

基于孤立森林与随机子空间的小干扰稳定评估

贺潇瑞

(东北电力大学电气工程学院, 吉林 吉林 132012)

摘要: 环境激励下的电力系统随机响应蕴含着系统丰富的动态信息, 通过随机子空间技术提取出随机响应中的小干扰稳定特征参数, 但是提取过程中往往容易混入异常数据, 导致无法判断出系统的真实状态。针对此问题提出了基于随机子空间与孤立森林的小干扰稳定评估方法, 使用孤立森林算法对随机子空间技术提取的小干扰稳定特征参数结果进行异常值检测, 有效改善了随机子空间技术获取系统的特征参数(振荡频率、阻尼比)的准确性。最后对 IEEE 16 机系统进行实例仿真, 仿真结果表明了该方法的可行性。
关键词: 随机响应; 小干扰稳定特征参数; 随机子空间; 异常值检测; 孤立森林

Small signal stability assessment based on isolation forest with stochastic subspace identification

HE Xiaorui

(School of Electrical Engineering, Northeast Electric Power University, Jilin 132012, China)

Abstract: The stochastic response of power system under environmental excitation contains rich dynamic information of the system, and the small signal stability characteristic parameters in the stochastic response can be extracted by the stochastic subspace technique, but the extraction process is often easy to mix in abnormal data, which leads to the inability to judge the real state of the system. To address this problem, a small signal stability evaluation method based on random subspace and isolation forest is proposed, and the isolation forest algorithm is used to detect outliers in the results of small signal stability characteristic parameters extracted by random subspace technique, which effectively improves the accuracy of obtaining the characteristic parameters (oscillation frequency and damping ratio) of the system. Finally, an example simulation is performed in the IEEE16 machine system, and the simulation results show the feasibility of the method.

key words: random response; small signal stability characteristics parameters; stochastic subspace; outlier detection; isolation forest

0 引言

随着现代电力系统的快速发展, 大电网互联充分展现了其优越性, 同时产生的小干扰稳定问题也给电网安全运行带来了巨大的挑战, 大量的研究表明如何快速准确地获取小干扰稳定特征参

数是抑制低频振荡问题的关键^[1-3]。

目前国内外研究主要包括基于数学模型与基于系统量测信号的分析方法^[4-5], 由于电力系统规模的增大, 使得动态模型受到“维数灾”的限制。基于数学模型的分析方法仅适用于离线分

析, 且不再适用于分析高维度系统。近几年, 广域测量技术的快速发展使得在线提取小干扰稳定特征参数成为了可能, 相较于大扰动事故后暂态响应信号, 系统在环境激励下的随机响应蕴含着大量系统动态特征信息且易于提取^[6], 因此出现了大量基于随机响应的特征提取方法, 如自回归滑动平均模型^[7]、随机子空间算法^[8]和基于NExt-DMD机电振荡参数提取方法等。然而在应用上述方法进行在线实时小干扰稳定特征参数提取时, 由于算法模型和量测过程中不可避免的误差, 有时会产生离散异常数据, 影响工作人员对系统实时状态做出正确的判断, 因此亟需一种方法对振荡参数集中的异常数据进行监控。

针对上述辨识特征参数结果混入异常值的问题, 本文引入一种基于孤立森林的异常检测技术。区别于基于聚类与距离的无监督检测方法, 孤立森林算法放弃了对正常数据的数学建模过程, 是一种基于集成学习的异常检测算法^[9]。目前该算法被广泛应用于水文监测^[10]、网络数据流量分析^[11]、电力运维数据清洗^[12]和直流漏电检测^[13]等方面。

本文在含一定负荷波动的随机响应信号中, 通过随机子空间算法在线获取相应小干扰稳定参数指标, 为了减小异常值对辨识数据的影响, 对每一个模式下辨识的结果使用孤立森林算法进行异常值检测, 以此来获得更加精确的参数指标。对IEEE 16机68节点系统进行仿真, 将辨识的结果与异常检测后的结果进行对照, 证明了该方法的有效性。

1 理论基础

1.1 基于随机子空间的特征参数辨识

发电机侧相量测量单元 (PMU) 的广域测量系统 (WAMS) 能够实时提取功率、频率和相角等电网动态信息。当电网受到环境激励时, 会产生杂乱的类噪声随机响应信号。作为目前较为成熟且应用比较广泛的技术, 基于系

统辨识理论的随机子空间算法 (SSI) 能够通过辨识获得系统模态信息, 并分析系统小干扰稳定特性。

实际电力系统中, 由于PMU采样得到的数据在时间上具有离散的性质, 因此系统状态空间模型可以表示为

$$\begin{cases} x_{k+1} = A_d x_k + w_k, & A_d = e^{A\Delta t} \\ y_k = C x_k + v_k \end{cases} \quad (1)$$

