

基于二次矩阵补全的低压配电网相序识别算法

洪文慧, 李钦豪, 张勇军, 杨 银

(华南理工大学 电力学院, 广东 广州 510641)

摘要:针对低压配电网数据完整性不足的场景,提出了一种基于二次矩阵补全的低压配电网相序识别算法。首先,分析了基于电流拟合的相序识别原理与模型;其次,研究了基于奇异值门限(SVT)算法的矩阵补全算法及其对缺失数据的一次补全方法;然后,为了进一步提升矩阵补全精度,提出了基于SVT算法的二次补全算法,并在此基础上构建了基于二次矩阵补全的低压配电网相序识别方法;最后,通过实际台区用电数据对所提算法进行算例分析。算例分析结果表明,所提算法适用于电流采集数据缺失的情况,能有效地降低电流矩阵补全的误差,从而提升低压配电网相序识别的准确率。

关键词:低压配电网;拓扑识别;相序识别;奇异值门限;矩阵补全

中图分类号:TM 73

文献标志码:A

DOI:10.16081/j.epae.202204028

0 引言

低压配电网 LVDN (Low-Voltage Distribution Network) 作为直接跟用户对接的关键环节,其运行、维护的智能化水平与客户满意度密切相关^[1-3]。低压用户的相序信息支撑着 LVDN 的运行维护^[4-7],对台区线损分析、负荷精准建模、电压质量分析有着非常重要的意义。由于用户接入、改接等影响,台区用户的相序记录错误较多,而通过人工排查的工作量非常巨大,采用智能化手段进行相序识别十分必要。随着智能电表的普及应用,用户用电信息中隐含了台区内各个节点的拓扑连接信息,基于数据驱动的相序识别方法应运而生。

当前基于数据驱动的相序识别方法主要有拟合法、机器学习法、相关性分析法等。文献[8]应用电流时序数据构建数学优化模型,该模型将所有电表的相序归属关系设为 0-1 变量,目标为各相序的总电量与该相序下所有电表电量的差值最小,从而求出台区相序关系。文献[9]提出了一种基于用户电压数据的相户识别方法,通过分析电表电压与台区出线端的电压的关联性,识别台区相序关系。文献[10]提出带约束的 K-均值聚类算法对用户电压进行聚类,并通过求解聚类中心与各相序电压的差值最小化模型实现相序识别。数据驱动的相序识别算法效果很大程度上依赖于数据完整性。数据完整性反映测量点全时序或部分时段信息缺失情况,包括有效测量点比例和数据时序完整度。在实际工程中当

数据受到电表本身功能不支持、通信丢包、通信故障等影响时往往会出现数据缺失的情况。因此,如何在数据完整性不足的情况下提高相序识别准确率,成为当前 LVDN 数字化发展的一个重要课题。

近年来,矩阵补全 MC (Matrix Completion) 方法广泛应用于各个领域,如网络调查、图像处理、机器学习等。矩阵补全指的是根据矩阵中已知部分元素去估测未知部分元素的值,从而恢复出原来的矩阵,其能够迅速恢复出与原矩阵非常相近的矩阵。矩阵补全算法包括奇异值分解 SVD (Singular Value Decomposition) 算法^[11]、奇异值门限 SVT (Singular Value Threshold) 算法^[12]、鲁棒矩阵补全 RMCOSN (Robust Matrix Completion with Outliers and Sparse Noise) 算法^[13]、基本次梯度下降 BSD (Basic Subgradient Descent) 算法^[14]等。SVD 算法通过 SVD 进行矩阵补全;SVT 算法通过计算大于设定阈值的奇异值进行矩阵补全^[15];RMCOSN 算法通过范数最小化和迹范数最小化相结合进行矩阵补全;BSD 法通过建立基本次梯度下降算法的有限秩收敛性进行矩阵补全。相比较而言,SVT 算法仅对大于阈值的奇异值进行计算,收敛速度快、精度高^[12]。通常情况下,矩阵补全算法应用于低秩矩阵时才能取得较好的补全效果^[16]。而在低压台区,电流数据可能差异性很大,难以达到理想的补全效果。

针对上述问题,本文首先建立了基于电流拟合的相序识别模型,研究了基于 SVT 算法的一次矩阵补全优化方法,对整体电流数据进行一次补全,对缺失数据进行填充;然后比较了电流矩阵中各向量的相似性,利用相似度高的向量构成新矩阵,再采用 SVT 算法对该低秩矩阵进行二次补全,从而显著提高矩阵补全效果;最后通过实际台区算例验证了本文算法的准确性。

收稿日期:2021-07-06;修回日期:2022-03-02

在线出版日期:2022-04-15

基金项目:国家自然科学基金资助项目(52177085);广州市科技计划项目(202102021208)

Project supported by the National Natural Science Foundation of China(52177085) and the Science and Technology Planning Project of Guangzhou(202102021208)

