

DOI: 10. 20040/j. cnki. 1000-7709. 2023. 20222091

基于关联分析的水电机组调速器运行数据多维度校验方法

李文友

(雅砻江流域水电开发有限公司, 四川 成都 610051)

摘要: 为准确掌握水电机组调速器的实际运行情况,提出了一种基于关联分析的水电机组调速器运行数据多维度校验方法,即首先构建多维度关联校验架构,处理水电机组调速器多维度运行数据;然后采用关联分析法计算多维度运行数据变化序列之间的关联度,通过判断关联度的范围完成多维度运行数据的异常辨识;最后基于辨识后的异常数据特点分别从单点、多点及连续异常数据三个维度完成异常数据重构,实现水电机组调速器运行数据多维度校验。试验结果表明,该方法能够有效辨识异常数、重构缺失数据,通过校验降低数据异常率,提高了水电机组调速器运行数据的可用性。

关键词: 关联分析; 水电机组; 调速器数据; 多维度校验; 异常辨识; 异常重构

中图分类号: [TV734. 1]; TP391

文献标志码: A

文章编号: 1000-7709(2023)09-0175-05

1 引言

水电规模的日益扩大,严重影响电力系统的运行与规划^[1]。水力发电与传统发电方式不同,水电机组调速器的设备数量众多,单个设备容量极小,运行数据处理不当会出现运行数据缺失与不完整现象^[2]。然而水电机组调速器的历史运行数据是研究水电相关问题的基础,传统调度监控方式不能获得水电机组调速器运行的确切状态,可通过数据校验对水电机组调速器的运行数据进行查缺补漏,提高历史数据可信度^[3,4]。刘君等^[5]提出了基于 Zigbee 的循环冗余校验方法,该方法仅对异常数据进行辨识,未能修正或填补异常数据,导致数据后期的可用性差;朱智桑等^[6]提出基于隐蔽信道的数据校验技术,有效结合 Modbus-TCP 协议及隐蔽信道完成数据的哈希校验,但该方法未考虑数据异常的多种情况,未实现调速器数据的多维度校验,校验结果应用局限性较大。为此,本文提出了一种基于关联分析的水电机组调速器运行数据多维度校验方法,并通过试验验证了该方法的可行性和有效性。

2 水电机组调速器运行数据的多维度关联校验方法

水电机组调速器运行数据的变量种类多、数

据量庞大,为验证水电机组调速器运行数据的可靠性,本文基于关联分析方法提出水电机组调速器运行数据多维度校验架构,校验水电机组调速器的运行数据,具体见图 1。

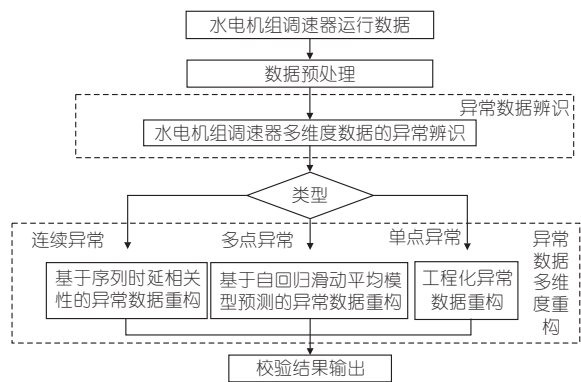


图 1 基于关联分析的多维度校验架构

Fig. 1 Multidimensional verification architecture based on association analysis

基于关联分析的水电机组调速器运行数据多维度校验方法,首先采用均值法对水电机组调速器多维度运行数据进行预处理,以保证各序列的数据量纲一致。然后运用关联分析方法计算水电机组调速器各参数无异常情况的变化序列之间的关联度及水电机组调速器运行数据无异常情况的序列与有异常情况的序列之间的关联度。通过序列之间的关联度实现异常数据的辨识^[7]。最后运用填充法、自回归滑动平均模型预测方法及序列

收稿日期: 2022-11-25, 修回日期: 2023-01-07

基金项目: 国家重点研发计划(2019YFE0105200)