式中, $x_k \in R^n$ 与 $y_k \in R^l$ 分别为系统的状态量与量测输出量; $w_k \in R^n$ 与 $v_k \in R^l$ 均为假定白噪声, 且满足 $E(w_k) = E(v_k) = 0$; $A \in R^{n \times n}$ 与 $C \in R^{l \times n}$ 分别代表系统状态矩阵与输出矩阵; Δt 为采样间隔。

SSI大致有两种不同的驱动方式, 即协方差驱动 (SSI-COV) 与数据驱动 (SSI-data)。相较于SSI-data, 由于SSI-COV在计算速度上更快, 因此更适合在线监测。

本文采用SSI-COV实现方式, 详细的实现方法参见文献^[17], 主要步骤: ①将量测数据提取出来构造Hankel矩阵, 并根据Hankel矩阵构造Toeplitz矩阵; ②对Toeplitz矩阵进行奇异值分解; ③将Toeplitz矩阵分解为控制矩阵与观测矩阵; ④依据②、③求得离散状态矩阵 A_d 。

对矩阵 A_d 进行特征值分解

$$A_d = \Psi \Lambda \Psi^{-1} \rightarrow \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \lambda_3 & \\ & & & \lambda_4 \end{bmatrix} \quad (2)$$

根据离散系统与连续系统特征值的关系得

$$\begin{aligned} \Lambda_c &= \ln(\Lambda) / \Delta t \\ \lambda_c &= a_c + j b_c \end{aligned} \quad (3)$$

由此可以推导出该系统的振荡频率 f_i 与阻尼比 ξ_i 分别为

$$\begin{aligned} f_i &= \sqrt{a^2 + b^2} / 2\pi \\ \xi_i &= -a / |\lambda_c| \end{aligned} \quad (4)$$

本文以发电机PMU的有功功率量测作为随机响应输入数据, 并使用SSI算法对随机数据进行辨识, 得到感兴趣的机电振荡模式所对应的频率、阻尼比, 以此作为异常值检测算法的训练数

数据集。

1.2 孤立森林算法

孤立森林具有集成学习的思想，能够集成多个随机二叉树，然后对每一个二叉树的结果进行综合决策，从而得到更加精确的结果，具体的过程为：

1) 通过有放回或无放回的方式，从训练数据集中均匀抽取 ψ 个样本作为子样本集。

2) 从子样本集中随机选取一个特征，并在该特征值的最大值和最小值之间随机产生一个数值 q 作为分割点，以此对样本集进行二叉划分，其中小于 q 的归置左侧，大于 q 的归置右侧。

3) 分别在新产生的左右数据子集中重复执行步骤2)，直到数据不能继续分割或者分割次数达到最大深度 $\log_2 q$ 时，停止分割。

生成完孤立树后，在数据样本 x 中遍历每一个孤立树，得出每一次的路径长度 $h(x)$ ，再记平均路径长度为

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n} \quad (5)$$

由 $h(x)$ 与 $c(n)$ 可以推导出每条待测数据的异常指数，其表达式为

$$S(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \quad (6)$$

式中， $E(h(x))$ 为样本 x 在所有孤立树中的路径长度的数学期望。

所得出的 $S(x, n)$ 的取值范围在0~1之间，当趋近于1时，认为是异常；当趋近于0时，认为是正常；当趋近于0.5时，认为无明显异常。

2 算法流程

随着现代信息技术与电网智能化的发展，很多发电厂和变电站都已安装广域测量系统(WAMS)，能够在时间和空间上可靠地收集到电网动态信息。本文利用SSI对广域测量系统收集到的随机响应数据进行辨识，在辨识的过程中发现辨识结果中存在密度稀疏，且距离正常点较远的异常点。因此应用孤立森林算法对辨识结果进行异常值检测，具体流程如图1所示。