1 基于电流拟合的相序识别模型

基于电流拟合的低压配电台区相序关系识别是拓扑识别中的重要方法之一^[4],基本原理是电路中的基尔霍夫电流定律。具体地,每个时刻下,各相总电流与该相的电表电流之和基本相等,如式(1)所示。

$$I_{\varphi}^t = \sum_{m \in \Omega_{\varphi}} I_m^t + \xi_{\varphi}^t \quad \varphi = A, B, C \quad (1)$$

式中: φ 为相序,本文中 φ 均归属配变低压侧; $t=1, 2, \dots, T$, T 为时刻的总数; I_{φ}^t 为相序 φ 在 t 时刻的电流值; Ω_{φ} 为相序 φ 上所有电表的集合; I_m^t 为相序 φ 上的电表 m 在 t 时刻的电流值; ξ_{φ}^t 为相序 φ 在 t 时刻的电流误差。

引入0-1变量 $x_{g,\varphi}$ 表示电表 g 与相序 φ 的归属关系,如果电表 g 归属于相序 φ ,则 $x_{g,\varphi}=1$,否则 $x_{g,\varphi}=0$ 。则式(1)可转化为:

$$I_{\varphi}^t = \sum_{g=1}^G x_{g,\varphi} I_g^t + \xi_{\varphi}^t \quad (2)$$

式中: I_g^t 为电表 g 在 t 时刻的电流值; G 为电表总数。

令 $\mathbf{X}_{\varphi} = [x_{1,\varphi}, x_{2,\varphi}, \dots, x_{G,\varphi}]^T$, $\mathbf{e} = [I_{\varphi}^t]_{T \times G}$, $\xi_{\varphi} = [\xi_{\varphi}^1, \xi_{\varphi}^2, \dots, \xi_{\varphi}^T]^T$, $\mathbf{I}_{\varphi} = [I_{\varphi}^1, I_{\varphi}^2, \dots, I_{\varphi}^T]^T$, 定义矩阵 $\mathbf{E}$ 、 $\mathbf{X}$ 、 ξ 、 $\mathbf{I}$ 如式(3)所示。

$$\mathbf{E} = \begin{bmatrix} \mathbf{e} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \mathbf{e} \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_A \\ \mathbf{X}_B \\ \mathbf{X}_C \end{bmatrix}, \quad \xi = \begin{bmatrix} \xi_A \\ \xi_B \\ \xi_C \end{bmatrix}, \quad \mathbf{I} = \begin{bmatrix} \mathbf{I}_A \\ \mathbf{I}_B \\ \mathbf{I}_C \end{bmatrix} \quad (3)$$

则相户关系识别问题转化为关于 $\mathbf{X}$ 的0-1变量求解问题,求解 $\mathbf{X}$ 的具体模型如下:

$$\begin{cases} \min f = \|\mathbf{I} - \mathbf{E}\mathbf{X}\|_2^2 \\ \text{s.t.} \begin{cases} \sum_{\varphi=A,B,C} x_{g,\varphi} \leq 1 \\ x_{g,\varphi} \in \{0, 1\} \end{cases} \quad g = 1, 2, \dots, G \end{cases} \quad (4)$$

2 基于SVT算法的矩阵补全方法

大多数低秩矩阵可根据多数已知元素恢复少量未知元素,这可通过求解一个简单的凸优化问题实现:假设原矩阵为 $\mathbf{N}$, $\mathbf{N}$ 中有 q 个元素 $\{N_{ij} | (i, j) \in \Omega\}$ 是已知的,其中 Ω 是基数为 q 的一个随机子集。大多数秩为 r 的矩阵 $\mathbf{N}$ 可通过式(5)完全恢复。

$$\begin{cases} \min \|\mathbf{H}\|_* \\ \text{s.t.} \quad H_{ij} = N_{ij} \quad (i, j) \in \Omega \end{cases} \quad (5)$$

式中: $\mathbf{H}$ 为恢复后的矩阵; $\|\mathbf{H}\|_*$ 为 $\mathbf{H}$ 的核范数,它等于 $\mathbf{H}$ 中所有奇异值之和。矩阵需满足低秩性才能保证较好的补全效果,SVT算法对矩阵秩 r 的要求是:假设有一个秩为 r 的矩阵 $\mathbf{Y}_{n_1 \times n_2}$,令 $n = \max(n_1, n_2)$,则存在常数 C ,使矩阵秩 r 和已知元素个数 q 需遵循^[16]:

$$q \geq C n^{6/5} r \ln n \quad (6)$$

式(5)所示的优化问题属于凸优化,可以将其重塑为一个半定规划,在某种意义上,这是非确定性多项式在多项式时间复杂度内遇到的问题(NP-hard)秩最小化问题中的最紧凸松弛问题:

$$\begin{cases} \min \text{rank}(\mathbf{H}) \\ \text{s.t.} \quad H_{ij} = N_{ij} \quad (i, j) \in \Omega \end{cases} \quad (7)$$

式中: $\text{rank}(\cdot)$ 表示矩阵的秩。

定义 P_{Ω} 为:

$$P_{\Omega}(\mathbf{H}) = \begin{cases} H_{ij} & (i, j) \in \Omega \\ 0 & \text{其他} \end{cases} \quad (8)$$

