

基于支持向量机的电力系统状态估计 多类型数据异常检测

郭嘉辉 侯月婷 丁磊 金朝阳

(山东大学电网智能化调度与控制教育部重点实验室 济南 250061)

摘要:为了解决异常数据严重影响电力系统状态估计性能的问题,提出了一种基于支持向量机(SVM)的电力系统预测辅助状态估计(FASE)多类型数据异常检测方法。首先,针对传统 FASE 的预测准确率欠佳的问题,提出了基于极限学习机的 FASE 方法,并利用 SVM 并基于预测数据、量测数据与估计值,实现了对坏数据、负荷突变和单相接地等多种类型的数据异常检测。其次,针对惩罚因子和核函数参数会影响分类精度的问题,提出采用灰狼算法对 SVM 参数进行优化,在兼顾计算速度的同时提高了数据异常检测的准确率。最后,在 IEEE 33 和丹麦 DTU 7K 47 节点主动配电网系统上进行仿真测试,所提方法在正常工况下提升 26.08%与 26.76%,计算速度提升 46.05%,在数据异常情况下准确率综合提升 32.04%与 29.27%,结果表明,所提方法具备较强的通用性与实时性,可以有效地检测电力系统中各种类型的数据异常,并提高状态估计的性能。

关键词:预测辅助状态估计;异常检测;极限学习机;支持向量机;灰狼算法

中图分类号: TM73 **文献标识码:** A **国家标准学科分类代码:** 470.4054

Multi-type data anomaly detection in power system state estimation using support vector machine

Guo Jiahui Hou Yueting Ding Lei Jin Zhaoyang

(Key Laboratory of Power System Intelligent Dispatch and Control, Shandong University, Jinan 250061, China)

Abstract: To address performance issues in state estimation caused by anomalous data in power systems, this paper proposes a multi-type data anomaly detection method for power system state estimation based on support vector machine (SVM). Firstly, extreme learning machine is proposed to enhance prediction accuracy in forecasting-aided state estimation (FASE). Then, SVM is utilized to detect various types of data anomalies, including bad data, load changes, and single-phase grounding, by incorporating prediction data, measurement data, and estimated values. To optimize SVM parameters and improve classification accuracy, the grey wolf algorithm is suggested to tackle the issue of penalty factor and kernel function parameters. Finally, simulations are conducted on real distribution systems, specifically the IEEE 33 and DTU 7K 47 systems, using data from different scenarios. The proposed method achieves accuracy improvement of 26.08% and 26.76% and calculating speed improvement of 46.05% under normal operating conditions, and a comprehensive enhancement of 32.04% and 29.27% in accuracy under data abnormality scenarios. The results demonstrate that the proposed method is highly generalizable and real-time, effectively detects various types of data anomalies, and enhances the performance and accuracy of state estimation.

Keywords: forecasting-aided state estimation; anomaly detection; extreme learning machine; support vector machine; grey wolf algorithm

0 引言

随着电力系统规模的不断扩大和电力需求的增加,确

保电力系统的经济性和安全性变得尤为重要。为了实现这一目标,现代电力调度系统需要全面、快速地了解电力系统的实时运行情况,并准确分析和预测其运行趋势。其

中,电力系统状态估计^[1]起着重要的作用,它通过对电力系统实时运行数据的检测,以提供有效的信息。目前,最具前景的状态估计方法是预测辅助状态估计(forecasting-aided state estimator, FASE)。目前,传统的 FASE 方法有基于卡尔曼框架的滤波器和霍尔特双参数线性指数平滑法^[2]等。然而,随着可再生能源的渗透率不断提升,电网潮流波动性大幅上升,导致传统的 FASE 方法精度下降明显。近年来,基于机器学习的 FASE 方法不断被提出,如人工神经网络^[3](artificial neural networks, ANN)、支持向量回归^[4](support vector regression, SVR)和长短期记忆算法^[5](long short-term memory, LSTM),以上方法在面临趋势大幅潮流波动性时仍能保持较高的预测精度。然而,其缺点在于训练时间较长,建立的预测模型相对复杂,难以适应电力系统在线状态估计的需求。基于此,亟需寻找一个准确且快速的机器学习预测方法。

另一方面,尽管基于机器学习的 FASE 在提高估计精度方面有诸多的优势,但电力系统中经常遭遇各种数据异常情况,如坏数据、负荷突变和故障等,这会严重影响 FASE 的估计精度和鲁棒性。故需要异常检测来对 FASE 的输出结果进行检测与校正,这对保证电力系统的安全稳定运行具有重要意义。对此,海内外学者做出了不懈的努力,提出利用偏度和具有固定阈值的残差(滤波估计量测误差)来检测和区分坏数据与负荷突变,但基于偏度的异常区分方法无法在大电网中准确区分两种数据异常,固定阈值的情况也并不具有普遍适用性。此后,学者们提出采用具有较高鲁棒性的状态估计方法来抑制坏数据与负荷突变对估计精度的影响^{[6][7]}。并且为了最大程度地消除数据异常的影响,需要对不同类型的数据异常进行准确的检测与识别。因此,Valverde 等^[8]提出利用偏度新息(预测量测误差)来区分坏数据与负荷突变,较固定阈值残差与偏度的方法在大电网中分辨精度有了显著提升。Geetha 等^[9]提出了固定阈值残差结合固定阈值新息的数据异常区分方法,其区分效果较利用偏度新息又有了显著的提升,但是依然无法区分故障带来的数据异常。Jin 等^[10]利用 FASE 的残差新息比和静态估计器的归一化残差,提出了一种新的异常检测框架,能够较为精确地区分坏数据、负荷突变以及故障等 3 种数据异常,且对不同级别的噪声、不同程度的测量冗余和不同规模与复杂性的网络都具有鲁棒性。然而由于缺乏较好的阈值设定方法,其基于历史数据设定的平均区分阈值可能与实际的阈值有所出入。