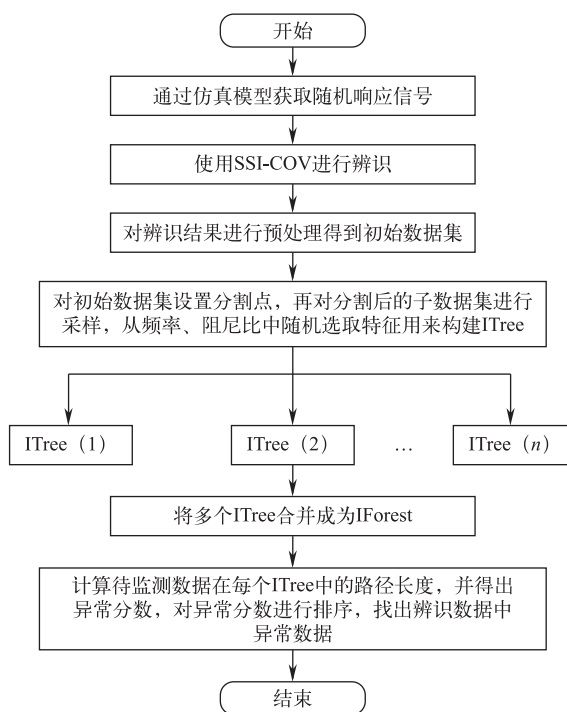


图1 算法流程图

Fig. 1 Algorithm flow chart

3 算例仿真

本文以IEEE 16机68节点系统的仿真数据进行分析，应用本文提出的算法在MATLAB提供的PST上进行仿真验证，选定基本运行方式，由小干扰稳定分析算法表明系统中存在15个机电振荡模式，包括4个对系统影响较大的区间振荡模式，见表1。

表1 SSAT 分析结果

Tab. 1 Analysis results of SSAT

模式	频率 /Hz	阻尼比 (%)
1	0.387 6	11.12
2	0.512 7	4.08
3	0.598 6	12.70
4	0.748 5	7.56

为了模拟出系统受到负荷波动后实际运行工况，本文在算例中的每个负荷加入了以基础值5%随机负荷波动。取16台发电机连续30 min的随机响应信号作为SSI算法的输入，考虑到随机响应数据本身的随机特性，采取滑动窗方式提取机电振荡参数，计算窗口以600 s处的数据点作

为起点, 1 800 s的数据点作为终点, 选择60 s的随机数据作为滑动分析窗口, SSI参数辨识结果见表2。

表 2 SSI 辨识结果

Tab. 2 Identification results of SSI

模式	频率 /Hz		阻尼比 (%)	
	均值	标准差	均值	标准差
1	0.412	0.028	11.529	0.053
2	0.513	0.018	5.934	0.046
3	0.609	0.039	12.651	0.068
4	0.755	0.029	8.553	0.068

由于箱型图既不需要对数据做出限制, 也不受异常值的影响, 并且可以直观地表现出数据的离散分布情况, 因此可以用来判断异常值的存在。对模式4下的辨识结果使用箱型图进行分析, 再比对原模式4的辨识数据图, 结果如图2所示。可以观察到部分数据点位于箱型图左右两侧, 即存在大于箱型图设定的上界和小于下界的数值, 因此可以判断出辨识数据存在异常值。

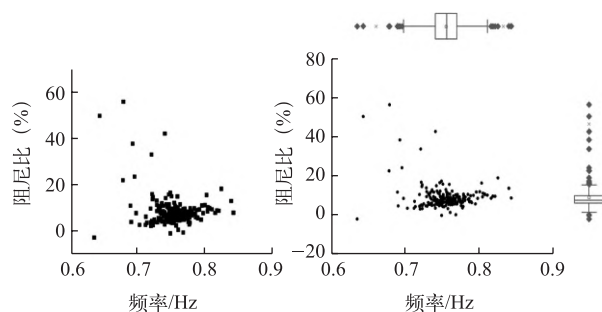


图 2 箱型图分析结果

Fig. 2 Analysis result of box plot

对每一种区间振荡模式下的辨识结果使用孤立森林算法进行异常值检测, 根据检测后得到的正常数据与异常数据绘制出四种模式下的数据离散分布图, 与原辨识数据进行对比, 结果如图3所示。离散分布图中正常数据和异常数据分别由黑色和红色表示。