式(5)所示问题可以表示为:

$$\begin{cases} \min \|\mathbf{H}\|_* \\ \text{s.t.} \quad P_{\Omega}(\mathbf{H}) = P_{\Omega}(\mathbf{N}) \end{cases} \quad (9)$$

引入中间矩阵 $\mathbf{Y}$, SVT法从 $\mathbf{Y}^0 = 0$ 开始进行迭代计算,本文采用奇异值中值法选取奇异值阈值 τ ,即将 $\mathbf{N}$ 的奇异值按照数值从大到小排列后取中值,则有:

$$\tau = \text{median}(\sigma_1, \sigma_2, \dots, \sigma_r) \quad (10)$$

式中: $\sigma_1 - \sigma_r$ 为数组中的元素; $\text{median}(\cdot)$ 为中值函数,用于取数组中的中值元素^[17]。

定义整数 k_0 ,其满足:

$$\frac{\tau}{\delta \|\mathbf{P}_{\Omega}(\mathbf{N})\|_2} \in (k_0 - 1, k_0] \quad (11)$$

式中: δ 为步长。

结合SVT算法可以推导得到 $\mathbf{H}^k$ 、 $\mathbf{Y}^k$,如式(12)所示。

$$\mathbf{H}^k = 0, \mathbf{Y}^k = k \delta \mathbf{P}_{\Omega}(\mathbf{N}) \quad k = 1, 2, \dots, k_0 \quad (12)$$

推导得到 $\mathbf{H}^k$ 、 $\mathbf{Y}^k$ 后,对 $\mathbf{Y}$ 进行SVD,首先指定奇异值个数 s ,然后计算前 s 个最大奇异值和相应的奇异向量。因此,必须确定在第 k 次迭代时 $\mathbf{Y}^{k-1}$ 的奇异值个数 s_k 。设 $s_k = r_{k-1} + 1$,其中 r_{k-1} 为 $\mathbf{H}^{k-1}$ 在第 $k-1$ 次迭代中非零奇异值的个数, $r_{k-1} = \text{rank}(\mathbf{H}^{k-1})$;然后计算 $\mathbf{Y}^{k-1}$ 的前 s_k 个奇异值,如果这 s_k 个奇异值中存在小于 τ 的奇异值,则此时的 s_k 即为第 k 次迭代时 $\mathbf{Y}^{k-1}$ 的奇异值个数,否则 s_k 需要反复加上预定义的整数 l ,直至出现小于 τ 的奇异值为止。结合问题的特点,建议 l 取5。

计算得到 s_k 后,对 $\mathbf{Y}^{k-1}$ 进行SVD,得到 $\mathbf{U}_{k-1}$ 、 Σ_{k-1} 、 $\mathbf{V}_{k-1}$ 。其中, Σ_{k-1} 为将 $\mathbf{Y}^{k-1}$ 的奇异值从大到小排列,取前 s_k 个奇异值作为对角线元素的对角线矩阵; $\mathbf{U}_{k-1}$ 、 $\mathbf{V}_{k-1}$ 分别为所取奇异值对应的左奇异向量和右奇异向量。

$\mathbf{U}_k$ 和 $\mathbf{V}_k$ 分别为 $n_1 \times r$ 维和 $n_2 \times r$ 维的正交矩阵, $\mathbf{Y}^k$ 的奇异值 σ_k 为正。对于每个 $\tau \geq 0$,引入软阈值算子 D_{τ} 如下:

$$D_{\tau}(\mathbf{H}_k) = \mathbf{U}_k D_{\tau}(\Sigma_k) \mathbf{V}_k^* \quad (13)$$

$$D_{\tau}(\Sigma_k) = \text{diag}(\{(\sigma_k - \tau)_+\}) \quad (14)$$

式中: $(\sigma_k - \tau)_+ = \max(0, \sigma_k - \tau)$ 。

当满足式(15)所示条件时,算法停止。

$$\frac{\|P_{\Omega}(H_k - N)\|_F}{\|P_{\Omega}(N)\|_F} \leq \varepsilon \quad (15)$$

式中: $\|\cdot\|_F$ 表示 F 范数; ε 为固定容差,本文中取为 10^{-4} 。

至此, H^k 可以通过式(12)进行迭代求解。

3 应用二次矩阵补全的相序识别算法

通过采集台区 n_{co} 个用户在 N 个等间隔不同时刻的电流构成用户二维电流矩阵 N , 其元素 $N(i, j)$ 表示用户 j 在 i 时刻的电流数据。考虑到电流数据经常会发生数据缺失的问题,需要研究电流矩阵的补全方法,以保证相序识别效果。

