theta^* (t_k + T_k)$$

可求得剩余停堆时间：

$$T_k = \frac{\xi_2 - \phi^*}{\theta^*} - t_k$$

6.3.2 算法验证

采用图 6-5 中数据及 6.2.2 节中堡垒机数据集作为验证数据集。

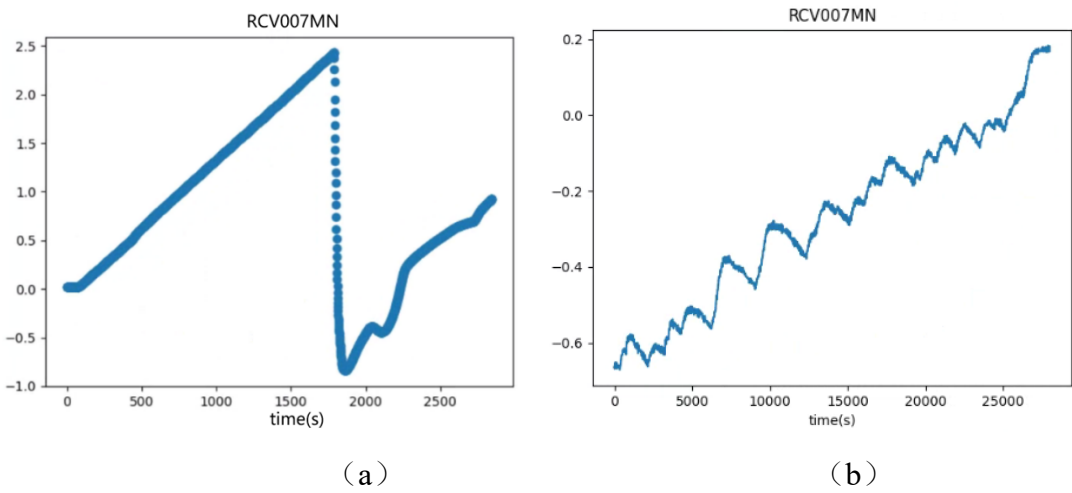
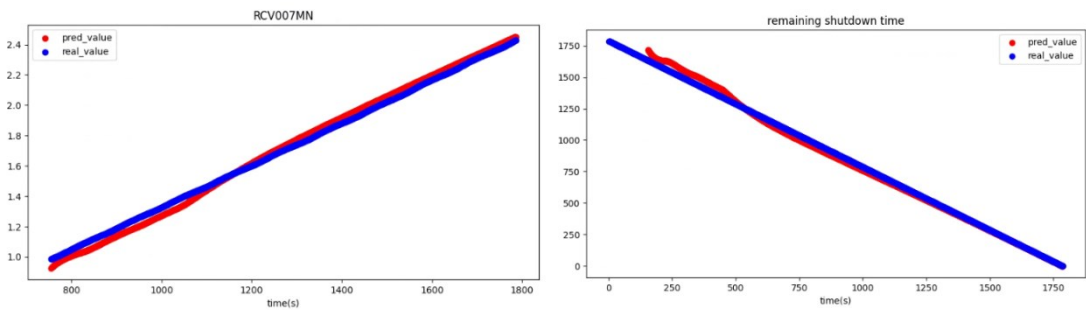


图 6-6 目标变量的仿真状态曲线

图 6-6 中 (a)图为下泄故障，是仿真故障工况数据集；(b) 图为稳压器液位正常工况数据。

表 6-2 预测验证数据描述

目标变量	停堆阈值	验证数据时间段	目标参数
RCV007MN（仿真数据）	2.43m	无	P7
RCV007MN	0.1m	2018 年 6 月 4 日 06 时-10 时	P7