综上所述,针对电力系统状态估计数据异常严重影响状态估计性能的问题,提出了一种基于支持向量机(support vector machine, SVM)的电力系统预测辅助状态估计(FASE)多类型数据异常检测方法。首先,提出基于极限学习机(extreme learning machine, ELM)的 FASE 方法,在大幅提升预测精度的同时保证了估计的实时性。

在此基础上,基于 SVM 从预测数据、量测数据以及估计结果中提取不同数据异常的特征,从而精确地检测与区分包括坏数据、负荷突变、单相接地故障等多种数据异常。最后,针对惩罚因子和核函数参数会影响分类精度的问题,提出采用灰狼算法(grey wolf algorithm, GWO)优化支持 SVM 参数,在兼顾计算速度的同时提高了数据异常检测的准确率。在 IEEE 33 和丹麦 DTU 7k 47 节点主动配电网系统上进行的仿真测试,结果表明本文所提方法具备较强的通用性与实时性,可以有效地检测电力系统中各种类型的数据异常,并提高状态估计的性能。

1 基于 ELM 的预测辅助状态估计

1.1 ELM 预测模型

FASE 的预测精度对异常检测准确率有较大的影响。然而,传统方法依然受到电网中量测设备精度以及配置的限制,而人工智能算法能通过非线性学习在一定程度上解决传统方法在异常值检测中的缺陷。ELM^[11]是一种高效的机器学习算法,在很大程度上改进了传统的反向传播算法(back propagation, BP)。

ELM 算法的核心思想是通过随机生成的权值、阈值和矩阵的广义逆理论(Moore-Penrose)求解输出权值参数,从而得到最终的输出结果,其网络结构如图 1 所示。

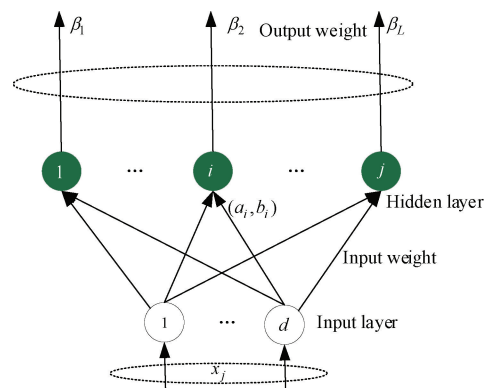


图 1 ELM 模型示意图

Fig. 1 Schematic diagram of ELM model

相对于传统的单隐层前馈神经网络来说,ELM 的网络结构、训练过程更为简单,并且具有更强的泛化能力。该神经网络中包含 L 个隐含层节点,对于输入向量,网络的输出可以表示为:

$$f_L(x) = \sum_{i=1}^L \beta_i g(\omega_i x + b_i) \quad (1)$$

式中: $\omega_i, b_i \in R^n$ 均为单隐层神经网络隐含层中参数, β_i 为连接第 i 个隐含层节点和输出层节点的权值; ω_i 为连接输入层和隐含层第 i 个节点的权值; b_i 为隐含层第 i 个节点的偏置, $g(\cdot)$ 为激活函数。

将式(1)改写为矩阵方程:

$$H\beta = Y \quad (2)$$

式中： \mathbf{H} 为极限学习机的隐含层输出矩阵； \mathbf{Y} 为期望输出向量。

输出权重可表示为：

$$\beta^* = \mathbf{H}^+ \mathbf{Y} \quad (3)$$

式中： \mathbf{H}^+ 为 \mathbf{H} 的 Moore-Penrose 广义逆矩阵。与 BP 神经网络相似，ELM 的训练过程也需要计算权重矩阵，但不同的是，ELM 所需的权重矩阵是逆矩阵。因此，对于 ELM 算法，只需要确定隐层神经元的数量即可得到最优解。

1.2 扩展卡尔曼滤波器(EKF)

卡尔曼滤波在实际应用中要求系统为线性系统，而实际情况下大多数系统为非线性系统，EKF 则较好地解决了非线性系统状态估计的问题。

EKF 的基本原理是将非线性方程分别围绕状态变量滤波值和预测值进行泰勒线性展开，并将二次及以上项省略，从而使得非线性系统线性化，之后利用卡尔曼滤波进行状态估计。具体推导过程参考文献[12]。

1.3 ELM-EKF 预测辅助状态估计

传统 EKF 假设系统的状态和测量误差均为高斯分布，对非高斯噪声分布的适应性较差，且线性化的过程中需要大量的计算，复杂耗时。

针对这些问题，ELM-EKF 在传统的 EKF 方法上做了改进，使其具有了快速学习的能力，ELM-EKF 可以更快速地学习系统的非线性特征，从而提高状态估计的效率和速度，且 ELM 使用随机生成的隐藏层权重和偏置来进行非线性映射，避免了传统 EKF 线性化的复杂计算。通过 ELM 的非线性映射能力，ELM-EKF 可以更好地适应系统非线性的变化，对于模型误差和噪声的影响具有一定的鲁棒性，能够提供更稳定和准确的状态估计结果。

2 基于 SVM 的状态估计多类型数据异常检测方法

针对电力系统的分布特性，传统的异常数据检测辨识方法根据状态估计的不同阶段，可分为 3 类：估计计算前的检测辨识、估计计算中的检测辨识和估计计算后的检测辨识。

1) 估计计算前的检测辨识，避免重复状态估计带来的繁琐计算，其灵活性强，适用性广。但在理论体系与方法完备性方面有待完善。

2) 估计计算中的检测辨识，可同时检测到多个异常数据，效率较高。然而所得结果未达最优。

3) 估计计算后的检测辨识，检测辨识效果好。但计算量大，残差淹没将导致检测灵敏度降低。

利用残差新息比(innovation reduction factor, Inred)的异常检测框架^[10]，能够检测到由线路中断引起的拓扑突然变化。该框架基用残差新息比判别坏数据与状态突变，其定义为：

$$\text{Inred} =$$

$$\lim_{\max(\epsilon) \rightarrow \infty} \frac{\max(|\mathbf{z}_{k,\text{true}} + \boldsymbol{\epsilon} - h(\hat{\mathbf{x}}_{k|k})| / \sqrt{\mathbf{P}_{k|k}^{\text{ss}}})}{\max(|\mathbf{z}_{k,\text{true}} + \boldsymbol{\epsilon} - h(\hat{\mathbf{x}}_{k|k-1})| / \sqrt{\mathbf{P}_{k|k-1}^{\text{ss}}})} \quad (4)$$