一般而言,数据补全的原理是根据矩阵中行列元素的相关性进行补全,因此低秩矩阵补全的概率更高^[16]。在LVDN中,存在不少电流一直是0的空房用户^[7];此外,受相关潜在因素(如环境温度、电价、生活作息等)的影响,同一台区的用电数据一般遵循某种固定的用电规律,即电流曲线在日、周、年时间尺度上呈周期性分布。因此,由相同时间间隔的电流数据构造的理想电流矩阵 $N \in \mathbf{R}^{i \times j}$ 应满足低秩假设,即 $\text{rank}(N) \ll \min(i, j)$ 。

3.1 二次矩阵补全策略

在某些时刻,用户用电的电流数据相似度较高。由于低秩矩阵的补全效果更明显,因此,原始电流矩阵中相似度高的向量可以形成一个新矩阵,在新矩阵上进行二次补全,可以提高矩阵补全的精度。

令原始电流矩阵 $N = [I^1, I^2, \dots, I^t, \dots, I^T]$, 其中 I^t 为时刻 t 的用户电流数据列向量, $I^t = [I_1^t, I_2^t, \dots, I_g^t, \dots, I_G^t]^T$ 。在对 N 进行一次补全后,基于向量的余弦相似度表征 $I^1 - I^T$ 之间的相似度,以向量 I^1 和 I^2 为例,两者之间的相似度 ρ 为:

$$\rho = \frac{I^1 I^2}{\|I^1\| \|I^2\|} = \frac{\sum_{g=1}^G I_g^1 I_g^2}{\sqrt{\sum_{g=1}^G (I_g^1)^2} \sqrt{\sum_{g=1}^G (I_g^2)^2}} \quad (16)$$

对 N 采用SVT算法进行一次矩阵补全得到一次补全矩阵 H' , 由式(16)得到 H' 中相似度高的 M 个向量。附录A图A1为相似度较高的电流向量示例,图中不同颜色的曲线表示由不同时刻数据形成的电流向量。由图可见,所有曲线的趋势基本一致,表明在这些时刻用户的用电行为相似。因此,为了提高矩阵补全精确度,本文将相似度高的 M 个电流向量组成一个 $G \times M$ 维的相似向量矩阵 N_0 , 通过将 N_0 中的

每个数据乘以修正权重 w_g^t 对 N_0 进行修正, w_g^t 的定义为:

$$w_g^t = \begin{cases} 1 & I_g^t \text{ 未缺失} \\ w & I_g^t \text{ 缺失后补全} \end{cases} \quad (17)$$

式中: w 为小于1的常数; I_g^t 为 N_0 中数据。

3.2 算法流程

本文提出的基于二次矩阵补全的相序识别算法流程如下。

1) 采集目标台区低压母线和所有电表在多个时刻的电流时序数据,形成原始电流矩阵 N 。

2) 利用SVT算法对数据缺失的原始电流矩阵 N 进行一次补全得到一次补全矩阵 H' , 从 H' 中选取相似度高的 M 个向量,形成相似向量矩阵 N_0 ($N_0 \subset H'$)。

3) 将 N_0 中的每个数据乘以修正系数 w_g^t 对 N_0 进行修正,从而得到修正矩阵 N_0' ; 然后利用SVT算法对 N_0' 进行二次补全得到二次补全矩阵 H_0 。

4) 将 H' 中的 N_0 部分替换为 H_0 得到最终的补全矩阵 H 。

5) 按照式(18)计算得到最终的补全矩阵与原始电流矩阵的相对误差 r_{com} 。

$$r_{com} = \frac{|\|N\|_F - \|H\|_F|}{\|N\|_F} \quad (18)$$

6) 利用 N 和 H 分别生成不同的 E 和 I , 再代入模型(式(4))求出相序关系 X 。

本文通过对比原始电流矩阵与最终的补全矩阵的相序识别准确率表征矩阵补全对相序识别的效果,相序识别准确率 λ 为:

$$\lambda = \frac{n_{\text{same}}}{G} \times 100\% \quad (19)$$

式中: n_{same} 为通过缺失后 / 补全后的数据识别正确的电表个数。

4 算例分析

4.1 算例数据

本文从广州市某实际台区(台区1)的历史负荷数据中,选取了189个用户在2019年3—7月这5个月期间的数据作为算例样本数据集,以1周为周期提取1组数据,采样间隔为15 min。台区1的示意图如附录A图A2所示。

4.2 台区缺失电流数据补全结果分析

4.2.1 随机数据缺失

1) 矩阵补全效果。

为验证本文算法的补全效果,对采集到的电流数据进行随机缺失处理后再补全,算例中所有矩阵均通过仿真得到。根据式(6)对矩阵低秩性的定义, C 取0.005时,式(6)始终成立。因此,本文算例中的

矩阵都具备低秩性。

对某用户在3月某一周的电流数据进行随机缺失10%的处理,实际数据与采用本文算法进行补全后的数据曲线图如附录A图A3所示。由图可见,实际数据曲线与利用本文算法进行补全后的数据曲线基本重合,表明数据缺失率为10%时本文算法的补全误差极小,补全效果非常理想。