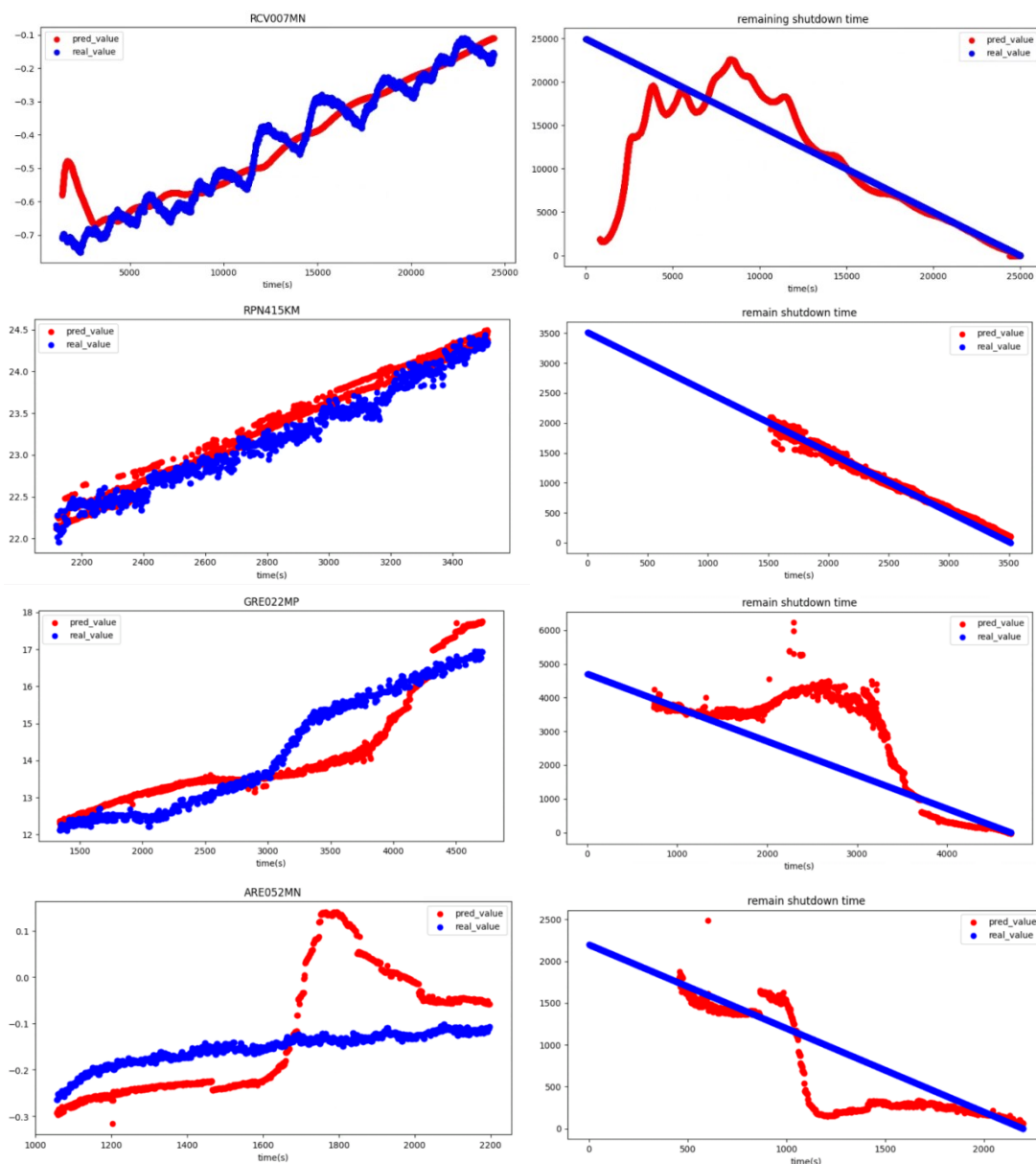


图 6-7 左图：趋势预测效果 右图：剩余未停堆时间

图 6-7 中左侧图中蓝色值表示目标变量的真实值，红色值表示目标变量十分钟的预测值。右侧图中曲线代表左侧目标变量在不同时刻所到达报警阈值的实际剩余停堆时间；红色点代表其所到达报警阈值的估计剩余停堆时间。