式中： $\mathbf{z}_{k,\text{true}}$ 为真实量测向量； $\boldsymbol{\epsilon}$ 为误差； $\mathbf{P}_{k|k}^{\text{ss}}$ 和 $\mathbf{P}_{k|k-1}^{\text{ss}}$ 分别为 $\mathbf{z}_{k,\text{true}} + \boldsymbol{\epsilon} - h(\hat{\mathbf{x}}_{k|k})$ 和 $\mathbf{z}_{k,\text{true}} + \boldsymbol{\epsilon} - h(\hat{\mathbf{x}}_{k|k-1})$ 的协方差矩阵。

Inred 具有如下属性：

Inred $\rightarrow 1$ ；SCADA 坏数据

Inred $\rightarrow (0, 1]$ ；PMU 坏数据

Inred $\rightarrow 0$ ；负载突然变化

Inred > 0 ；故障状态

由于该框架缺乏较好的阈值设定方法，为了提升数据异常的检测精度，不仅需要提出一个自动化的阈值设定方法，还需要更精确的预测分类率，另外需要增加对数据的判别类型。

SVM^[13] 是一个很好的解决方法，其目标是找到一个能够使间隔最大化的决策超平面，以达到更好地分类新样本的泛化能力。

2.1 SVM 分类的核函数

数据异常的复杂性使得低维空间中无法对异常进行线性区分。为解决此问题，可将数据转化到高维空间，使其线性可分。核函数为映射关系 $\phi(x)$ 的内积，在映射中有关键作用。通过构造增维的核函数，有助于更好地区分数据。分类中常用核函数将二维数据映射至三维或更高维，可有效应对线性不可分的数据异常，提升异常数据的识别和分类效果。

常用的核函数有线性核函数、多项式核函数、径向基核函数(radical basis function, RBF)和 Sigmoid 核函数。本文采取 RBF 核函数，表达式为 $k(x, x_i) = \exp(-\gamma |x - x_i|^2)$ ，其中 γ 为参数。具备适应低维、高维、小样本和大样本等情况的能力，并且具有较宽的收敛域，可在不同数据集和情境下取得良好分类效果，是一种理想的分类函数。

SVM 的目标函数和约束条件表示为：

$$\min \frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \sum_{i=1}^l \xi_i \quad (5)$$

$$\text{s. t. } \mathbf{y}_i (\boldsymbol{\omega}^T \mathbf{x}_i + b) \geq 1 - \xi_i \quad (6)$$

式中： $\boldsymbol{\omega}$ 、 \mathbf{x} 为向量； b 为常数； l 为样本数； \mathbf{y}_i 表示分类对象的属性； ξ_i 为非负松弛项； C 为惩罚系数。

SVM 的分类原理如图 2 所示。

2.2 参数优化原理

SVM 的区分精度受 C 值、 γ 值等参数的影响较大，故需要针对此进行参数优化。模型参数优化是通过极小化目标函数使得模型输出和实际观测数据之间达到最佳的拟合程度的过程。由于情景与模型本身的复杂性，常规优化算法难以达到参数空间上的全局最优。

给定训练样本集 $T = \{(x_i, y_i)\}_{i=1}^n \in (\mathbf{X} \times \mathbf{Y})^n$ ，其中

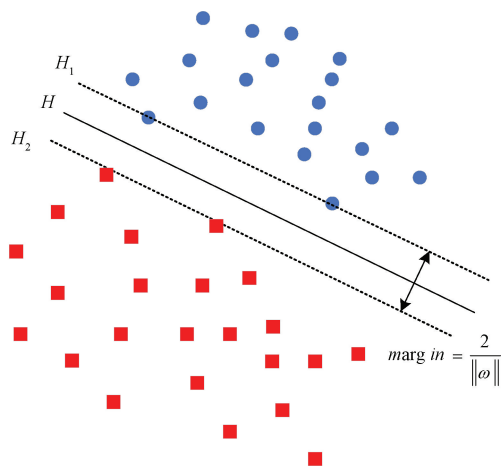


图2 SVM分类原理

Fig. 2 Principle diagram of SVM classification

$x_i \in \mathbf{X} = R^n, y_i \in \mathbf{Y} = R$, 最优超平面可以通过给定的正则化参数 C 值求解以下凸规划问题, 即求解式(7)的最小值^[15]。

$$\phi(\omega) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n L(f(x_i), y_i) \quad (7)$$

式中: C 为惩罚系数, 即对误差的宽容度; $L(\cdot)$ 表示损失函数。 C 越高, 说明越不能容忍出现误差, 容易发生过拟合。 C 越小, 容易发生欠拟合。 C 过大或过小, 都会导致泛化能力下降。

通过核函数把非线性数据映射到高维空间, 在高维空间进行线性回归。 其二次规划优化形式可以转化为:

$$\min W(\alpha, \alpha^*) = \frac{1}{2} \sum_{i=1, j=1}^n (\alpha_i, \alpha_j^*)^T \kappa(x_i, x_j) (\alpha_i, \alpha_j^*) + \epsilon \sum_{i=1}^n (\alpha_i, \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i, \alpha_i^*) \quad (8)$$

