

DOI: 10.19783/j.cnki.pspc.191254

基于流形理论与 L1+L2 约束的智能电网故障定位

李泽兰^{1,2}, 曹一家¹, 彭寒梅¹, 孟凡东¹

(1. 湖南大学电气与信息工程学院, 湖南 长沙 410082; 2. 湖南电气职业技术学院, 湖南 湘潭 411100)

摘要: 现有的故障定位算法只能适应于小数据样本并且对于数据的依赖程度高。若数据中存在异常数据会带来较大的定位误差, 且故障区域与故障位置需要分别进行求解。采用流行学习方法对智能电网中全局信息进行数据提取, 充分利用了高维数据中包含的故障信息, 并构造了故障区域与故障位置相统一的模型, 通过 L1+L2 约束摆脱了对数据的依赖。算法中的参数统一采用 Adagard 方法进行学习, 自动调节学习效率, 提高参数学习的准确度和速度。最后, 在 PSCAD 中搭建 IEEE39 节点模型, 并在不同的场景下对算法的鲁棒性和定位精度进行了验证。

关键词: 流行学习; L1+L2 约束; Adagard 方法; PSCAD; 鲁棒性

Smart grid fault location based on manifold theory and L1+L2 constraints

LI Zelan^{1,2}, CAO Yijia¹, PENG Hanmei¹, MENG Fandong¹

(1. College of Electrical and Information Engineering, Hunan University, Changsha 410082, China;

2. Hunan Electrical College of Technology, Xiangtan 411100, China)

Abstract: Existing fault location algorithms only adapt to small samples and heavily depend on data. If abnormal data exists, it will cause larger positioning errors, and the fault area and fault location need to be solved separately. In this paper, the manifold learning method is used to extract global information in a smart grid. This makes full use of the fault information contained in the high-dimensional data. A unified model of fault area and fault location is constructed, which can get rid of the dependence on data through L1 + L2 constraints. The parameters in the algorithm are uniformly learned by the Adagard method, which can automatically adjust the learning efficiency and improve the accuracy and speed of parameter learning. The IEEE39 bus system built in PSCAD proves the algorithm's robustness and high precision in different scenarios.

This work is supported by National Natural Science Foundation of China (No. 51520105011).

Key words: manifold learning; L1 + L2 constraint; Adagard methods; PSCAD; robust

0 引言

准确的故障定位能及时切断故障, 防止故障蔓延而引发连锁故障; 为电网的安全、可靠运行提供有效保护。传统的故障定位主要有两大类: 一类利用故障发生时产生的特有信号进行定位^[1-3], 另一类构造相关电气量方程迭代求解故障位置^[4-6]。

由于智能电网全局可观, 电网中产生的数据可达到大数据级别。文献[7-9]指出智能电网大数据可用于配电网运行状态评估及故障辨识领域, 提高数据的冗余度。但利用智能电网的大数据进行故障定位必须解决两个关键问题: (1) 在类型繁多、价值密

度低数据下提取有效故障信息数据; (2) 利用提取出的故障参数构造高精度的智能电网故障定位模型。传统故障定位是基于小数据样本进行, 无法处理高维、异构的智能电网大数据, 且数据样本中一旦出现少量数据被污染, 就会引发较大的定位偏差, 不适用于智能电网故障定位。

针对此问题学者专家做了大量的研究, 文献[10]提出了利用多因素处理智能电网的高维数据, 但提取过程高度依赖人为设定, 文献[11]在此基础上提出了 PCA 自动提取故障特征, 但要求样本必须均匀分布, 不具备通用性。文献[12]采用不直接使用样本数据的方法, 基于离群点对智能电网状态进行分析, 但算法只能对故障区域进行定位, 无法实现故障距离的精确定位。综上所述, 目前没有适用于智

能电网的数据故障特征提取和定位的方法。

因此本文提出了基于流形理论的自适应故障特征选择方法, 对冗余高维信息进行特征融合, 组成故障定位特征向量; 并考虑噪声及信息传输失败等实场景, 构建具有鲁棒性的 L1+L2 约束的故障距离辨识模型, 并提出采用 Adagard 方法获得模型参数, 提高模型实现速度, 实现故障距离的精确定位。

1 基于流形学习提取故障特征

智能电网通过数据监控采集系统(SCADA)获取故障集合, 文献[13]指出智能电网每个节点保守产生 7 维数据, 则一次故障将产生 $7D$ (D 为节点数) 维数据, 而大量节点的故障特征信息与故障位置弱相关, 整体的价值密度低, 智能电网故障特征数据不能直接使用, 需要对故障特征进行提取后使用。

流形学习是通过分析数据点之间的邻近性来揭示数据集内在隐藏结构, 通过寻找光滑映射 $f: X \rightarrow Y$, 得到高维数据的低维表达, 实现维数约简, 其中 $X \in R^N$, $Y \in R^n$, 且 $n < N$, Y 保留数据 X 邻近结构不变。

1.1 调和平均数测地距

数据 x_i, x_j 邻近性描述准确度决定维数约简的精准性, 文献[14]指出可用欧氏空间或测地距对数据邻近性进行描述, 定义欧氏空间距离为两点之间的平方根, 测地距 $G(x_i, x_j)$ 为图论中采用 FLloyd 或 Dijkstra 算法计算所得两点之间最短的距离。

本文采用调和测地距对数据邻近性进行描述, 来调和数据的采样差异, 调和测地距公式为

$$\text{dis}(x_i, x_j) = \frac{G(x_i, x_j)}{\sqrt{H(i)H(j)}} \quad (1)$$

式中: $G(x_i, x_j)$ 为图论的测地距; $H(i)$ 和 $H(j)$ 为调和系数, 如式(2)和式(3)所示。

$$H(i) = \frac{n-1}{\sum_{k=1}^n \frac{1}{G(x_i, x_k)}} \quad k \neq i (k=1, 2, \dots, n) \quad (2)$$

$$H(j) = \frac{n-1}{\sum_{k=1}^n \frac{1}{G(x_j, x_k)}} \quad k \neq j (k=1, 2, \dots, n) \quad (3)$$

以瑞士卷为例说明三者的区别, 图 1 为瑞士卷, 黑点为数据集, 欧氏空间距离用蓝色虚线标识, 蓝色实线为测地距 $G(x_i, x_j)$ 。

在采样均匀时, 测地距能避免短接问题。如图 1 瑞士卷中两个圆点在数据集中处于不同的层级,

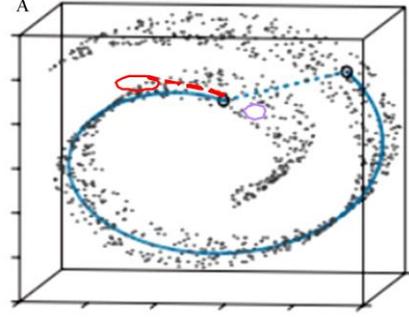


图 1 测地距分析图

Fig. 1 Analysis map of geodesic distance

两者之间相差较远, 蓝色虚线欧氏空间距离会将两点直接相连短接在一起, 而蓝色实线测地距描述时能真实反映两者相差较远的关系; 但是在采样不均匀的情况下, 文献[7-9]指出直接采用测地距描述时会产生过聚类问题。如图 1 左边黑色圆点, 远端红色圈中采样密度高、样本点多, 而近端紫色圈处, 没有采样点。直接采用测地距进行描述时会把远端红色圈内视为临近点, 与实际的数据关系不符合。本文提出的调和测地距, 会通过调和系数对测地距修正, 进行正则化处理, 获取准确数据关系描述。红色圈中 x_i 采样密度较大, 则调和系数 $H(i)$ 比较小, 则根据式(1), 最终调和测地距会变大, 准确地反映了两点之间的距离, 避免了过聚合的问题。本文采用的调和测地距与测地距和定义欧氏空间距离相比能更加精准地反映数据之间的结构关系。