由图 6-7 所知，线性模型不仅能对目标变量未来十分钟的数值进行较为准确的预测，在剩余停堆时间的预测上，也能随着时间的推进，给出较为准确的估计值。

6.4 基于自回归多步预测模型的关键参数趋势预测

核电厂数据的目标传感器变量存在线性相关性和时间相关性，仅仅分析单个参数变化趋势在系统中应用时，会忽略参数间的关联特征，影响实际应用中预测结果的精度。为此，本项目拟研究一种基于线性模型的的多输入多输出的多步滚

动预测的方法。该方法的一个优点是可以同时捕获多个变量之间的相关性和每个参数的时间关联性，以提高关键参数预测的能力。

6.4.1 自回归多步预测模型

自回归多步预测模型是在单步预测模型的基础上进行的，单步预测模型直接以预测目标值作为预测模型输出，即以 $x(t)$ 为预测模型输入，以 $x(t+a)$ 为预测模型输出，预测步长 a 取为预报提前时间对应的时间间隔数，一步预测模型如下：

$$x(t+a) = Ax(t) + b = f[x(t)]$$

考虑变量之间的相关性，在本项目中， $x(t) = [x_1(t), x_2(t), \dots, x_n(t)]^T$ ， $x_i(t)$ 为 t 时刻第 i 个传感器变量， A 为待求解系数矩阵， b 为待求解向量。

通过最小化一步预测误差的均方值来求解参数，其目标函数为：

$$\begin{aligned} J &= \frac{1}{2} \sum_{i=1}^M \|x(i) - \hat{x}(i)\|_2^2 \\ &= \frac{1}{2} \sum_{i=1}^M \|x(i) - f(x(i-a))\|_2^2 \end{aligned}$$

其中， $x(i)$ 表示观测值， $\hat{x}(i)$ 表示预测值。通过对目标函数求导得出解析解：

$$[A^* \quad b^*] = Y^T X^T \cdot (XX^T)^{-1}$$

其中：

$$X = \begin{bmatrix} x_1(t) & x_1(t-1) & \cdots & x_1(t-l) \\ \vdots & \vdots & \ddots & \vdots \\ x_n(t) & x_n(t-1) & \cdots & x_n(t-l) \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

$$Y = [x(t+a)^T, x(t+a-1)^T, \dots, x(t+a-l)^T]^T$$

l 表示时间窗口长度。

基于上述一步预测模型，按以下方式迭代实现多步预测：

$$\begin{aligned} \hat{X}(t+a) &= f[X(t)] \\ \hat{X}(t+2a) &= f[\hat{X}(t+a)] \\ &\vdots \\ \hat{X}(t+(k+1)a) &= f[\hat{X}(t+ka)] \end{aligned}$$