式中: $\kappa(x_i, x_j)$ 为满足 Mercer 条件的核函数, 通常应用 RBF 核函数。 γ 值为 RBF 核函数的固有参数, 决定数据映射到新的特征空间后的分布, γ 与支持向量的个数呈反比。 而支持向量的个数影响训练与预测的速度。

由以上分析可知, 为使 SVM 具有良好的应用效果, 需对其正则化参数 C 、核参数及 γ 值进行优化。

GWO 是一种全局随机搜索算法, 模拟了自然界中灰狼的领导层次结构和捕猎机制^[14]。 通过个体狼与猎物之间的距离来评估适应度。 根据适应度值, 选择并命名 3 个最佳狼为 α 、 β 和 δ 。 由这 3 只狼领导的 ω 只狼代表种群中其余的解。 α 、 β 和 δ 的更新迭代被称为狩猎。 围捕行为建模如下:

$$\mathbf{X}_{(t+1)} = \mathbf{X}_{prey(t)} - A \cdot |C \cdot \mathbf{X}_{prey(t)} - \mathbf{X}(t)| \quad (9)$$

式中: A 和 C 为随机系数; $\mathbf{X}_{prey(t)}$ 和 $\mathbf{X}(t)$ 分别代表猎物和狼当前迭代的位置向量; t 表示当前迭代次数。 A 和 C 可以通过如下方程得到:

$$\begin{cases} A = 2a \cdot r_1 - a \\ C = 2 \cdot r_2 \\ a = 2(1 - t/T_{max}) \end{cases} \quad (10)$$

式中: r_1 和 r_2 都是在 $[0, 1]$ 的随机向量; a 为收敛因子, 在优化过程中线性地从 2 减少到 0; T_{max} 为最大迭代次数。 在狩猎过程中, 3 只个体狼, α 、 β 和 δ , 靠近猎物, 其他狼跟随它们。 因此, \mathbf{X}_α 、 \mathbf{X}_β 和 \mathbf{X}_δ 被视为猎物位置并插入到式(10)中, 得到更新方程:

$$\begin{cases} \mathbf{X}_{1(t+1)} = \mathbf{X}_{\alpha(t)} - A_1 \cdot |C_1 \cdot \mathbf{X}_{\alpha(t)} - \mathbf{X}(t)| \\ \mathbf{X}_{2(t+1)} = \mathbf{X}_{\beta(t)} - A_2 \cdot |C_2 \cdot \mathbf{X}_{\beta(t)} - \mathbf{X}(t)| \\ \mathbf{X}_{3(t+1)} = \mathbf{X}_{\delta(t)} - A_3 \cdot |C_3 \cdot \mathbf{X}_{\delta(t)} - \mathbf{X}(t)| \end{cases} \quad (11)$$

$$\mathbf{X}_{GWO(t+1)} = \frac{\mathbf{X}_{1(t+1)} + \mathbf{X}_{2(t+1)} + \mathbf{X}_{3(t+1)}}{3} \quad (12)$$

式中: $\mathbf{X}_{GWO(t+1)}$ 为灰狼的平均位置, 根据 3 只个体狼 α 、 β 和 δ 的位置进行更新。

使用 GWO 算法可以很好的对 SVM 的各项参数进行优化, 达到参数空间上的全局最优。

3 基于 SVM 的电力系统状态估计多类型数据异常检测算法的基本原理

本文提出的基于 SVM 的电力系统状态估计多类型数据异常检测算法, 算法流程如图 3 所示。 利用 ELM 进

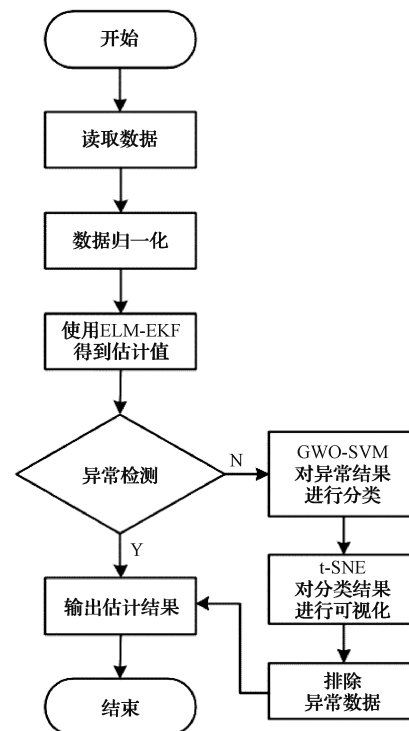


图3 基于支持向量机的电力系统状态估计多类型数据异常检测算法流程

Fig. 3 Flowchart of multi-type data anomaly detection in power system state estimation using support vector machine

行潮流预测,结合使用 GWO 优化参数的 SVM 检测数据异常。最后使用 t-SNE^[15] 进行数据降维,增强其分类可视性。

具体步骤如下:

- 1) 使用 Z-Score 算法对输入的量测数据进行归一化。
- 2) 使用 ELM 进行潮流预测,结合 EKF 得到估计值。
- 3) 使用 Inred 进行异常检测,若发现异常,则使用 GWO-SVM 进行异常结果分类。
- 4) 使用 t-SNE 对分类结果进行可视化。
- 5) 排除异常数据,得到最终状态估计结果。

4 算例分析

为验证提出的基于支持向量机的电力系统状态估计多类型数据异常检测方法的有效性,本章节选取 IEEE 33 和丹麦 DTU 7K 47 节点主动配电网系统点测试系统对所采用的 SVM 分类器的性能进行分析,故障为线路接地故障,分为单相接地、双相接地、三相接地故障。所有的测试均基于 MATLAB 环境实现,测试平台为配置有英特尔酷睿 i7-9750CPU(@2.6 Hz)16 GB 内存的台式机。为了保证结果的一般性,所有结果均为 500 次蒙特卡洛模拟结果的平均值。

对于量测噪声,根据量测的类型,将其标准偏差设置如下。

SCADA: P、Q 量测值(0.02%),电压(0.002%);