1.2 基于流形学习的故障特征提取实现

假设在系统中共保存了 m 次故障, 每次故障产生 N 维数据, 构成原始的 $m \times N$ 数据集。

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mN} \end{bmatrix} \quad (4)$$

原始数据价值密度低, 提取获得低维数据 \mathbf{Y} , 基于流形学习理论, \mathbf{Y} 为保留原始数据 \mathbf{X} 全局构造的关系表达。两者结构关系矩阵满足: $\mathbf{d}(\mathbf{X}) = \mathbf{d}(\mathbf{Y})$, 其中 $\mathbf{d}(\mathbf{X})$ 是根据式(1)求取的, 结构关系矩阵为

$$\mathbf{d}(\mathbf{X}) = \begin{bmatrix} dx_{11} & dx_{12} & \dots & dx_{1N} \\ dx_{21} & dx_{22} & \dots & dx_{2N} \\ \vdots & \vdots & dx_{ij} & \vdots \\ dx_{m1} & dx_{m2} & \dots & dx_{mN} \end{bmatrix} \quad (5)$$

式中, $dx_{ij} = \text{dis}(x_i, x_j) = \sum_{k=1}^N (x_{ik} - x_{jk})^2$, $\mathbf{d}(\mathbf{Y})$ 同理。

提取实质是求解出低维空间与高维空间满足

$d(\mathbf{X}) = d(\mathbf{Y})$ 光滑映射求解, 根据多维尺度(MDS)原理, 对 \mathbf{X} , \mathbf{Y} 的数据结构关系做内积处理, 将映射函数的求解转换为特征值与特征向量的求解问题。

原始数据 \mathbf{X} 和低维数据 \mathbf{Y} 内积满足

$$-\frac{1}{2}\mathbf{H}d(\mathbf{Y})\mathbf{H} = -\frac{1}{2}\mathbf{H}d(\mathbf{X})\mathbf{H} =$$

$$-\frac{1}{2}\mathbf{H} \begin{bmatrix} (y_1 - y_1)^2 & (y_1 - y_2)^2 & \cdots & (y_1 - y_m)^2 \\ (y_2 - y_1)^2 & (y_2 - y_2)^2 & \cdots & (y_2 - y_m)^2 \\ \vdots & \vdots & (y_i - y_j)^2 & \vdots \\ (y_m - y_1)^2 & (y_m - y_2)^2 & \cdots & (y_m - y_m)^2 \end{bmatrix} \mathbf{H} =$$

$$-\frac{1}{2}\mathbf{H} \begin{bmatrix} y_1^2 & y_1^2 & \cdots & y_1^2 \\ y_2^2 & y_2^2 & \cdots & y_2^2 \\ \vdots & \vdots & y_i^2 & \vdots \\ y_m^2 & y_m^2 & \cdots & y_m^2 \end{bmatrix} - 2 \begin{bmatrix} y_1 y_1 & y_2 y_1 & \cdots & y_m y_1 \\ y_1 y_2 & y_2 y_2 & \cdots & y_m y_2 \\ \vdots & \vdots & y_i y_j & \vdots \\ y_1 y_m & y_2 y_m & \cdots & y_m y_m \end{bmatrix} +$$

$$\begin{bmatrix} y_1^2 & y_2^2 & \cdots & y_m^2 \\ y_1^2 & y_2^2 & \cdots & y_m^2 \\ \vdots & \vdots & y_i^2 & \vdots \\ y_1^2 & y_2^2 & \cdots & y_m^2 \end{bmatrix} \mathbf{H} = -\frac{1}{2}\mathbf{H}\mathbf{Z}\mathbf{H} + \mathbf{H}\mathbf{Y}\mathbf{Y}^T\mathbf{H} - \frac{1}{2}\mathbf{H}\mathbf{Z}^T\mathbf{H}$$

(6)

$\mathbf{H} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}^T$, \mathbf{I} 为 n 阶单位矩阵, $\mathbf{1}$ 为 n 维单位列向量, 为中心化矩阵, 所以 $\mathbf{H}\mathbf{Z}\mathbf{H}$ 和 $\mathbf{H}\mathbf{Z}^T\mathbf{H}$ 为 n 阶零阶矩阵。式(6)简化为

$$-\frac{1}{2}\mathbf{H}d(\mathbf{Y})\mathbf{H} = -\frac{1}{2}\mathbf{H}d(\mathbf{X})\mathbf{H} =$$

$$-\frac{1}{2}\mathbf{H} \begin{bmatrix} y_1 y_1 & y_2 y_1 & \cdots & y_m y_1 \\ y_1 y_2 & y_2 y_2 & \cdots & y_m y_2 \\ \vdots & \vdots & y_i y_j & \vdots \\ y_1 y_m & y_2 y_m & \cdots & y_m y_m \end{bmatrix} \mathbf{H} =$$

$$-\frac{1}{2}\mathbf{H}\mathbf{Y}\mathbf{Y}^T\mathbf{H}$$

内积 $\Phi = \mathbf{H}d(\mathbf{X})\mathbf{H}$ 为对称矩阵, 其特征值分解为

$$\Phi = \mathbf{A}\mathbf{L}\mathbf{A}^T \quad (8)$$

式中: \mathbf{A} 为特征向量; \mathbf{L} 为特征值, 根据式(7), \mathbf{Y} 和内积 Φ 为近似矩阵, 根据近似矩阵特性 \mathbf{Y} 可以通过 \mathbf{X} 的数据结构关系特征值分解特征向量和特征值得到。

$$\mathbf{Y} = \mathbf{A}_n \mathbf{L}_n^{1/2} \quad (9)$$

式中: \mathbf{A}_n 为矩阵 Φ 前 n 个最大特征值的特征向量; $\mathbf{L}_n^{1/2}$ 为对角阵最大特征值的平方根, 在内积处理中一直满足 $d(\mathbf{X}) = d(\mathbf{Y})$, 且 $n < N$; 根据式(9)所求 \mathbf{Y} 即为原始数据 \mathbf{X} 的低维表达。

2 故障定位统一模型

故障位置辨识模型定义为: 利用提取的故障特

征数据建立与故障距离的映射关系。

$$\mathbf{Z} = f(\mathbf{Y}) + \varepsilon \quad (10)$$

式中: \mathbf{Y} 为输入变量, 是流行学习降维后所得的故障特征数据空间构成的 $m \times n$ 维向量; \mathbf{Z} 为输出变量。目前的故障位置辨识都是给定区域的前提下对位置进行定位^[15-17], 故障区域和故障位置没有在模型中同时给出。