作者简介: 李文友(1973-),男,硕士、高级工程师,研究方向为电力自动化及远程控制, E-mail: gaoyihuic6@163. com

时延相关性方法完成单点、多点及连续异常数据的重构,实现水电机组调速器运行数据多维度校验。

2.1 数据预处理

设 $X_i = \{X_i(k)\}$ 为数据未出现异常时水电机组调速器的多维度运行数据的变化序列,多维度状态参数包括水电机组调速器的有功功率、受油器摆度、组合轴承摆度、水导摆度及受油器振动^[8,9]。 $Y = \{Y(k)\}$ 为数据出现异常时水电机组调速器的数据在当前时刻与前一时刻的变化序列, $i = 1, 2, \dots, m, k = 1, 2, \dots, n$, 其中 n, m 为自然数。

因各序列的数据量纲不同,所以在水电机组调速器多维度数据进行关联分析前,需采用均值法对水电机组调速器的多维度数据进行预处理,处理后的变化序列分别由 $x_i(k), y(k)$ 表示,具体计算公式为:

$$\begin{cases} y(k) = Y(k) / \bar{Y}(k) \\ x_i(k) = X_i(k) / \bar{X}_i(k) \end{cases} \quad (1)$$

其中
$$\bar{Y}(k) = \frac{1}{n} \sum_{k=1}^n Y(k) \quad (2)$$

$$\bar{X}_i(k) = \frac{1}{n} \sum_{k=1}^n X_i(k) \quad (3)$$

综上所述,将 $X_i(k), Y(k)$ 转变为两个数据变化序列 $x_i(k)$ 与 $y(k)$ 。

2.2 基于关联分析的运行数据异常辨识

关联分析法是根据两个事物之间的相似程度,判断各事物之间的关联程度,若两个事物发展及变化趋势的拟合度较高,即这两个事物的关联度较高,反之,这两个事物关联度较低^[10]。基于关联分析的水电机组调速器多维度数据的异常辨识,采用关联分析法计算变化序列之间的关联度范围,若水电机组调速器多维度数据中出现异常,则两个变化序列的关联度超出设定范围,以此完成水电机组调速器多维度数据的异常辨识^[11,12]。

当水电机组调速器多维度数据未出现异常时,其历史数据变化序列关联度的数值较小,若水电机组调速器多维度数据出现异常时,则当前数据变化序列与历史数据变化序列的关联度不在设定范围内,且差异较大。两个数据变化序列为:

$$\begin{cases} x_i = \{x_i(k) \mid k = 1, 2, \dots, n\} \\ y = \{y(k) \mid k = 1, 2, \dots, n\} \end{cases} \quad (4)$$

式中, n 为水电机组调速器多维度数据的测量数目; m 为水电机组调速器多维度数据中无异常数据的总数。

假设 ξ_{ij} 为水电机组调速器多维度数据无异常情况时变化序列之间的关联系数,具体公式为:

$$\xi_{ij}(k) = \frac{M_k + \min_j \min_k |x_i(k) - x_j(k)|}{M_k + |x_i(k) - x_j(k)|} \quad (5)$$

$$M_k = \max_j \max_k |x_i(k) - x_j(k)| \rho \quad (6)$$

式中, $x_i(k), x_j(k)$ 分别为数据未出现异常时水电机组调速器第 i 种参数与第 j 种参数的变化序列, $j = 1, 2, \dots, m; \rho$ 为分辨系数,其取值范围为 $(0, 1)$, ρ 值越小代表水电机组调速器多维度数据的变化序列之间关联度越高。

假设 $R(x_i, x_j)$ 为序列 x_i 与 x_j 之间的加权关联度,其具体由下式描述:

$$R(x_i, x_j) = \sum_{k=1}^n \frac{\omega_k \xi_{ij}(k)}{n} \quad (7)$$

式中, ω_k 为水电机组调速器多维度数据的测量点 k 所对应的测试权重。

若未出现异常时水电机组调速器多维度数据变化序列中数值均有效,则将各序列逐个作为参考序列,以此计算 x_i 与 x_j 两个序列之间的关联度, x_i 与 x_j 之间的关联度越高,序列 x_i 数据出现异常情况时,序列 x_j 数据异常的可能性越大,通过辨识到序列 x_i 的数据异常,快速得知与其关联性较高的数据序列均为异常。

假设 ϕ_i 为序列 $x_i(k)$ 与 $y(k)$ 之间的关联系数, ϕ_i 的具体公式为:

$$\phi_i = \frac{M_k + \min_i \min_k |y(k) - x_i(k)|}{M_k + |y(k) - x_i(k)|} \quad (8)$$