PMU: 幅值(0.002%),相角(0.01%)。

4.1 正常工况下基于机器学习的预测辅助状态估计仿真测试

机器学习方法在预测方面有广泛的应用,不同的机器学习预测方法有其各自的优点和缺点,需要根据具体问题的特点选择合适的方法。在 IEEE 33 节点系统下,仿真估计器采用 EKF,预测器采用 SVR、BP、ELM、LSTM 4 种不同方法作对比,比较其预测精度及仿真时间(包括训练时间和预测时间),为表示简便,下文省去“—EKF”。

SVR、BP、ELM 和 LSTM 为机器学习性能较好的回归算法。其中,SVR 适用于高维、非线性数据,训练时间较长,训练模型简单且精度较高。BP 神经网络适合处理非线性数据,但容易陷入局部最优解,故需进行多次试验。ELM 学习速度快,对于高维、非线性数据表现优秀,但模型解释性尚需提升。LSTM 适合处理序列数据,但对于高维数据的训练复杂度较高,需要进行适当的超参数调优。因此,在实际应用中需要根据具体问题和数据集的特点选择合适的方法。相较于 BP、SVR、LSTM,ELM 对于高维数据和非线性数据表现良好且泛化能力强,避免了过拟合的问题。

本文使用 3×134 数据集,采用 110 组数据做训练集,24 组做预测集。

SVR 选择 RBF 核函数,惩罚系数 C 和核函数参数 g

用网格搜索法,网格设置为 $[-10,10]$,步长为 0.5。

BP 输入层维度为 3×110 ,输出层维度为 1×24 ,隐藏层节点数为 10。激活函数默认为 sigmoid 函数,权重默认为随机初始化,训练次数为 1 000 次,训练目标最小误差为 0.001,学习速率为 0.1。

ELM 隐藏层节点数为 24,激活函数为 sigmoid 函数,权重随机初始化。

LSTM 求解器设置为 adam,125 次迭代,梯度阈值设置为 1,指定初始学习率 0.005,在 25 轮训练后通过乘以因子 0.2 来降低学习率。

在上述设置条件下预测效果如图 4 所示,统计其仿真耗时如表 1 所示。

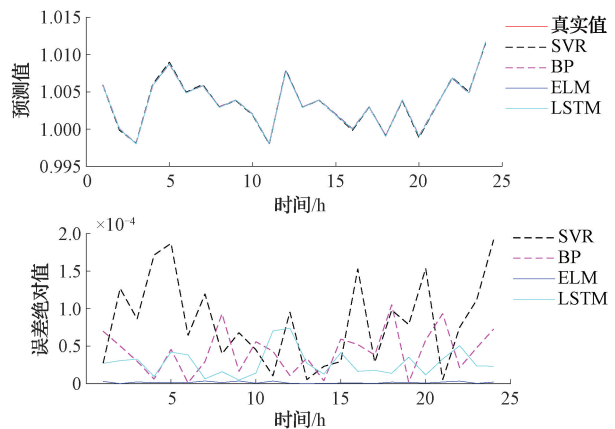


图 4 SVR、BP、ELM、LSTM FASE 效果对比

Fig. 4 Comparison chart of SVR, BP, ELM, and LSTM FASE performance

表 1 SVR、BP、ELM、LSTM 仿真时间比较

Table 1 Comparison of calculation time for SVR, BP, ELM, and LSTM

算法	训练耗时/s	估计耗时/s	准确率/%
SVR	2.011 0	0.003 7	96.32
BP	0.142 0	0.008 7	97.34
ELM	0.000 6	0.003 3	98.13
LSTM	3.626 7	0.009 5	97.68

由图 4 和表 1 可知,对于 SVR,通常在大数据集上训练时间较长,但在小数据集上表现较好,且预测时间较短。BP 神经网络的训练时间通常较长,且容易陷入局部最优解,需要进行多次试验来避免,但预测时间较短。对于 LSTM 神经网络,其适合处理序列数据,但对于训练复杂度较高的高维数据,需要进行适当的超参数调优,训练时间通常较长,但预测时间相比于训练时间较短。ELM 的训练时间较短,因为 ELM 对超参数调优的依赖性较低,模型的建立和训练过程更加简化,故预测时间较短,且准确性较高,相较于其他 FASE 方法估计速度综合提升 46.05%。

4.2 参数优化方法比较

不同的惩罚因子和不同的核函数会影响异常检测精度,本文采用7种不同的优化方法对SVM参数进行优化,根据不同的节点系统对其精度进行验证,各优化方法准确率如表2所示,所有数据均采用文献[11]中数据。

表2 不同优化方法在不同数据类型下异常检测准确率对比

Table 2 Comparison of anomaly detection accuracy with different optimization methods across various data types (%)

算法	IEEE 14	IEEE 39	IEEE 57
SVM($C=2, g=1$)	97.51	97.06	97.66
GS-SVM	98.59	98.71	99.33
PSO-SVM	98.67	98.92	99.19
GA-SVM	98.41	99.51	99.05
GWO-SVM	99.33	99.78	99.78
ABC-SVM	99.06	99.73	99.79
FA-SVM	98.67	99.57	99.50
CS-SVM	99.12	99.03	98.99

表2中量测数据均添加高斯噪声,而量测噪声方差是基于假设的计量精度和关于测量误差的历史数据来选择的,并非所有的量测噪声都能严格用高斯分布表示。故在IEEE 14节点系统数据中添加拉普拉斯噪声,再次计算各优化方法准确率,结果如表3所示。

表3 不同优化方法在不同噪声类型下异常检测准确率与耗时对比

Table 3 Comparison of anomaly detection accuracy and computational time with different optimization methods under various noise types