本文构造故障区域与故障位置相统一的故障定位模型, 定义 \mathbf{Z} 为 $m \times 2$ 维矩阵, $\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix}$, 其中

$\mathbf{Z}_1 = (z_{11}, z_{12}, \dots, z_{1n})$ 代表各个样本点的故障区域, $\mathbf{Z}_2 = (z_{21}, z_{22}, \dots, z_{2n})$ 代表各个样本点的故障距离, 则故障定位统一模型表示为

$$\begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \end{bmatrix}^T \mathbf{Y} + \varepsilon \quad (11)$$

2.1 故障区域辨识判别准则函数

对于故障区域辨识而言, 当同一区域发生故障时, 为同一类故障, 故障之间聚合度高, 用类内分布矩阵描述为

$$\mathbf{S}_w = \sum_{i=1}^c \sum_{x_k \in \text{class}_i} (u_i - y_k)(u_i - y_k)^T \quad (12)$$

对于不同区域的故障而言, 数据之间相似程度低, 类间差别用类间分布矩阵表示。

$$\mathbf{S}_b = \sum_{i=1}^c n_i (u_i - u)(u_i - u)^T \quad (13)$$

式中: c 为故障区域; n_i 为属于类别 i 的个数; u_i 为类别 i 的样本均值; u 为总样本均值。

$$u_i = \frac{1}{n_i} \sum_{y \in \text{class}_i} y_k \quad (14)$$

$$u = \frac{1}{m} \sum_{i=1}^m y_i \quad (15)$$

式中, $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{in})$ 。构造故障区域的映射函数 \mathbf{W}_1 , 故障区域辨识判别准则函数应使得类间距离足够大, 同一区域故障足够小, 判别函数定义为

$$\min_f J(\mathbf{W}_1) = \frac{\mathbf{W}_1^T \mathbf{S}_w \mathbf{W}_1}{\mathbf{W}_1^T \mathbf{S}_b \mathbf{W}_1} \quad (16)$$

2.2 故障位置辨识判别准则函数

文献[18-19]指出故障特征数据并非均匀分布, 只采用一个特征空间无法准确地捕捉数据信息。本文采用组合多尺度核方法构建再生核空间, 融合各个特征空间优势可增加故障位置辨识能力^[20-21]。

故障位置的特征空间由 K 个样条高斯核函数构成, 组合多尺度的映射模型为

$$f_{\theta}(y_i) = \sum_{j=1}^k \theta_j f(\varphi_j(y_i)), \theta_j \neq \theta_l (j \neq l) \quad (17)$$

式中, $\varphi_j(y_i)$ 为 y_i 的第 i 个核函数表达, 如式(18)所示。

$$\varphi_j(y_i) = K_i(y_i, y) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma_i^2}\right) \quad (18)$$

$f(\varphi_j(y_i))$ 为对应的核函数特征空间对故障距离的映射。

$$f(\varphi_j(y_i)) = W_1^1 y_{i1} + W_2^2 y_{i2} + \dots + W_2^n y_{in} \quad (19)$$

根据最小均方准则, 应使得预测故障位置与实际故障位置误差最小, 则辨识判别准则函数定义为

$$\min_f J(W_2) = \sum_{i=1}^n \left(d^{(i)} - \sum_{j=1}^k \theta_j f(\varphi_j(y_i)) \right)^2 \quad (20)$$

式中: $d^{(i)}$ 为发生第 i 次故障的实际故障位置; $\theta_j f(\varphi_j(y_i))$ 为根据多尺度映射后计算所得的故障位置。

3 参数学习

故障定位目标是通过参数 W_1 、 W_2 的学习, 输入故障特征向量至统一模型式(11)中, 同时输出精准故障区域与故障位置。

3.1 目标函数

参数学习的目标函数是映射后所求值与实际值之间差距最小, 则目标函数是区域和位置判别准则最小函数, 表达式为

$$\min_f J(W_1, W_2) = \sum_{i=1}^n \left(d^{(i)} - \sum_{j=1}^k \theta_j f(\varphi_j(y_i)) \right)^2 + \frac{W_1^T S_w W_1}{W_1^T S_b W_1} \quad (21)$$

3.2 L1+L2 约束

如果对式(21)直接采用最大似然法获得模型参数, 当数据存在异常值, 模型会大幅度偏离实际值。若采用 Breiman. L 提出的最佳子集方法^[22]进行模型学习, 虽然有较好的鲁棒性, 但预测精度却不令人满意。

本文对模型进行 L1+L2 约束, 保证数据存在异常值的情况下, 也可获得较高的定位精度。

定义 L1+L2 约束为

$$R = \begin{cases} r^2/2 & (|r| \leq \eta) \\ \eta|r| - \eta^2/2 & (|r| > \eta) \end{cases} \quad (22)$$

式中, $r = d^{(i)} - \sum_{j=1}^k \theta_j f(\varphi_j(y_i))$ 为预测模型与实际故障距离的残余差。 $|r| \leq \eta$, 数据为正常数据, 采用

L2 约束, 实现对正常数据精确拟合; $|r| > \eta$, 数据被视为异常数据, 采用 L1 约束, 使残余差仅为线性增长, 抑制损失增幅, 减少异常数据对模型精度的影响。为实现数据的平滑覆盖, 在 L1 约束中减去 $\eta^2/2$ 。

对于正常数据, 模型最优参数预测距离与实际距离的残余差足够小; 对于异常数据, 预测距离与实际距离的残余差足够大, 则 L1+L2 约束模型参数的目标函数为

$$\min_f J(W_1, W_2) = \frac{\sum_{i=1}^d (d^{(i)} - \sum_{j=1}^k \theta_j f(\varphi_j(y_i)))^2 / 2}{\sum_{i=1}^d \eta \left| d^{(i)} - \sum_{j=1}^k \theta_j f(\varphi_j(y_i)) \right| - \eta^2 / 2} + \frac{W_1^T S_w W_1}{W_1^T S_b W_1} \quad (23)$$

s.t. $d + k = m$

式中, d 、 m 分别为正常数据与异常数据的个数。

3.3 参数学习

本文采用 Adagard 目标参数 W_1 和 W_2 进行学习, 突破了动量梯度法和指数加权平均法求解参数的过程中学习效率恒定的缺陷, 动态地调整学习效率。Adagard 具体实施如下:

$$W_{1,2}^{i(\text{new})} = W_{1,2}^{i(\text{old})} - \frac{\eta}{\sqrt{G_{i(\text{old})} + \varepsilon}} \nabla J(W_1, W_2) \quad (24)$$

$$G_{i(\text{old})} = G_{i(\text{old-1})} + \nabla J(W_1, W_2) \quad (25)$$

式(22)各个参数的更新方法根据每个参数对于损失函数的影响程度自适应地调节学习速度, 对于损失函数影响大的给予一个快速下降学习率, 较小的给予较小的学习率, 调整的速率反比于累计的梯度平方值的平方根。

4 智能电网故障定位的步骤

本文提出的智能电网故障定位方案, 主要流程如下:

1) 数据分析与融合

采用流形理论对数据进行处理, 实现高维数据的约简, 实现数据融合。

2) 故障区域及故障位置辨识

利用约简后的故障特征构建故障区域与故障相统一的模型, 并给出判别准则函数。

3) 参数学习

根据判别准则函数构造 L1+L2 约束目标函数, 采用 Adagard 法自适应调整学习效率, 获得参数最终的模型表达。

5 仿真实验与分析

为了验证本文提出的广域故障定位策略, 在

PSCAD/EMTD 中建立如图 2 所示的 39 节点输电网络, 采用典型的 IEEE 39 节点仿真。

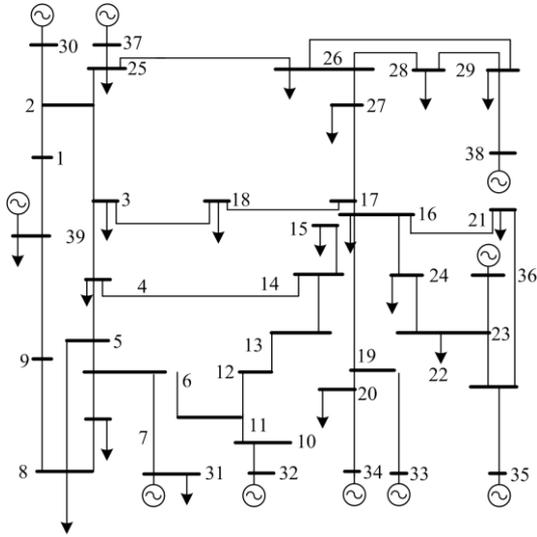


图 2 IEEE 39 节点系统

Fig. 2 IEEE 39-bus system

共 34 条线路, 10 台发电机, 12 个变压器, 根据文献[18]中的方法, 系统的行波测距装置安装位置单侧安装在线路 L12-L11, L10-L32, L19-L33, L20-L34, L23-L36, L25-L37, L2-L30, L29-L38, L19-L20 来获取暂态故障信息。

采样要求原始信号为不重叠采样信号且精度误差范围为 $\pm 1\text{LSB}$ 内。根据要求采样频率设置应在 200 kHz(采集的故障数据中行波的频率段为 10~100 kHz)获得不重叠的行波波头信息。采样速率与采样精度两者关系需满足式(26)。

$$X_m \sin \frac{2p}{S} = \frac{V_{\text{ref}}}{2^c - 1} \quad (26)$$

式中: c 为采样系统数模转换器的位数; V_{ref} 为基准电压; X_m 代表峰值; S 为采样速率。以此保证采样电压、电流等模拟数据, 经过数模转换后的数字信号所形成的故障特征集合满足采样精度要求。

5.1 算法运算量分析

整个的算法包含两大类部分运算量损耗: 一是前期模型学习; 二是后期输出运算。非流行的算法量分析前期模型学习复杂度为: $O(m^3 + m^3)$, 后期的运算输出为 $O(N)$, 总的计算量为 $O(m^3 + m^3 + N)$ 。而采用流行学习前期运算损耗由两个部分构成: 一是测地线计算, 复杂度为 $O(Km^2 \log m + Nm^2)$, 其中 K 为数据的邻近点个数, N 为特征数据的维度; 二是矩阵转换运算与特征值求解运算, 复杂度为 $O(m^3)$, 前期模型学习计算复杂度为 $O(Km^2 \log m + Nm^2 + m^3)$, 后期运算输出为 $O(n)$, 总的计算量为

$$O(Km^2 \log m + Nm^2 + m^3 + n)。$$

综上分析算法损耗量集中在前期的模型学习中, 而模型一旦学习成功, 只需将新故障直接代入已有的模型中进行后期的输出运算, 前期运算不会给后期运算造成影响, 由于后期运算为 $O(n)$ 线性运算, 所以模型可以在线诊断故障。

在前期模型学习中流行算法和非流行算法都与样本数立方成正比, 计算量损耗都比较大, 为进一步减小运算损耗, 本文在前期实际模型学习中采用基于小样本的流行学习叠加 Hebbian 规则提取数据, 其中小样本数目为 $c(c \ll m)$ 。将前期的计算复杂度降为 $O(cmN)$, 实现模型快速增量式学习。计算量对比仿真结果如图 3 所示。

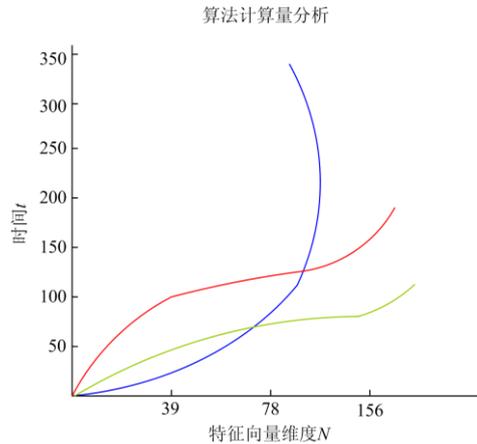


图 3 算法计算量分析

Fig. 3 Computation analysis of algorithm

图 3 中一共三条曲线, 蓝色代表采用非流行的算法运算量分析, 红色代表采用直接流行学习方法分析, 绿色曲线代表采用流行学习叠加 Hebbian 规则方法分析。三条曲线的运算时间都随着样本数目的增加而增加, 在样本维度较小时, 两种流行学习运算时间(直接和叠加 Hebbian 规则方法)都高于非流行学习, 其中流行学习方法最高; 当样本维度较大时, 两种流行学习运算时间都低于非流行学习算法, 而叠加 Hebbian 规则的流行学习, 运算量是三者中最小的。

智能电网发生故障时动态事件窗在 SCADA 系统中会将故障前断面信息和故障断面信息一起组合成故障信息, 为高维数据, 所以流行学习与非流行学习相比在应用于智能电网信息提取时所需运算量更少, 而基于小样本的流行学习叠加 Hebbian 规则在前期提取时, 能进一步缩减运行量至 $O(cmN)$, 所以运算总量是三个方法中最优的, 且适应于智能电网故障特征提取。

5.2 数据分析

在如图 2 所示的 39 节点网络中, 发生故障后, 故障数据传输到智能电网的 SCADA(数据采集与监视控制)中, 系统将收集智能电子设备发送过来的电网运行模拟量信息、监控装置开关量信息、事故警告信息三大类信息, 通过动态事件窗提取, 并根据不同类型数据的存储和管理规则组成故障特征数据集。

SCADA 中故障特征数据集类型分为两类, 一类是由开关量和电网模拟量构成的断面信息, 一类是警告信息, 对两类信息结构进行列表分析, 断面信息结构如表 1 所示, 事故警告信息结构如表 2 所示。

表 1 电网断面信息分析

Table 1 Information analysis of power grid section

名称	数据类型	长度
时间	TIMESTAMP	默认值
位置信息	VARCHAR	50
通道信息	VARCHAR	255
装置名称	VARCHAR	50
开关量状态	BOOLEAN	1
电气量	FLOAT	4

表 2 事故警告信息分析

Table 2 Analysis of accident warning information

名称	数据类型	长度
id	INT	11
时间	TIMESTAMP	默认值
告警内容	VARCHAR	50
告警端口	VARCHAR	255
保护ID	VARCHAR	50