通过比较原辨识数据图与异常处理后的数据离散分布图, 可以直观地发现孤立森林算法能够辨别出不同模式下异常数据, 将去除异常值后的数据进行整理, 得到的结果见表3。

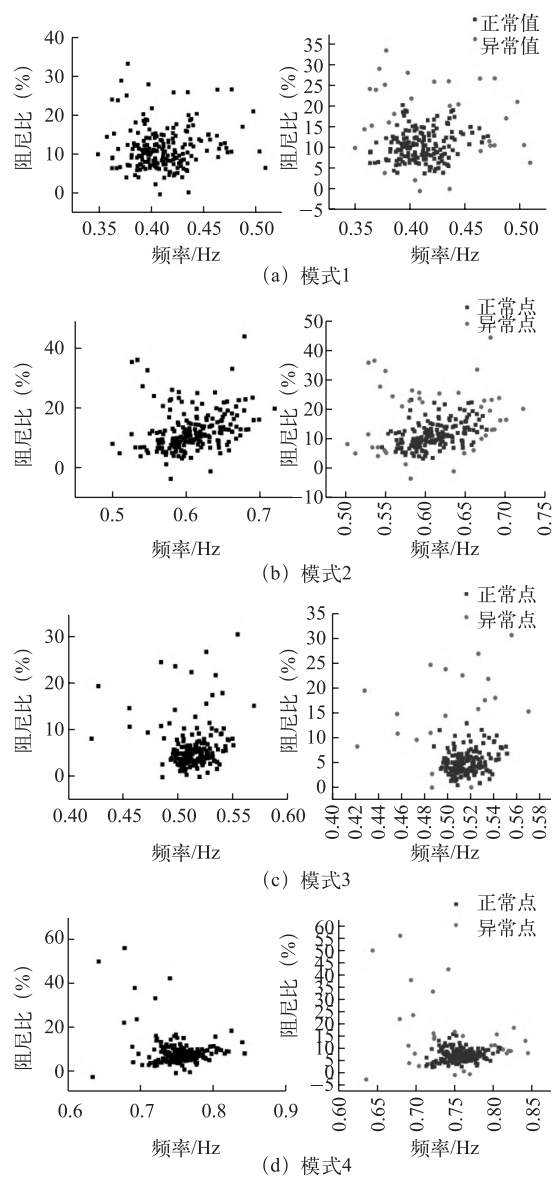


图 3 四种模式下的数据离散分布图

Fig. 3 Data dispersion distribution chart of four modes

表 3 异常检测后的辨识结果

Tab. 3 Identification results after anomaly detection

模式	频率 /Hz		阻尼比 (%)	
	均值	标准差	均值	标准差
1	0.410	0.021	10.562	0.036
2	0.515	0.012	4.891	0.022
3	0.609	0.029	11.423	0.043
4	0.756	0.019	7.234	0.025

在小干扰稳定特征参数中, 频率和阻尼比属于同一模式下的不同参数, 因此二者的标准差存在一定的差异。综合分析表2、表3可以发现,

当异常值存在时,标准差偏大,表明数据的浮动程度较大;经过孤立森林算法处理后,标准差减小,表明数据浮动程度明显降低,证明孤立森林算法能够有效处理振荡参数辨识后的异常数据。

4 结束语

针对提取随机响应数据中的小干扰稳定特征参数时存在异常值的问题,本文提出一种基于随机子空间和孤立森林的小干扰稳定评估方法。将该方法应用在IEEE 16机68节点系统中,实验结果表明通过孤立森林算法异常检测算法的引入,能够解决SSI辨识随机数据中的异常值问题,提高辨识结果准确度,利用本算法进行小干扰稳定评估,可以为电网运行及监测提供可靠的依据。

参考文献

- [1] 王健,于文娟,汤卫东,等.基于改进Prony算法的华中电网低频振荡在线监测分析系统研究[J].电气应用,2016,35(16):68-73.
- [2] YANG Deyou, WANG Lixin, CAI Guowei, et al. Synchronized ambient output-only based online inter-area transfer capability assessment considering small signal stability[J]. IEEE Transactions on Power Systems, 2021, 36(1): 261-270.
- [3] 曲欣,尚光伟,曹锐,等.基于相量测量装置的低频振荡在线监视系统研究[J].电气应用,2010,29(24):54-56.
- [4] 蔡国伟,杨德友,张俊丰,等.基于实测信号的电力系统低频振荡模态辨识[J].电网技术,2011,35(1):59-65.
- [5] 陈刚,何潜,段晓,等.电力系统低频振荡分析与抑制综述[J].南方电网技术,2010,4(3):17-22.
- [6] 谢剑,成业,王晓茹.基于NExT和PRCE方法的低频振荡分析[J].电工技术学报,2018,33(1):121-130.
- [7] 高晗,蔡国伟,杨德友,等.基于NExT-DMD的电力系统机电振荡参数提取[J].电网技术,2022,46(1):284-291.
- [8] DOSIEK L, PIERRE J W. Estimating electromechanical modes and mode shapes using the multichannel ARMAX model[J]. IEEE Transactions on Power Systems, 2013, 28(2): 1950-1959.
- [9] 李倩,韩斌,汪旭洋.基于模糊孤立森林算法的多维数据异常检测方法[J].计算机与数字工程,2020,48(4):862-866.
- [10] 赵臣啸,薛惠锋,王磊,等.基于孤立森林算法的取用水量异常数据检测方法[J].中国水利水电科学研究院学报,2020,18(1):31-39.
- [11] 吴志强,张胜,包晓玲,等.针对WSN异常数据检测改进的孤立森林方法[J].小型微型计算机系统,2021,42(1):127-131.
- [12] 李星南,施展,亢中苗,等.基于孤立森林算法和BP神经网络算法的电力运维数据清洗方法[J].电气应用,2018,37(16):72-79.
- [13] 周军,巩森,吴瑜坤,等.基于IForest和SVD的直流传感器去噪方法[J].电气应用,2022,41(11):91-96.

作者简介

贺潇瑞(1996—),男,硕士研究生,研究方向为电力系统稳定与控制。

(收稿日期:2023-01-28)