算法	含高斯噪声 准确率/%	含拉普拉斯 噪声 准确率/%	异常检测 耗时/s
SVM ($C=2, g=1$)	96.32	96.14	0.002 9
GS-SVM	97.91	97.36	0.004 6
PSO-SVM	99.27	97.51	0.004 5
GA-SVM	99.27	99.27	0.003 8
GWO-SVM	99.71	98.78	0.003 2
ABC-SVM	99.69	98.70	0.003 7
FA-SVM	99.51	98.68	0.004 8
CS-SVM	99.23	98.64	0.005 3

从以上对比结果可知,相比于其他方法,GWO-SVM优化算法兼顾了精度与计算速度。

4.3 异常检测仿真案例

综合4.1节与4.2节的仿真结果,本文提出的优化方法相较于其他6种优化方法有着较高的准确率与较快的计算速度。将本文提出的异常检测方法与其他3种现有异常检测方法进行比较。

在IEEE 14节点系统仿真中,定义坏数据误差率为量测数据与真实数据差值的绝对值占真实数据值的比例。首先设置基准配置条件,在一组数据中随机选择一个数据置零,即本数据的坏数据误差率设置为100%,负载变化率设置为60%,即负载变化为原负载的40%;所有数据中添加高斯噪声;数据包含预测数据。之后设置在基准配置条件中参数改变时的仿真实例,包括不同负荷变化率、不同误差率数据、添加非高斯噪声时准确率的变化。

1) 基准配置仿真实例

对于IEEE 14配网量测,在其中随机添加90组坏数据,误差率为100%(即在量测数据中随机选择一个数据置零)的情况,分类效果如图5所示,统计其分类准确率如表4所示。

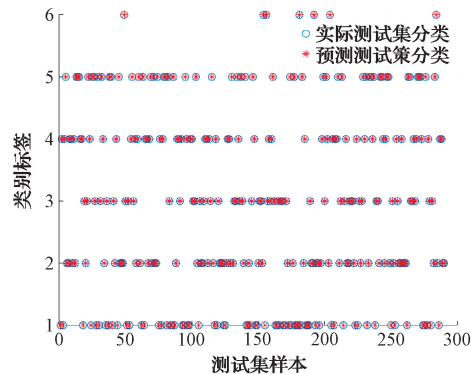


图5 IEEE 14节点数据分类结果分布

Fig. 5 Distribution of data classification results in IEEE 14-node system

表4 IEEE 14节点系统不同误差率的数据异常分类准确率
Table 4 Classification accuracy of data anomalies in IEEE 14-Bus system under different error rates

数据类型	准确率/%
正常	96.53
单相故障	99.72
两相故障	99.83
三相故障	99.91
负载突变	95.18
坏数据	99.87

由GWO-SVM方法训练,由t-SNE降维至二维,得到散点图如图6所示,再由SVM映射至高维空间分类

后在二维空间绘制分类边界,根据类别着色。t-SNE 分类仅将数据分类可视化,并不影响分类准确率。其中的散点图边界明显,所有数据分类区域独立,同类数据空间距离较小且各数据簇紧凑,表明 SVM 分类效果良好。

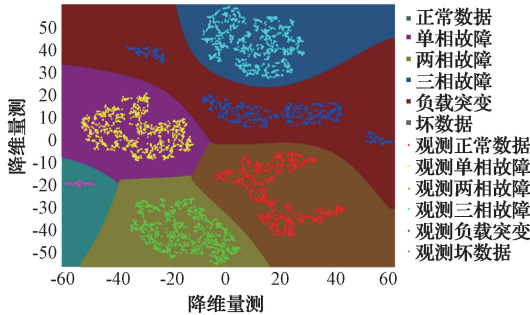


图 6 RBF 数据分类

Fig. 6 RBF data classification diagram

将本文所提方法与现有的其他 3 种方法的检测精度进行比较,检测准确率如表 5 所示。

表 5 本文方法与其他方法检测检测准确率对比

Table 5 Comparison of detection accuracy between proposed method and other methods

算法	正常数据 准确率/%	负载突变 准确率/%	坏数据 准确率/%	线路故障 准确率/%
文献[8]	70.5	57.02	17.75	4.84
文献[9]	70.5	11.94	99.81	4.84
文献[10]	95.5	96.32	99.57	100
本文	97.58	99.52	99.92	100

结果表明,文献[8]利用归一化新息和归一化新息的偏度来检测异常,在负荷突变(suddenly load change, SLC)情况下,其检测精度较低,因当 SLC 较小时,归一化残差的偏度会显著增加。文献[9]采用归一化新息和归一化残差来区分 SLC 和坏数据情况。然而,在 SLC 情况下,残差值较大,导致 SLC 和坏数据情况难以进行有效区分。本文所提方法由文献[10]中同样的数据类型验证其准确率,除 0.48% 的 SLC 和 0.08% 的坏数据未被检测到外,其余故障坏数据均被本文方法正确检测到。与此对比,文献[8-9]的误检率较高,文献[10]的误检率较低。本文所提出的方法误检率最低且不需要设定阈值,拥有较高的泛化性。因此,该方法同样适用于在线异常检测。

2)不同坏数据误差率对状态分类影响

基准配置条件下的仿真结果是在坏数据误差率为 100% 的条件下进行的,对于实际情况,其误差率并不确定。在量测噪声概率密度函数服从标准高斯分布情况下,本文通过引入不同坏数据误差率来比较各算法的分类准

准确率,测试结果如表 6 所示。

表 6 不同数据异常误差率下的检测准确率对比

Table 6 Comparison of detection accuracy under different data anomaly error rates

算法	不同坏数据误差率下的准确率/%				
	20%	40%	60%	80%	100%
文献[8]	56.51	60.89	65.37	66.28	67.92
文献[9]	59.36	62.57	66.23	66.34	68.25
文献[10]	91.83	93.21	96.23	98.97	99.32
本文	94.37	95.79	97.13	99.06	99.58