三大类数据由 RTU、ARM、电能质量测量仪器、行波测距仪、故障滤波器等不同设备提供。由表 1 和表 2 中对比发现, 其中第二列数据类型由字节、布尔整数等多种类型构成, 且第三列的存储长度并不一致, 故障特征数据呈现多样性的特点。

算例中一共 39 节点, 根据文献[23], 一次故障将产生 273(7×39)维数据, 故障设置在 L29-38 之间, 对故障原始数据采用斯皮尔曼相关系数对行关联度分析, 并绘制仿真图如图 4 所示。

斯皮尔曼相关系数低于 0.5 时, 可以认定该维数据与故障位置弱相关。斯皮尔曼相关系数高于 0.5 时, 则认定为与故障位置强相关。如图 4 所示为 29 节点行波电流、38 节点行波电压等 17 维数据信息, 其中故障位置两端电压电流斯皮尔曼相关系数并不一致, 由于故障点的设置并不在中点位置, 离故障点较近端稍高于远端, 电流相关系数稍高于电压是

由于变压器对于电流行波的传导性更优, 其他维信息与经验判断相符, 273 维数据中除去这 17 维数据外, 其他都与故障位置弱相关。整体的价值密度低, 智能电网故障特征数据不能直接使用, 需要对故障特征进行提取后使用。

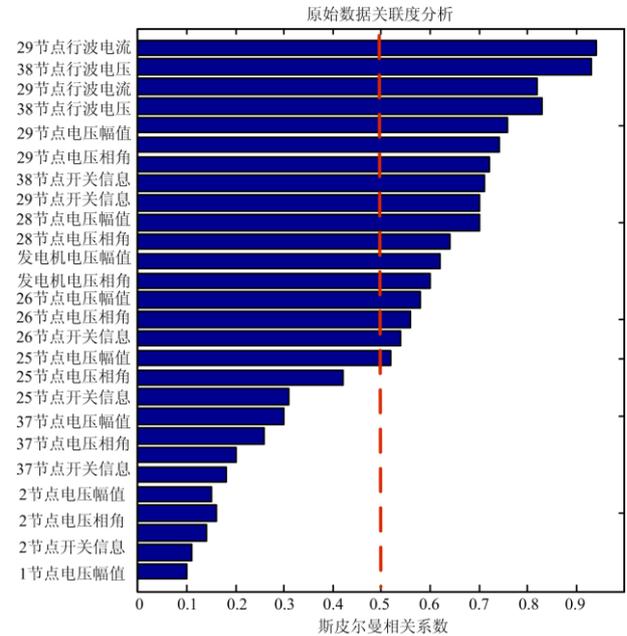


图 4 原始数据关联度分析

Fig. 4 Correlation analysis of original data

5.3 提取故障特征方法的验证

根据上文分析的原始数据多源异构, 价值密度低, 本文采用的流行学习对数据进行提取, 并对流行学习方法进行有效性验证。

1) 流行学习下特征数据提取的性能分析

对于提取后的故障数据性能采用自相关性和互相关性进行评价, 定义自相关函数为

$$I(k) = \frac{E[(x_i - \mu)(x_i - \mu)]}{\sigma^2} \quad (27)$$

互相关函数为

$$R(k) = \frac{E[(x_i - \mu)(x_{i+k} - \mu)]}{\sigma^2} \quad (28)$$

对于模型中原始故障数据采用流行学习进行特征提取后降低至 20 维。由于篇幅的限制, 提取其中 3 个特征向量进行自相关和互相关分析, 构成 3×3 的图形矩阵, 对角线上代表各数据自相关关系, 副对角线上代表数据之间的互相关系数, 如图 5 所示。

图 5 是基于流形学习方法的数据之间的相关性仿真, 第 1 行、第 1 列代表第 1 个数据特征向量的自相关函数, 第 1 行、第 2 列代表第 1 个数据特征与第 2 个数据向量之间的互相关关系, 依此类推,

观测图中自相关函数呈单调递增的趋势，而互相关函数并无单调特性，说明故障特征本身相关，故障特征之间无相关性，证明提取的数据特征不存在冗余性，提取 20 维数据有效。

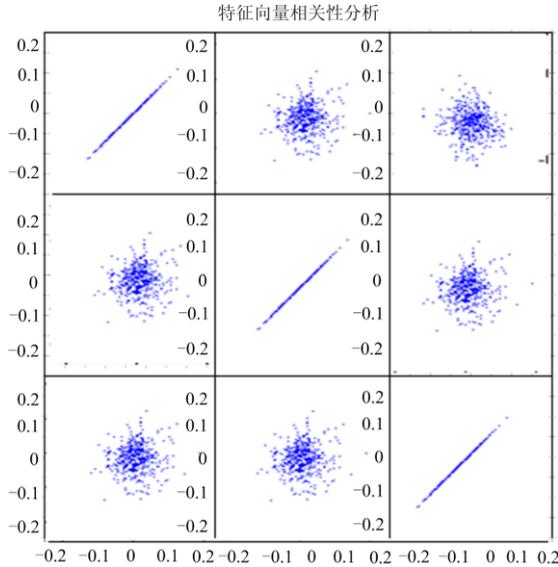


图 5 特征向量相关性分析
Fig. 5 Correlation analysis of eigenvector

文献[24-25]提出的带权重介数法的重要节点提取方法指出，39 节点中共有 8 个重要节点，则这 8 个数据产生 16 维数据(电压，电流)，文献做节点重要性判断时仅依靠线路的传输功率和电压等级，并没有考虑节点在电网整体结构中的位置，采用流形理论对数据进行全局考虑，所以提取了 20 维的数据，数据维度的提取，符合节点重要度原则下的特征选择，也验证了流形学习方法的合理性和有效性。