当电流数据缺失率在[1%,50%]范围内变化时,本文算法的平均相对误差折线图如图1所示。由图可见:当数据缺失率在[1%,45%]范围内时,本文算法的平均相对误差均在0.001以下,补全精度很高;然而当数据缺失率超过45%时,本文算法的平均相对误差迅速变大,算法补全效果下降。鉴于45%以上的数据缺失率属于比较极端的情况,可以认为本文算法对用户电流数据有良好的补全效果。

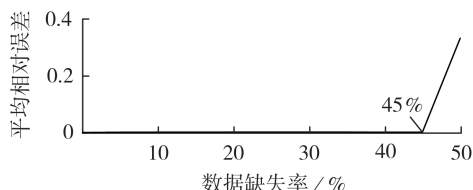


图1 数据缺失率变化范围为[1%,50%]时,本文算法的平均相对误差

Fig.1 Mean relative error of proposed algorithm when range of data missing rate is [1%,50%]

将本文算法与已有的缺失数据补全算法——SVT算法^[12]、BSD算法^[14]进行对比分析,当数据缺失率在[1%,45%]范围内变化时,3种算法的平均相对误差如图2所示。由图可见:利用本文算法进行二次补全后,平均相对误差始终为最低;与SVT算法相比,本文算法的平均相对误差降低了98%左右,表明对矩阵进行二次补全能够明显提升补全效果。

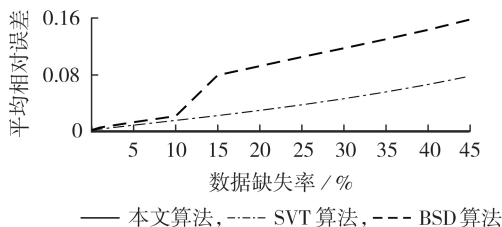


图2 随机数据缺失情况下,3种算法的平均相对误差对比
Fig.2 Comparison of mean relative error among three algorithms under random data missing condition

2)补全前后的相序识别效果。

考虑原始电流矩阵不同程度的缺失情况,利用采集到的电流数据进行仿真。在不同的数据缺失率下,采用本文算法前、后的相序平均识别准确率热力图分别如附录A图A4(a)、(b)所示。由图可见,采用本文算法进行补全前,相序平均识别准确率

为59%~88%,补全后相序平均识别准确率提升至68%~97%。采用本文算法对3月某一周的电流数据在不同的数据缺失率下进行补全,并对比补全后、未缺失、缺失未补全3种情况下的相序识别准确率,结果如附录A图A5所示。由图可见,在不同的数据缺失率下,采用本文算法进行补全后,相序识别准确率均能提高12%以上。

4.2.2 某时段数据全部缺失

电表在采集或者传输数据的过程中,可能会出现各种故障问题,难以在短时间内修复。因此,可能出现某一时段内的用户电流数据全部缺失的情况。鉴于此,对电流数据进行随机时段缺失处理,即随机连续缺失50、100、150、200、250个时刻(数据缺失率分别为10%、20%、30%、40%、50%)的电流数据,并对比采用本文算法进行补全前后的相序识别效果。

1)矩阵补全效果。

对某用户在3月某一周的电流数据进行连续缺失250个时刻的电流数据的处理,实际数据和采用本文算法进行补全后的数据曲线图如附录A图A6所示。由图可见:实际数据和补全后数据的曲线基本重合,表明在连续缺失200个时刻的电流数据的情况下,本文算法的补全效果良好;另外,在数据缺失时刻数由50增加到200的过程中,本文算法的平均相对误差均较低,在 $1.88 \times 10^{-4} \sim 1.7 \times 10^{-2}$ 范围内,但当连续缺失250个时刻的电流数据,即数据缺失率达到50%时,本文算法的精度明显下降。在实际工程中通常不会发生数据缺失率过高的情况,因此本文算法具有较高的工程实用价值。

将本文算法与SVT算法^[12]、BSD算法^[14]进行对比分析,在随机缺失50、100、150、200个时刻的电流数据的情况下,3种算法的平均相对误差如表1所示。可见,本文算法和SVT算法相比BSD算法,补全精度有明显提升,然而,二次补全相对一次补全而言优势不明显,这是因为SVT算法是根据矩阵中向量之间的相似性对矩阵进行补全,而原始电流矩阵在某一时段内的数据全部缺失时,矩阵中多列数据缺失(元素为0),补全时误差大。因此,相比4.2.1节中的数据随机缺失的情况而言,在某一时段内的数据

表1 某时段数据全部缺失的情况下,3种算法的平均相对误差对比

Table 1 Comparison of mean average error among three algorithms when all data in a certain period is missing

数据缺失 时刻数	平均相对误差		
	本文算法	SVT 算法	BSD 算法
50	0.000 188	0.000 007	0.038 270
100	0.000 111	0.000 014	0.040 478
150	0.000 344	0.000 023	0.043 154
200	0.017 023	0.004 237	0.046 293