由表 6 可知,随着误差率的增大,分类准确率越来越高。但即使误差率较低时,所提出方法分类仍能保持较高准确率。

3)非高斯噪声对状态分类的影响

上述仿真数据均为添加高斯噪声时统计的状态分类准确率,为验证本文所提出的 SVM 分类器的普遍适用性,在数据中添加拉普拉斯噪声,以验证分类准确率,结果如表 7 所示。

表 7 含拉普拉斯噪声的数据异常分类准确率对比

Table 7 Comparison of classification accuracy of data anomalies with Laplace noise

数据类型	文献[8]	文献[9]	文献[10]	本文
正常	67.88	68.35	93.26	94.23
单相故障	59.54	72.07	97.29	99.12
两相故障	60.27	75.85	97.87	99.23
三相故障	62.89	74.24	98.46	99.16
负载突变	55.69	60.65	94.73	97.97
坏数据	65.36	99.35	99.32	99.67

由表 7 可知,在数据添加拉普拉斯噪声时,虽相对于高斯噪声准确率略微下降,但 SVM 分类依然保持较高的准确率。

4)IEEE 33 节点与丹麦 DTU 7K 47 节点主动配电网系统仿真实例比较

为验证所提出的 SVM 分类器在电力系统特别是含有规模化分布式电源的电力系统下的普遍适用性,考虑在 IEEE 33 节点与丹麦 DTU 7K 47 节点主动配电网系统进行测试。

丹麦 DTU 7K 47 节点主动配电网系统结构如图 7 所示,该系统有 3 个风电场,由第 4 代可控风机组成,装机容量分别为 12、15 和 15 MW。

量测数据来自 PMU 和 SCADA,量测配置如表 8 所示。计算基于标么值,基准电压为 VB=10 kV,基准容量为 Sb=10 MVA。系统每 15 min 进行一次状态估计计算。

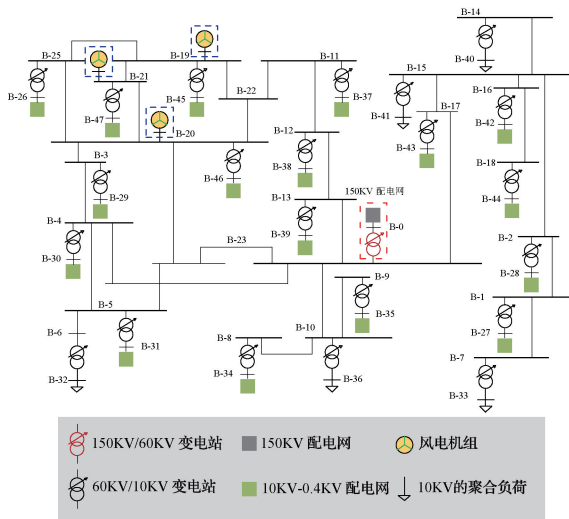


图7 丹麦DTU 7K 47节点主动配电网系统
Fig.7 Danish DTU 7K 47-node active distribution network system

表8 DTU 7K 47节点测试系统量测配置

Table 8 Measurement configuration of DTU 7K 47-node system

量测设备	量测量	误差/%	放置位置
SCADA	有功功率	0.02	所有节点
	无功功率	0.02	
	电压幅值	0.002	
PMU	电压幅值	0.002	0,22,40

设置坏数据误差率为100%，注入功率变化率为60%，分类准确率如表9所示。

由表9可知，对于IEEE 33节点与丹麦DTU 7K 47节点主动配电网系统数据进行分类，相对于IEEE 14节点系统，所提基于SVM的异常检测方法对不同数据异常的分类准确率略微下降但依然保持较高分类准确率。在IEEE 33节点与丹麦DTU 7K 47节点主动配电网系统中在正常工况下提升26.08%与26.76%，在数据异常情况下准确率综合提升32.04%与29.27%。

表9 不同节点系统下不同数据类型的分类准确率对比

Table 9 Comparison of classification accuracy for different data types in different node systems (%)

节点系统	算法	正常	单相故障	两相故障	三相故障	负载突变	坏数据
IEEE33	文献[8]	68.91	57.11	57.21	58.91	54.95	65.97
	文献[9]	69.22	70.52	71.23	73.15	60.33	98.29
	文献[10]	92.57	93.82	95.71	96.42	90.21	98.34
	本文	95.13	97.72	98.83	98.99	94.21	99.10
DTU 7K 47	文献[8]	66.15	63.25	64.03	64.82	61.62	63.22
	文献[9]	65.40	67.21	67.42	69.18	70.19	97.35
	文献[10]	89.38	86.06	88.62	91.61	87.23	96.57
	本文	91.42	94.55	95.66	96.05	92.35	98.94

5 结论

本文提出了基于ELM的FASE方法与基于GWO-SVM的电力系统状态估计数据异常检测方法。基于不同场景下的大量仿真得到如下结论。

1)提出的基于GWO-SVM的数据异常检测方法具有较强的通用性与实时性,实现了对坏数据、负荷突变和单相接地故障等多种类型数据异常的检测,较现有方法有更高的检测精度与更多的检测种类,其检测精度受测试系统影响较小,且对坏数据误差率变化、非高斯量测噪声等具有较高的鲁棒性。

2)ELM-EKF与现有常见的基于机器学习的预测辅助状态估计方法相比,在电力系统状态估计上具有更高的估计精度与速度。

3)采用7种不同的优化方法对SVM中参数进行优化,验证了所提出的GWO-SVM优化算法具有较高的预测精度与较快的计算速度。

本文使用单一模型进行异常检测,而应用基于多种机器学习集成框架是目前异常检测领域的新趋势,作者未来将着重研究探讨集成学习模型在状态估计异常检测中的发展与应用。

参考文献

- [1] 俞文帅,黄蔓云,卫志农,等.基于动态分区和多估计准则的电力系统自适应鲁棒状态估计[J].电力系统自动化,2021,45(19):173-180.
YU W SH, HUANG M Y, WEI ZH N, et al. Adaptive robust state estimation for power system based on dynamic partitioning and multiple estimation criteria [J]. Automation of Electric Power Systems, 2021,45(19):173-180.
- [2] 朱茂林,刘灏,毕天姝,等.考虑输入量不良数据的发电机动态状态估计方法[J].电力系统自动化,2022,46(7):94-103.