2) 流形学习与 PCA^[26-27]提取数据的性能分析

PCA 及其改进算法作为数据提取的重要方法，在图像分类、目标识别、视频追踪等领域广泛使用。将流行学习方法与 PCA 方法进行比较分析获得仿真图如图 6 所示。

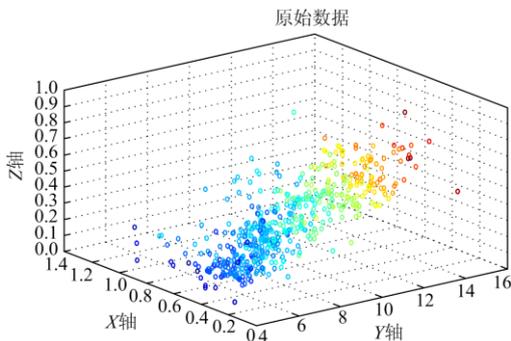


图 6 原始数据
Fig. 6 Raw data

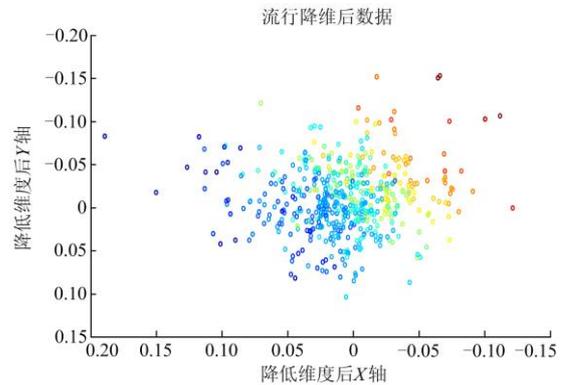


图 7 基于流形学习处理后数据
Fig. 7 Data processing based on manifold learning

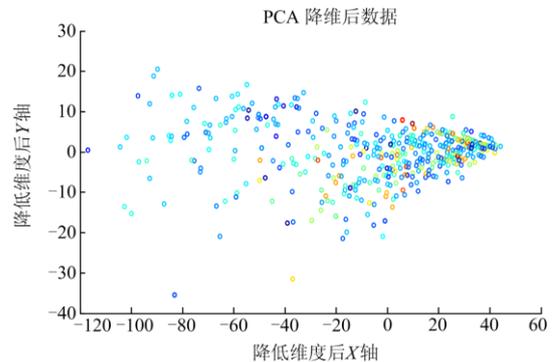


图 8 基于 PCA 处理后数据
Fig. 8 Data processing based on PCA learning

图 6 为原始数据，图 7 和图 8 分别为采用流形学习和 PCA 方法提取后数据的映射结果，PCA 及流形学习都将数据维度压缩至 20 维，原始数据结构的分布规律是前半部分为深蓝色，中段为浅蓝色，后段为浅黄色，最后段为土黄色。基于流形学习的数据提取能保持规律，而基于 PCA 的提取已经不再具备上述特点，验证了基于流形学习方式提取的数据具备整体数据结构特征。

分析故障距离与提取特征之间的相关性，流形理论下相关性为图9，PCA的相关特性为图10。

流形数据特征与故障距离之间关系，数据特征 1 与故障距离呈现出较强的相关性，数据特征向量 2 和数据特征 3 的相关性依次减弱。

PCA 各数据特征与故障距离之间的关系都较弱，PCA 第 1 特征向量相关性差于流形理论的第 3 特征向量，流形学习方法效果优于 PCA 的方法。

其原因为 PCA 方法的基本思想是预设原数据满足高斯分布前提下，去寻找数据空间中一组极大线性无关组进行数据处理。PCA 方法在简单化数据分析的同时，也扭曲了对数据内在规律性的真实反映，导致所提取的特征向量最后与故障距离无关。

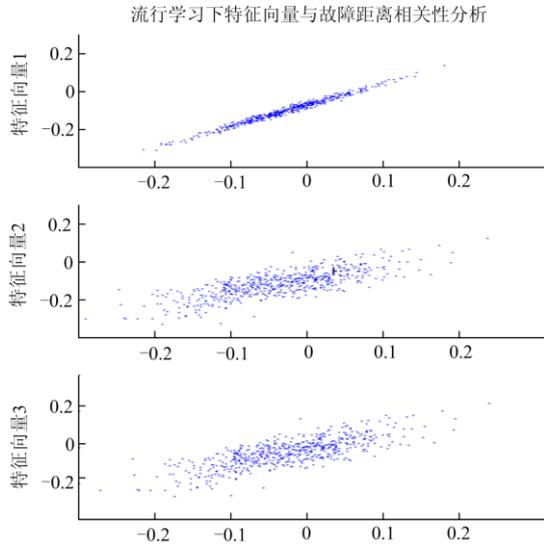


图 9 基于流行理论特征向量与故障距离的相关性分析
Fig. 9 Analysis of correlation between eigenvector and fault distance based on manifold learning

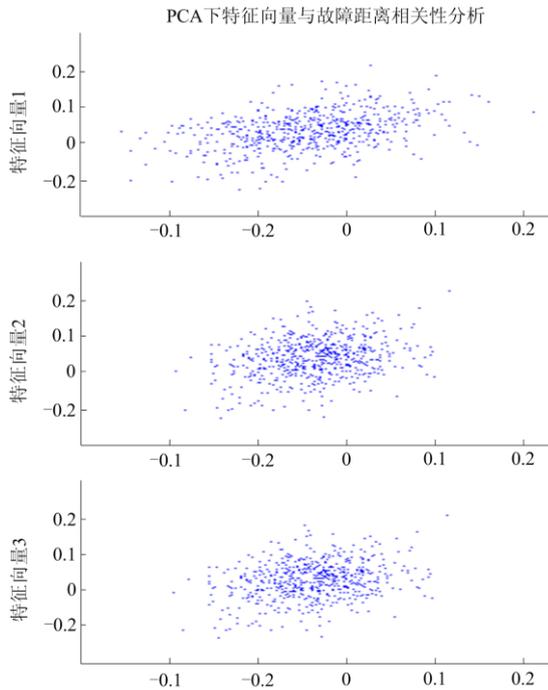


图 10 基于PCA特征向量与故障距离的相关性分析
Fig. 10 Analysis of correlation between eigenvector and fault distance based on PCA learning

而流形学习在进行数据提取时去除先验假设, 完成一般意义下对数据的挖掘, 获得数据与数据之间真实的关系。

5.4 算例系统的故障定位方法验证

算法的有效性用误差率进行描述, 误差率定义为

$$error = \left| \frac{d_{实际} - d_{模型估算}}{d_{实际}} \right| \quad (29)$$

为了方便定位, 故障距离为以行波测量仪器安装位置为参考点。验证模型精准度时, 为避免得出的误差只针对某一组特定的数据取值较优, 采用交叉验证获得平均误差, 对多种工况下进行模型精确度验证。

1) 不同故障类型下的故障定位

如表 3 所示, 在不同接地故障的情况下都获得了较好的误差率, 其中以单相故障的距离预测性能最佳, 其原因是在模型建立时, 训练数据以单相接地故障最多, 所以拟合精度也以单相接地故障最佳, 其他两种采集的样本数据较少, 数据不平衡导致了模型精准度的偏差, 可在后续中增补此类型的数据, 提高模型训练数据的精准度。鉴于此种情况, 为了获得无偏差的故障距离定位模型, 采集的类型数据样本数量之间差距不要过大。

表 3 不同故障类型误差率分析

故障类型	平均误差率	最小误差
单相接地	0.038	0.024
两相接地	0.064	0.036
相间	0.079	0.044
三相接地	0.087	0.052

2) 不同接地电阻下的故障定位

如表 4 所示, 对于不同接地电阻, 误差变化不敏感, 表明在接地电阻不同的情况下, 模型同样能获得较好的仿真精度。

3) 不同接地电阻与不同接地下的故障定位

图 11(a)为误差率与阻抗距离关系图, 图 11(b)为误差率、阻抗值与故障距离变化的三维图形, 图 11(c)为误差等高线投影; 图 11(b)的颜色由误差率决定, 不同的颜色代表不同的误差率区域, 为便于分析, 各个区域中间标注代表性的坐标点。图 11(b)三维图误差率是几个不同颜色板块的混接, 表明误差率与接地阻抗、接地电阻值之间并不是简单的线性关系。