全部缺失的情况下,利用本文算法进行二次补全时,很难找到相似向量,难以体现二次补全的优势。

2) 补全前后的相序识别效果。

当原始电流矩阵发生不同程度的时段数据缺失时,采用本文算法前、后的相序平均识别准确率热力图分别如附录A图A7(a)、(b)所示。由图A7(a)可见,随着数据缺失时段的长度由50个时刻增到200个时刻,相序平均识别准确率从91%下降至63%。由图A7(b)可见:当缺失时段长度在100个时刻以内时,本文算法的平均相序识别准确率都在90%以上;除了7月的第1周连续缺失200个时刻的电流数据的情况,在其他的连续缺失50~200个时刻的电流数据的情况下,本文算法的相序平均识别准确率都在90%左右。对3月某一周采集的电流数据的数据进行随机时段缺失处理,并对比采用本文算法补全后、未缺失、缺失未补全3种情况下的相序识别准确率,结果如附录A图A8所示。由图可见,采用本文算法进行补全后,相序识别准确率比补全前最高提升了20%左右,且补全效果十分稳定。

4.2.3 通信延时影响分析

考虑到电表数据受到通信延时的影响,在原电表电流数据上叠加由量测时延造成的高斯误差,并令电流数据的标准差为0.003^[18]。通过仿真可知,当数据缺失率为5%~45%时,本文算法的平均相对误差为 $3.21 \times 10^{-4} \sim 1.20 \times 10^{-3}$ 。在3月某一周的电流数据中叠加高斯误差,采用本文算法对叠加高斯误差后的数据在不同的数据缺失率下进行补全,对比补全后、未缺失、缺失未补全3种情况下的相序识别准确率,结果如附录A图A9所示。由图可见,利用本文算法进行补全后,相序识别准确率比缺失未补全的相序识别准确率最高增长了12%左右。因此,在通信延时的影响下,本文算法仍能有效提升数据缺失情况下的相序识别准确率。

4.3 补充算例验证

考虑到1个台区的算例分析结果具备偶然性,本文从广州市另一台区(台区2)的历史负荷数据中,选取了210个用户在2020年5—8月的数据作为算例样本数据集,以1周为周期提取1组数据,采样间隔为15 min。台区2的示意图如附录A图A10所示。采用本文算法对台区2的采样数据在不同的数据缺失率下进行补全,并对比补全后、未缺失、缺失未补全3种情况下的相序识别准确率,结果如附录A图A11所示。由图可见,采用本文算法对台区2的采用数据进行补全后,相序识别率最高提升了15%左右,效果显著。

5 结论

本文针对低压台区电表数据出现缺失引起的相

序关系识别不准确问题,提出了一种基于二次矩阵补全的LVDN相序识别方法,主要结论如下:

1) 基于台区电流数据和时刻所形成的矩阵低秩性提出二次补全算法,与一次补全算法相比,其在数据随机缺失的情况下具有更高的数据补全精度,在某时段数据全部缺失的情况下亦有良好的补全效果;

2) 提出了基于二次补全的LVDN相序识别算法,利用矩阵补全后的电流数据进行相序识别,识别准确率得到了明显提高。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 唐捷,蔡永智,周来,等. 基于数据驱动的低压配电网线户关系识别方法[J]. 电力系统自动化,2020,44(11):127-134.
TANG Jie, CAI Yongzhi, ZHOU Lai, et al. Data-driven based identification method of feeder-consumer connectivity in low-voltage distribution network[J]. Automation of Electric Power Systems, 2020, 44(11): 127-134.
- [2] 张勇军,刘斯亮,江金群,等. 低压智能配电网技术研究综述[J]. 广东电力,2019,32(1):1-12.
ZHANG Yongjun, LIU Siliang, JIANG Jinqun, et al. Research review on low-voltage intelligent distribution network technology[J]. Guangdong Electric Power, 2019, 32(1): 1-12.
- [3] 刘皓明,朱芳芳,钱程晨. 收网算法及其在低压配电网潮流优化中的应用[J]. 电力自动化设备,2014,34(4):43-48.
LIU Haoming, ZHU Fangfang, QIAN Chengchen. Shrinking net algorithm and its application in power flow optimization of low-voltage distribution system[J]. Electric Power Automation Equipment, 2014, 34(4): 43-48.
- [4] 张丽强,丛伟,董罡,等. 基于多元线性回归的单相电表相别判断方法[J]. 电力自动化设备,2020,40(5):144-156,187.
ZHANG Liqiang, CONG Wei, DONG Gang, et al. Method for single-phase electric meter phase identification based on multiple linear regression[J]. Electric Power Automation Equipment, 2020, 40(5): 144-156, 187.
- [5] 李坤,周来,张勇军,等. 基于量测数据质量的低压台区拓扑识别结果可信度评价[J]. 电力系统自动化,2021,45(17):99-107.
LI Kun, ZHOU Lai, ZHANG Yongjun, et al. Credibility evaluation of topology identification results in low-voltage distribution network based on quality of measured data[J]. Automation of Electric Power Systems, 2021, 45(17): 99-107.
- [6] FOGGO B, YU N. A comprehensive evaluation of supervised machine learning for the phase identification problem[J]. World Academy of Science, Engineering and Technology International Journal of Computer and Systems Engineering, 2018, 12(6):419-427.
- [7] ZHOU L, ZHANG Y, LIU S, et al. Consumer phase identification in low-voltage distribution network considering vacant users[J]. International Journal of Electrical Power & Energy Systems, 2020, 121:106079.
- [8] ARYA V, SEETHARAM D, KALYANARAMAN S, et al. Phase identification in smart grids[C]//2011 IEEE International Conference on Smart Grid Communications. Brussels, Belgium: IEEE, 2011:25-30.
- [9] MITRA R, KOTA R, BANDYOPADHYAY S, et al. Voltage correlations in smart meter data[C]//Proceedings of the 21th

- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, Australia: ACM, 2015: 1999-2008.
- [10] WANG W, YU N, FOGGO B, et al. Phase identification in electric power distribution systems by clustering of smart meter data[C]//2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA). Anaheim, CA, USA: IEEE, 2016: 259-265.
- [11] 李通, 舒勤. 一种基于SVD的交通流量数据补全算法[J]. 数字技术与应用, 2016(1): 129-130.
LI Tong, SHU Qin. A traffic flow data completion algorithm based on SVD[J]. Digital Technology and Application, 2016(1): 129-130.
- [12] 冯雅莉. 矩阵补全算法与图像恢复的研究[D]. 广州: 广东工业大学, 2019.
FENG Yali. The study of matrix completion algorithms and image recovery[D]. Guangzhou: Guangdong University of Technology, 2019.
- [13] HSU D, KAKADE S M, ZHANG T. Robust matrix decomposition with sparse corruptions[J]. IEEE Transactions on Information Theory, 2011, 57(11): 7221-7234.
- [14] WEN F, YING R, LIU P, et al. Matrix completion via nonconvex regularization: convergence of the proximal gradient algorithm [EB/OL]. (2019-03-02)[2021-03-08]. <https://doi.org/10.48550/arXiv.1903.00702>.
- [15] 梁温. 稳健低秩矩阵补全的交替方向乘法[D]. 北京: 北京交通大学, 2020.
LIANG Wen. Alternating direction method of multipliers for robust low rank matrix completion[D]. Beijing: Beijing Jiaotong University, 2020.
- [16] CAI J, CANDES E J, SHEN Z. A singular value thresholding algorithm for matrix completion[J]. SIAM Journal on Optimization, 2010, 20(4): 1956-1982.
- [17] 刘佳音, 于晓光, 金鹏飞, 等. 基于奇异值分解降噪方法的大型风机故障诊断研究[J]. 辽宁科技大学学报, 2016, 39(4): 284-291.
LIU Jiayin, YU Xiaoguang, JIN Pengfei, et al. Research of fault diagnosis for large fan blower based on singular value decomposition denoising method[J]. Journal of University of Science and Technology Liaoning, 2016, 39(4): 284-291.
- [18] 刘安迪, 李妍, 谢伟, 等. 基于多源数据多时间断面的配电网线路参数估计方法[J]. 电力系统自动化, 2021, 45(2): 46-54.
LIU Andi, LI Yan, XIE Wei, et al. Estimation method of line parameters in distribution network based on multi-source data and multi-time sections[J]. Automation of Electric Power Systems, 2021, 45(2): 46-54.

作者简介:



洪文慧

洪文慧(1998—),女,硕士研究生,主要研究方向为低压拓扑识别(**E-mail**: 1371267547@qq.com);

李钦豪(1989—),男,工程师,博士,通信作者,主要研究方向为无功电压控制、智能配电网运行控制等(**E-mail**: liqinhao@scut.edu.cn);

张勇军(1973—),男,教授,博士生导师,博士,主要研究方向为智能电网与能源互联网的规划、运行与控制等(**E-mail**: zhangjun@scut.edu.cn)。

(编辑 任思思)

Quadratic matrix completion based phase sequence identification algorithm for low-voltage distribution network

HONG Wenhui, LI Qinhao, ZHANG Yongjun, YANG Yin

(School of Electric Power, South China University of Technology, Guangzhou 510641, China)

Abstract: For the scenario of insufficient data integrity, a quadratic matrix completion based phase sequence identification algorithm for low-voltage distribution network is proposed. Firstly, the principle and model of phase sequence identification based on current fitting are analyzed. Secondly, the matrix completion algorithm based on SVT (Singular Value Threshold) algorithm and its one-time completion method for missing data are studied. Thirdly, a secondary complement algorithm based on SVT algorithm is proposed in order to further improve the accuracy of matrix completion, based on which, a secondary matrix complement based phase sequence identification method for low-voltage distribution network is constructed. Finally, the proposed algorithm is analyzed by an example based on the actual station power data. The case analysis results show that the proposed algorithm is suitable for the condition of missing current acquisition data, and can effectively reduce the error of the current matrix completion, thereby improving the accuracy of the phase sequence identification of low-voltage distribution network.