- ZHU M L, LIU H, BI T SH, et al. Dynamic state estimation method for generators considering bad data in input [J]. *Automation of Electric Power Systems*, 2022,46(7):94-103.
- [3] 王颖,朱南阳,谢浩川,等.基于对比学习辅助训练的超短期风功率预测方法[J].*仪器仪表学报*,2023,44(3):89-97.
- WANG Y, ZHU N, XIE H, et al. Ultra-short-term wind power forecasting based on contrastive learning-assisted training [J]. *Chinese Journal of Scientific Instrument*, 2023,44(3):89-97.
- [4] 邵凯旋,何怡刚,汪磊.基于多尺度熵分析与改进SVM的变压器故障识别[J].*电子测量与仪器学报*,2022,36(6):161-168.
- SHAO K X, HE Y G, WANG L. Fault identification of transformer based on multiscale entropy analysis and improved SVM [J]. *Journal of Electronic Measurement and Instrumentation*, 2022, 36 (6) : 161-168.
- [5] 郭敏,张浩.基于改进Bi-LSTM网络下的多时变状态锂电池剩余寿命预测方法[J].*国外电子测量技术*,2023,42(10):59-68.
- GUO M, ZHANG H. Remaining useful life prediction method of lithium-ion battery in multiple time-varying states based on an improved Bi-LSTM network [J]. *Foreign Electronic Measurement Technology*, 2023, 42(10):59-68.
- [6] 张静,宋宝林,谢松霖,等.基于状态估计的高速受电弓鲁棒预测控制[J].*电工技术学报*,2021,36(5):1075-1083.
- ZHANG J, SONG B L, XIE S L, et al. Robust predictive control of high-speed pantograph based on state estimation [J]. *Transactions of China Electrotechnical Society*, 2021, 36(5):1075-1083.
- [7] 王玉彬,夏明超,李鹏,等.基于改进鲁棒自适应UKF的配电网动态状态估计方法[J].*电力系统自动化*,2020,44(1):92-100.
- WANG Y B, XIA M CH, LI P, et al. Dynamic state estimation method of distribution network based on improved robust adaptive unscented Kalman filter [J]. *Automation of Electric Power Systems*, 2020, 44(1): 92-100.
- [8] VALVERDE G, TERZIJA V. Unscented Kalman filter for power system dynamic state estimation[J]. *IET Generation, Transmission & Distribution*, 2012, 5(1):29-37.
- [9] GEETHA S J, CHAKRABARTI S, RAJAWAT K, et al. An asynchronous decentralized forecasting-aided state estimator for power systems [J]. *IEEE Transactions on Power Systems*, 2019, 34 (4) : 3059-3068.
- [10] JIN Z, ZHAO J, DING L, et al. Power system anomaly detection using innovation reduction properties of iterated extended kalman filter [J]. *International Journal of Electrical Power & Energy Systems*,2022,136: 107613.
- [11] 丁同,傅晓锦.基于改进粒子群优化算法和极限学习机的锂离子电池健康状态预测[J].*国外电子测量技术*,2023,42(5):163-173.
- DING T, FU X J. Improved particle swarm optimization and an extreme learning machine are used to predict the health state of lithium-ion batteries [J]. *Foreign Electronic Measurement Technology*, 2023, 42(5):163-173.
- [12] 官明辉,乌江,焦朝勇.基于模糊自适应扩展卡尔曼滤波器的锂电池SOC估算方法[J].*电工技术学报*,2020,35(18):3972-3978.
- GONG M H, WU J, JIA CH Y. SOC estimation method of lithium battery based on fuzzy adaptive extended kalman filter [J]. *Transactions of China Electrotechnical Society*, 2020, 35(18):3972-3978.
- [13] 王亮,王一鸣,侯威,等.基于半指数支持向量回归的电力负荷预测[J].*国外电子测量技术*,2022,41(12):164-170.
- WANG L, WANG Y M, HOU W, et al. Electricity load forecasting based on half-exp support vector regression [J]. *Foreign Electronic Measurement Technology*, 2022,41(12):164-170.
- [14] 孙林,王东山,秦伦明,等.基于灰狼算法的局部放电信号快速自动降噪算法[J].*电子测量技术*,2023,46(17):131-138.
- SUN L, WANG D SH, QIN L M, et al. Fast automatic noise reduction algorithm for PD signals based on GWO [J]. *Electronic Measurement Technology*,2023,46(17):131-138.
- [15] 张淑清,段晓宁,张立国,等. Tsne降维可视化分析及飞蛾火焰优化ELM算法在电力负荷预测中应用[J].*中国电机工程学报*,2021,41(9):3120-3130.
- ZHANG SH Q, DUAN X N, ZHANG L G, et al. Tsne dimension reduction visualization analysis and moth flame optimized ELM algorithm applied in power load forecasting [J], *Proceedings of the CSEE*, 2021,41(9):3120-3130.

作者简介

郭嘉辉,硕士研究生,主要研究方向为人工智能技术在电力系统状态估计中的应用。

E-mail:sdguojiahui@163.com

侯月婷, 硕士研究生, 主要研究方向为机器学习技术在电力系统状态估计中的应用。

E-mail:994765819@qq.com

丁磊, 教授, 博士生导师, 主要研究方向为低惯量电力系统运行控制、新能源发电并网控制和电力系统主动解

列等。

E-mail:dinglei@sdu.edu.cn

金朝阳(通信作者), 副教授, 硕士研究生导师, 主要研究方向为低惯量电力系统关键参数估计、电力系统状态估计等。

E-mail:zhaoyangjin@sdu.edu.cn