表 4 不同接地电阻下的误差率分析

接地电阻	平均误差率	最小误差
0.1	0.036	0.024
1	0.042	0.026
10	0.058	0.032
100	0.064	0.034

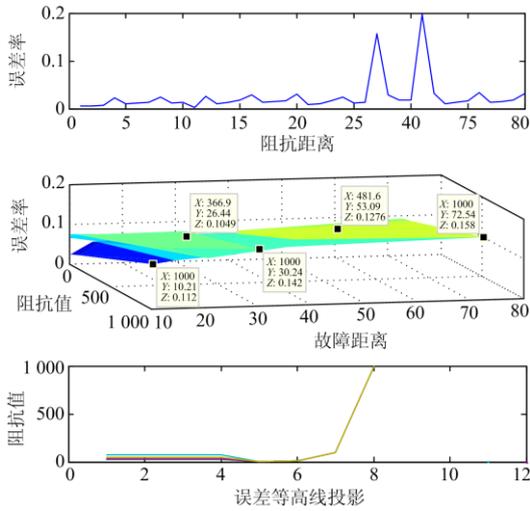


图11 误差率综合分析

Fig. 11 Synthesis analysis of error rate

由图 11(a)和图 11(b)综合分析可知, 图 11(a)表明故障误差率曲线与故障距离不成正比关系; 故障距离增加的时候, 误差率并不随之增加, 误差率并不会随着故障距离及接地阻抗值显著恶化; 分析误差变化的曲线, 误差率较高点为黄色区域, 其横坐标为 100~1000 Ω 段, 纵坐标为故障距离为 30~80 m 段, 为阻抗值较高与接地故障距离较大区域。

分析原因, 当接地故障与故障距离增加时, 行波的波头电压及电流两个参数的模分量幅值减少^[15], 使得参数信息不准确, 由于数据本身的偏差导致误差率升高。在故障距离为 72.54 m 与接地阻抗为 1 000 Ω 时, 由于行波在线路中的损耗和衰减以及接地阻抗的影响, 使得提取到的幅值信息急剧衰减至零值附近^[25-29], 则提供的信息可视为无效信息, 近似为噪声, 使得模型的参数缺损, 模型在最恶劣的情况下, 其最大误差率仍仅为 1.58%, 证明了数据的有效性。

5.5 鲁棒性分析

对加入异常信号的测试数据构建模型的残余差, 分析如图 12 所示。

残余图横轴为每个样本点, 与横轴垂直的直线代表残余的置信空间, 直线上的圆点代表每个样本点由多尺度核方法的故障定位模型回归所获得的值与实际值之间的残余差。从残余图可以看出, 除红色第四个数据外, 其余数据的置信空间都包含零点, 残余差仅在零点附近波动, 说明多尺度核方法的故障定位模型与原始数据(故障距离)之间的差值较小, 能较好地逼近训练样本。而红色样本为放入的干扰样本, 根据图 12 分析干扰样本的置信空间不过

零点, 残余差较大, 证明多尺度核方法能对存在异常数据的样本规律进行拟合, 并根据数据实际内部规律构建模型, 证明模型具有较好的抗干扰能力, 并具有较好的鲁棒性。

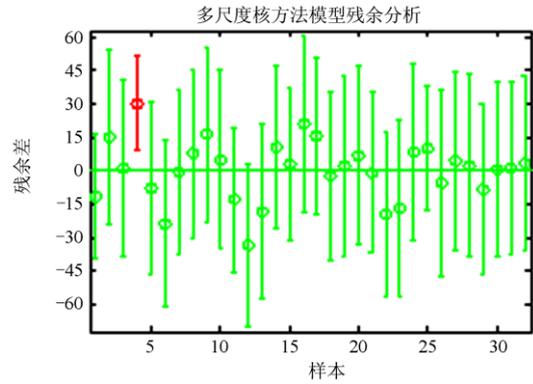


图 12 模型残余差分析

Fig. 12 Model analysis of residual error

6 结论

本文提出了一种流行学习和 L1+L2 约束相结合的故障定位与故障类型判别方法。首先对智能电网高维海量数据采用流行学习方法提取出能反映数据内蕴特征的低维数据; 其次构造了故障区域(一阶段)与故障位置(二阶段)相统一的定位模型, 一阶段故障区域采用类间分布和类内分布矩阵关系, 进行故障区域判定, 二阶段采用核函数精准地反映故障特征与故障位置的非线性关系, 通过 L1+L2 约束保证了在参数异常的情况下仍然能高精度地定位精度, 与传统算法相比摆脱了对于数据的依赖, 具有较高的鲁棒性。算法中一阶段和二阶段的参数学习中选择 Adagard 算法, 根据参数对于损失函数的影响自动地调节学习速率, 提高了参数学习的准确度和速度, 最后在 PSCAD 中搭建 IEEE39 节点系统, 通过仿真验证算法的有效性。

参考文献

[1] PARK J D, CANDELARIA J, MA Liuyan. DC ring-bus microgrid fault protection and identification of fault location[J]. IEEE Transactions on Power Delivery, 2018, 28(4): 2576-2584.

[2] 费春国, 李春信. 基于电压振幅和支持向量回归机的高压电力输电线故障定位[J]. 电力系统保护与控制, 2018, 46(13): 27-32.

FEI Chunguo, LI Chunxin. Fault location in high voltage power transmission based on voltage amplitude and support vector regression[J]. Power System Protection