假设 $D(x_i, y)$ 为无异常数据序列 $x_i(k)$ 与异常数据变化 $y(k)$ 之间的关联度,其具体公式为:

$$D(x_i, y) = \sum_{k=1}^n \frac{\omega_k \phi_i(k)}{n} \quad (9)$$

其中
$$\omega_k = 1 / (e\lambda)^{L_k} \quad (10)$$

式中, e 为无限不循环小数; λ 为权重区分系数; L_k 为与测量点 k 相对的匹配出线度。

若 $D(x_i, y)$ 在设定范围内,说明水电机组调速器运行数据中未出现异常,若 $D(x_i, y)$ 超过设定范围,说明水电机组调速器运行数据出现缺失或不完整等异常情况。即通过判断关联度 $D(x_i, y)$ 与加权关联度 $R(x_i, x_j)$ 中各序列之间的关系,完成水电机组调速器运行的多维度数据的异常辨识。

2.3 水电机组调速器异常运行数据重构

经过水电机组调速器多维度数据的异常辨识后,提升了水电机组调速器中多维度运行数据的可用性,但水电机组调速器中多维度运行数据易出现缺失现象,当存在过多异常数据需被剔除时,会降低水电机组调速器中多维度运行数据的可用

性。所以对辨识后的异常数据特点可将其分为单点异常、多点异常及连续异常三种特征。基于这三种特征对水电机组调速器运行数据进行多维度重构,实现最终的水电机组调速器运行数据的多维度校验。

2.3.1 基于填充法的单点异常数据工程化重构

运用填充法对水电机组调速器运行过程中出现的单点异常数据实现工程化重构,具体为采用一个经验常量或均值替换异常数据,该方法操作简单、节省时间并且精准性较好。

(1)常量填充。基于以往经验对水电机组调速器运行过程中出现的单点异常数据运用一个固定常量值完成数据填充。

(2)均值填充。将水电机组调速器出现单点异常时的前后多维度运行数据进行均值求解,将获得的平均值完成数据填充。

2.3.2 基于自回归滑动平均模型预测的多点异常数据重构

因为水电机组调速器运行异常点的两侧数据是准确的,所以可从两侧数据对异常数据进行逆向或正向的重构,以此获得逆向与正向的重构结果。由于自回归滑动平均模型预测方法会随着时间的延长,预测准确率会逐渐降低,因此运用不同长度预测结果的可接受率,分配给预测结果不同的权重,完成逆向与正向两个预测值在重构结果中的权重分配,以此获得更准确的重构结果。

(1)基于自回归滑动平均模型预测结果的可接受率。对于水电机组调速器完整的历史运行数据,运用自回归滑动平均模型预测 u 步,若第 p 步的预测误差 E_p 比误差可接受阈值 ϵ_{error} 小,则代表该预测结果准确性较高。对大量水电机组调速器的历史运行数据预测仿真 N 次,则第 p 步预测结果的可接受率 γ_p 的计算公式为:

$$\gamma_p = [\text{Count}(|E_p| < \epsilon_{\text{error}}) / N] \times 100\% \quad (11)$$

式中,Count()为计数器函数,若满足该函数中的逻辑关系,则计数器加 1。

由式(11)可知, γ_p 与 p 呈负相关关系。

(2)基于自回归滑动平均模型预测方法的双向预测。设水电机组调速器缺失数据时段为 $[g, h]$,基于自回归滑动平均模型预测方法对该时间段内水电机组调速器运行数据实现双向预测,预测结果由 $\{\vec{P}_g, \vec{P}_{g+1}, \dots, \vec{P}_h\}$ 与 $\{\bar{P}_g, \bar{P}_{g+1}, \dots, \bar{P}_h\}$ 描述。 \vec{P}_g, \bar{P}_g 分别为 g 时正向、逆向预测结果。

(3)加权重构值计算。 b 时 \vec{P}_b 的预测时间长度为 $b-g+1$,可接受率为 γ_{b-g+1} ,而 b 时 \bar{P}_b 的预

测时间长度为 $h-b+1$,可接受率为 γ_{h-b+1} 。由此获得 b 时水电机组调速器异常数据重构结果的计算公式为:

$$P_b = \frac{\vec{P}_b \gamma_{b-g+1} + \bar{P}_b \gamma_{h-b+1}}{\gamma_{b-g+1} + \gamma_{h-b+1}} \quad (12)$$