最终通过多步预测值触发停堆逻辑的时刻计算得到自动停堆的剩余时间。

6.4.2 算法验证

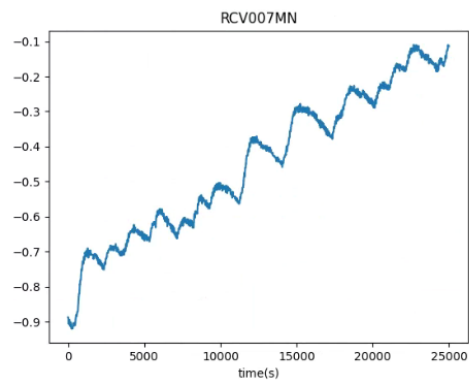


图 6-8 目标变量 RCV007MN 正常工况

采用目标变量 RCV007MN 在 2018 年 6 月 4 日 06 时-10 时的正常工况数据作为验证集，如图 6-8 所示。

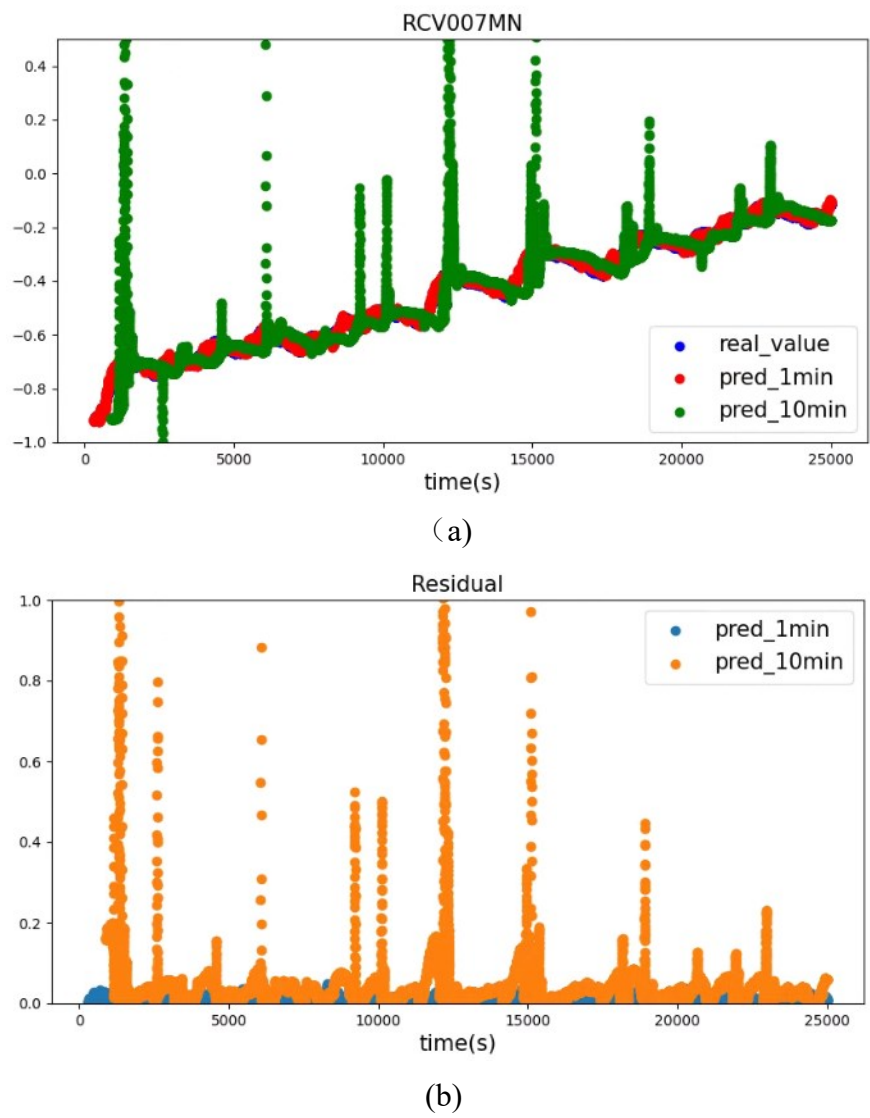


图 6-9 目标变量 RCV007MN 正常工况

图 6-9(a)图中蓝色点表示该目标变量的真实值，红色点表示该目标变量未来 30s 的预测值，绿色点是由 30s 一步预报器串联起来的多步预测器的估计值，表示该目标变量未来 10 分钟的预测值。图 6-9(b)图中蓝色点表示 30 秒预测值与真实值的残差，橙色点表示 10 分钟预测值与真实值的残差。

在图 6-9 中 30s 的预测值曲线与真实值几乎完全重合，可得预测时间越短，预测精度越高。随着预测时间的增加，滚动误差不断累积，会产生不断累积的时延，影响预测的精度。虽然自回归多步预测模型可以给出较为可参考的预测值，但其迭代过程中，其变量之间的相关性越来越低，使得最终很难求出有效的剩余停堆时间。