- and Control, 2018, 46(13): 27-32.
- [3] COLE S, BEERTEN J, BELMANS R. Generalized dynamic VSC MTDC model for power system stability studies[J]. IEEE Transactions on Power Systems, 2010, 25(3): 1655-1662.
- [4] BRETAS A S, SALIM R H, SALIM K C O. Further improvements on impedance-based fault location for power distribution systems[J]. IET Generation, Transmission & Distribution, 2010, 9(11): 467-478.
- [5] 高淑萍, 胡振宇, 张保会, 等. 一种适用于 MMC-HVDC 直流输电线路的保护新原理[J]. 电力系统保护与控制, 2018, 46(13): 13-20.
GAO Shuping, HU Zhenyu, ZHANG Baohui. A new relay protection principle of MMC-HVDC transmission line[J]. Power system Protection and Control, 2018, 46(13): 13-20.
- [6] BORGHETTI A, BOSETTI M. On the use of continuous-wavelet transform for fault location in distribution power systems[J]. Electrical Power and Energy Systems, 2006, 28(9): 608-612.
- [7] 韩笑, 狄方春, 刘广一. 应用智能电网统一数据模型的大数据应用架构及其实践[J]. 电网技术, 2016, 40(10): 3206-3212.
HAN Xiao, DING Fangchun, LIU Guangyi. A big data application structure based on smart grid data model and its practice[J]. Power System Technology, 2016, 40(10): 3206-3212.
- [8] 王德文, 杨力平. 智能电网大数据流式处理方法与状态监测异常检测[J]. 电力系统自动化, 2016, 40(14): 122-128.
WANG Dewen, YANG Liping. Stream processing method and condition monitoring anomaly detection for big data in smart grid[J]. Automation of Electric Power Systems, 2016, 40(14): 122-128.
- [9] 胡伟, 李勇, 曹一家, 等. 基于 LOF 和 SVM 的智能配电网故障辨识方法[J]. 电力自动化设备, 2016, 36(6): 7-12.
HU Wei, LI Yong, CAO Yijia, et al. Fault identification based on LOF and SVM for smart distribution network[J]. Electric Power Automation Equipment, 2016, 36(6): 7-12.
- [10] 王彬, 王治华, 周宁慧, 等. 基于集合论估计的电网状态辨识(五)拓扑错误识别[J]. 电力系统自动化, 2016, 40(9): 9-15.
WANG Bin, WANG Zhihua, ZHOU Ninghui, et al. Power system state identification based on set theory estimation part five topology error identification[J]. Automation of Electric Power Systems, 2016, 40(9): 9-15.
- [11] 林予彰, 石立宝, 姚良忠, 等. MASOUD.一种多准则综合的电压脆弱性评估方法[J]. 电力系统自动化, 2014, 38(23): 54-59.
LIN Yuzhang, SHI Libao, YAO Liangzhong, et al. A multi-criteria integrated method for voltage vulnerability assessment[J]. Automation of Electric Power Systems, 2014, 38(23): 54-59.
- [12] 周宁慧, 王彬, 王治华, 等. 基于集合论估计的电网状态辨识(四)离群点识别[J]. 电力系统自动化, 2016, 40(8): 22-28.
ZHOU Ninghui, WANG Bin, WANG Zhihua, et al. Power system state identification based on set theory estimation part four outlier identification[J]. Automation of Electric Power Systems, 2016, 40(8): 22-28.
- [13] 张莹, 梁军. 基于分布参数模型的串联补偿双回线单线故障定位算法[J]. 电力系统自动化, 2017, 41(1): 134-139.
ZHANG Ying, LIANG Jun. Distributed parameter model based single-line fault location algorithm for series-compensated double-circuit transmission lines[J]. Automation of Electric Power Systems, 2017, 41(1): 134-139.
- [14] 朱真峰, 翟艳祥, 叶阳东, 等. 一种线性的在线 AUC 优化方法[J]. 计算机研究与发展, 2018, 55(12): 2725-2733.
ZHU Zhenfeng, ZHAI Yanxiang, YE Yangdong, et al. A linear method for online AUC maximization[J]. Computer Research and Development, 2018, 55(12): 2725-2733.
- [15] HABIB H F, YOUSSEF T, CINTUGLU M H, et al. Multi-agent-based technique for fault location, isolation, and service restoration[J]. IEEE Transactions on Industry Applications, 2018, 3(53): 1841-1851.
- [16] COLE S, BEERTEN J, BELMANS R. Generalized dynamic VSC MTDC model for power system stability studies[J]. IEEE Transactions on Power Systems, 2010, 3(25): 1655-1662.
- [17] PERSONAL E, GARCÍA A, PAREJO A. A comparison of impedance-based fault location methods for power underground distribution systems[J]. Electrical Power and Energy Systems, 2016, 25(11): 30-36.
- [18] 孙毅, 李世豪, 崔灿, 等. 基于高斯核函数改进的电力用户用电数据离群点检测方法[J]. 电网技术, 2018, 42(5): 1595-1606.
SUN Yi, LI Shihao, CUI Can, et al. Improved outlier detection method of power consumer data based on Gaussian kernel function[J]. Power System Technology, 2018, 42(5): 1595-1606.
- [19] 张骥, 侯进. 高斯核函数快速插值的头发实时仿真与

- 渲染[J]. 计算机辅助设计与图形学学报, 2017, 29(2): 320-327.
- ZHANG Pian, HOU Jin. Real-time hair simulation and rendering based on fast Gaussian kernel interpolation[J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(2): 320-327.
- [20] 王双成, 高瑞, 杜瑞杰. 小时间序列的动态朴素贝叶斯分类器学习与优化[J]. 控制与决策, 2017, 32(1): 163-166.
- WANG Shuangcheng, GAO Rui, DU Ruijie. Learning and optimization of dynamic naive Bayesian classifiers for small time series[J]. Control and Decision, 2017, 32(1): 163-166.
- [21] 王书宁, 戴建设, 胡萍. 极小极大拟合准则下的线性模型选择方法[J]. 自动化学报, 1994(1): 29-35.
- WANG Shuning, DAI Jianshe, HU Ping. Linear model selection method under minimax fitting criterion[J]. Acta Automatica Sinica, 1994(1): 29-35.
- [22] 梁睿, 徐成, 王飞, 等. 复杂输电网中基于全网行波信息的测距装置最优配置[J]. 电工技术学报, 2016, 31(21): 30-38.
- LIANG Rui, XU Cheng, WANG Fei, et al. Optimal deployment of fault location devices based on wide area travelling wave information in complex power grid[J]. Transactions of China Electrotechnical Society, 2016, 31(21): 30-38.
- [23] WANG Huifang, ZHANG Chenyu, LIN Dongyang, et al. An artificial intelligence based method for evaluating power grid node importance using network embedding and support vector regression[J]. Frontier of Information Technology & Electronic Engineering, 2019, 20(6): 816-829.
- [24] 朱国威, 王先培, 贺瑞娟, 等. 基于重要度评价矩阵的电网关键节点辨识[J]. 高电压技术, 2016, 42(10): 3347-3353.
- ZHU Guowei, WANG Xianpei, HE Ruijuan, et al. Identification of vital node in power grid based on importance evaluation matrix[J]. High Voltage Engineering, 2016, 42(10): 3347-3353.
- [25] TIAN Wende, ZHANG Guixin, ZHANG Xiang, et al. PCA weight and Johnson transformation based alarm threshold optimization in chemical processes[J]. Chinese Journal of Chemical Engineering, 2017, 26(8): 1653-1661.
- [26] CHEN Yongqing, ZHANG Lina, ZHAO Binbin. Identification of the anomaly component using BEMD combined with PCA from element concentrations in the Tengchong tin belt, SW China[J]. Geoscienc Frontiers, 2019, 10(4): 1561-1576.
- [27] 于华楠, 马聪聪, 王鹤. 基于压缩感知估计行波自然频率的输电线路故障定位方法研究[J]. 电工技术学报, 2017, 32(23): 140-148.
- YU Huanan, MA Congcong, WANG He. Transmission line fault location method based on compressed sensing estimation of traveling wave natural frequencies[J]. Transactions of China Electrotechnical Society, 2017, 32(23): 140-148.
- [28] 彭吕斌, 何剑, 谢开贵, 等. 特高压交流和直流输电系统可靠性与经济性比较[J]. 电网技术, 2017, 41(4): 1098-1107.
- PENG Lübin, HE Jian, XIE Kaigui, et al. Comparison of reliability and economy between UHVAC and UHVDC transmission systems[J]. Power System Technology, 2017, 41(4): 1098-1107.
- [29] FAN Wen, LIAO Yuan. Wide area measurements based fault detection and location method for transmission lines[J]. Protection and Control of Modern Power Systems, 2019, 4(1): 53-64. DOI: 10.1186/s41601-019-0121-9.

收稿日期: 2019-10-13; 修回日期: 2020-05-02

作者简介:

李泽兰(1983—), 女, 通信作者, 博士, 讲师, 研究方向为电力系统故障诊断, 电力系统的自愈控制; E-mail: 73696732@qq.com

曹一家(1969—), 男, 博士, 教授, 研究方向为电力系统安全与控制, 大电网智能优化调度, 分布式智能系统理论、电力信息集成与网络化控制。E-mail: yjcao@csust.edu.cn

(编辑 周金梅)