P_b 的计算结果即为实现水电机组调速器运行过程中多点异常数据重构的填充结果。

2.3.3 基于序列时延相关性的连续异常数据重构

因为水电机组调速器运行数据之间包含相关性,所以能运用基于序列时延相关性的方法,对水电机组调速器运行中连续的异常数据进行重构。

(1)计算水电机组调速器运行数据的时延相关性。将序列 $E = \{e_1, e_2, \dots, e_r\}$ 与序列 $F = \{f_1, f_2, \dots, f_r\}$ 设为水电机组调速器运行数据的时间序列, r 为任意自然数, F 对于 E 延迟 l 时的相关系数为 $R(l)$,其计算公式为:

$$R(l) = \frac{\sum_{t=l+1}^r (f_{t-l} - \bar{f})(e_t - \bar{e})}{\sqrt{\sum_{t=1}^{r-1} (f_t - \bar{f})^2} \sqrt{\sum_{t=l+1}^r (e_t - \bar{e})^2}} \quad (13)$$

其中 $\bar{e} = \frac{1}{r-l} \sum_{t=l+1}^r e_t$; $\bar{f} = \frac{1}{r-l} \sum_{t=l+1}^r f_t$

式中, t 为数据缺失开始时刻; l 的取值范围为 $(0, n/2)$,以此获得相关系数 $R(0), R(1), \dots, R(n/2)$ 。

将 l' 设为 $R(l)$ 最大时所对应的最大时延相关点,若 $R(l')$ 大于相关性阈值 σ ,则表示 E 与 F 之间存在时延相关性。

(2)基于时延相关性重构水电机组调速器运行数据。待数据重构的水电机组调速器为 W_0 ,其他水电机组调速器为 W_1, W_2, \dots 。运用在时段 $[g-r, h]$ 内水电机组调速器运行中水流速度数据。 W_1 的最大时延相关点为 l'_1 ,其对应的相关系数为 $R(l'_1)$,设定 $R(l'_1) = \max(R(l'_1), R(l'_2), \dots)$,可采用 W_1 的水流速度曲线重构 W_0 的历史运行数据。因为 W_0 与 W_1 之间存在时延相关性,则可将 W_0 与 W_1 水流速度之间的关系具体表达为:

$$P_g^0 \approx c + aP_{g+l'}^1 \quad (14)$$

式中, P_g^0 为 W_0 在 g 时的水流速度; a, c 均为未确定系数; $P_{g+l'}^1$ 为 W_1 在 $g+l'$ 时的水流速度。

W_0 缺失段的数据可依据 W_1 的水流速度运用式(14)完成水电机组调速器运行数据的重构。

3 试验分析

试验选取额定功率为 42.44 MW 的某水电

站中的水电机组为例,采用 Matlab 模拟软件进行模拟试验,存储水电机组调速器运行数据的数据库为 Micro-Clusters2018。该水电站水电机组的水轮机调速器型号为 JY-HYTS-202,其操作油压为 2.5 MPa,额定油压为 16 MPa,调节规律为适应式 PID 调节,频率给定调整范围为 $50(1+10\%)$ Hz,交流工作电源为 $AC220V\pm 15\%$,直流工作电源为 $DC220(1+15\%)$ V,机频信号电压与网频信号电压范围均为 $AC0.3V\sim AC150V$ 。该水电站运用计算机系统对水电机组调速器的运行数据进行监测,通过本文方法对异常数据进行辨识与重构实现数据校验。

为验证本文方法的异常数据辨识的优劣程度,采用本文方法对监测数据进行异常辨识,具体结果见表 1。由表 1 可知,本文方法能够对多种类型的水电机组调速器运行数据进行异常辨识,并获取其异常辨识结果。

表 1 水电站调速器运行数据辨识结果

Tab. 1 Identification results of governor operation data of hydropower station

参数	监测数据	历史数据	辨识结果
有功功率/MW	21.56	3.48	数据异常
受油器摆度 A/ μm	62.46	30.56	数据异常
受油器摆度 B/ μm	61.84	21.96	数据异常
组合轴承摆度 A/ μm	113.48	110.96	数据正常
组合轴承摆度 B/ μm	146.23	92.15	数据异常
水导摆度 A/ μm	43.06	24.26	数据异常
水导摆度 B/ μm	38.84	40.34	数据正常
受油器振动/ μm	101.52	54.32	数据异常

试验检验采用本文方法重构的水电机组调速器出力异常数据与水电机组调速器实际出力曲线的拟合度,以此验证本文方法重构异常数据的有效性。以 2021 年 6 月 13 日该水电站某一水电机组调速器实际出力数据为例,依据本文方法辨识到该水电机组调速器出力数据缺失状况,并采用本文方法对缺失数据进行重构,具体效果见图 2。由图 2 可知,重构数据与实际数据的差值较小,曲线拟合度较高,能够提高水电机组调速器运行数据的可用性。

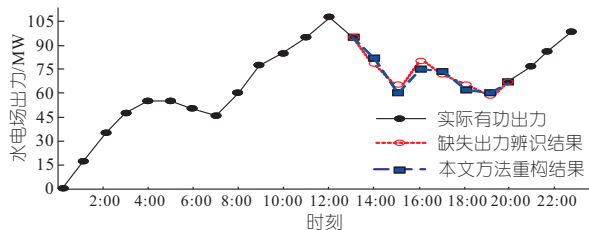


图 2 本文方法重构效果

Fig. 2 Reconstruction effect of this method

对 2021 年 6 月 13 日水电机组调速器运行数据进行校验,具体结果见图 3。由图 3 可知,水电

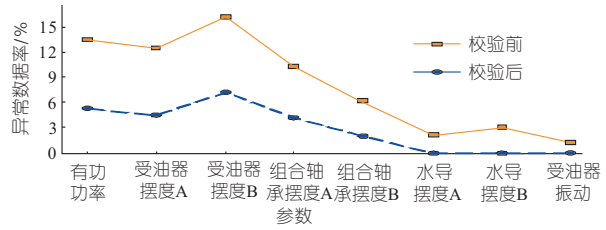


图 3 水电机组调速器运行数据多维度校验结果

Fig. 3 Multi-dimensional verification results of governor operation data of hydropower unit

机组调速器运行数据进行校验前,各监测参数的数据异常率较高,尤其有功功率、受油器摆度及组合承摆度这三个参数的数据异常率均超过 6%;对水电机组调速器运行数据采用本文方法进行多维度数据辨识与重构后,数据异常率均有所下降,水导摆度与受油器振动的运行数据甚至得到完全校正。

为验证本文方法实现水电机组调速器运行数据多维度校验的应用性能,抽取 2021 年 6 月 12 日~2021 年 7 月 12 日期间水电机组调速器的有功功率进行校验。采用本文方法多维度校验水电机组调速器有功功率的结果见图 4。由图 4 可知,运用本文方法能够辨识到水电机组调速器有功功率的结果的异常数据,并通过多维度数据重构方法对缺失数据及错误数据进行填补重构,完成了水电机组调速器运行数据的多维度校验,增强了运行数据的完整性及准确性。

水电机组调速器运行数据多维度校验				
运行数据库		数据校验		
有功功率				
参数类别	年	月	日	日至
	2021	06	12	2021 07 12
数据库数据	采样时间	数据辨识	数据重构	校验结果
2.48	2021.06.12 10:06:31	数据异常	20.48	校验完成
40.69	2021.06.15 10:13:26	数据异常	20.69	校验完成
3.48	2021.06.18 10:08:37	数据异常	23.48	校验完成
20.48	2021.06.21 10:21:15	数据正常	-	校验完成
2.39	2021.06.24 10:16:58	数据异常	21.39	校验完成
0.56	2021.06.27 10:09:47	数据异常	20.56	校验完成
21.86	2021.06.30 09:58:12	数据正常	-	校验完成
62.24	2021.07.03 10:08:32	数据异常	22.24	校验完成

图 4 多维度校验水电机组调速器有功功率的结果

Fig. 4 Results of multi-dimensional verification of the active power of the governor of hydropower units

4 结论

针对水电机组调速器的历史运行数据在采集、测量及传送过程中因故障或受到干扰易出现数据缺失、异常与不完整的现象会影响水电机组调速器数据应用的可靠性,提出了一种基于关联分析的水电机组调速器运行数据多维度校验方法,该方法能有效辨识异常数、重构缺失数据,提高了数据质量,可真实反映水电机组调速器的运