Key words: low-voltage distribution network; topology identification; phase sequence identification; singular value threshold; matrix completion

附录 A

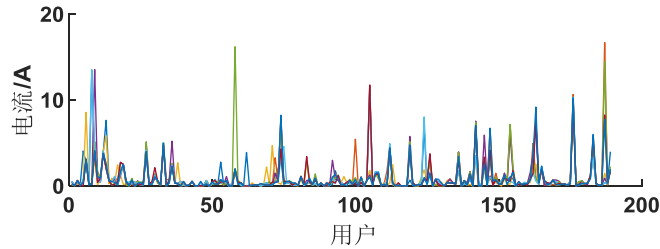


图 A1 相似度较高的电流向量绘制成的曲线

Fig.A1 Curve drawn by current vectors with high similarity

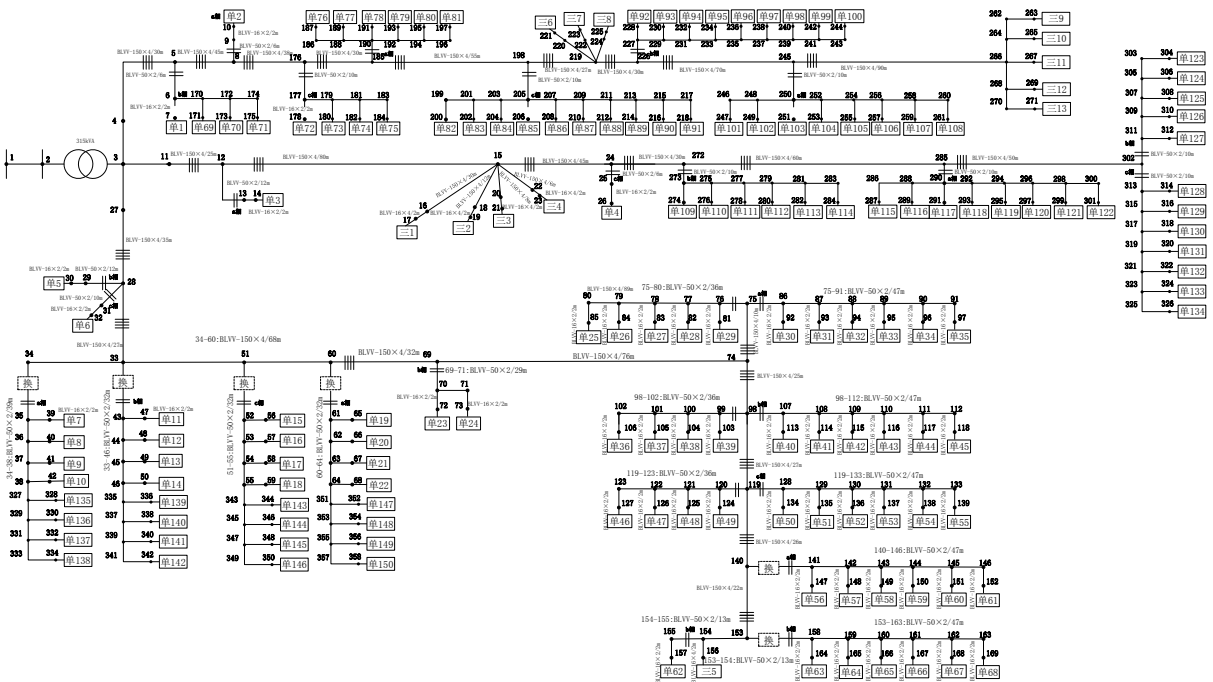


图 A2 台区 1 的拓扑连接图

Fig.A2 Topology connectivity diagram of Platform Area 1

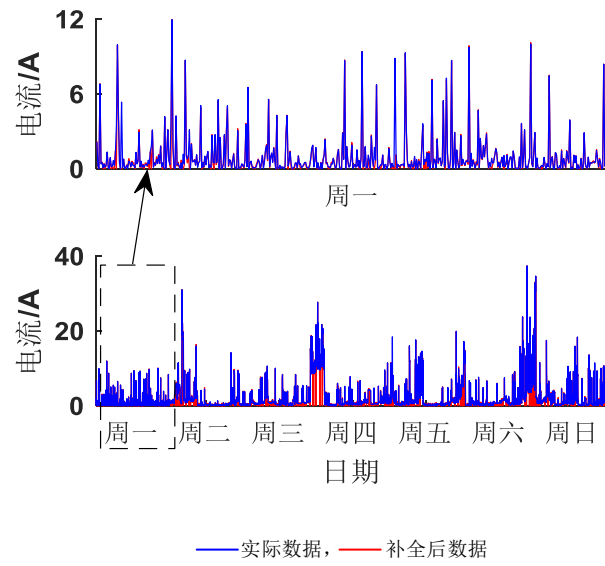


图 A3 数据缺失率为 10%时的实际数据和补全后数据曲线

Fig.A3 Actual data and completed data curves with data missing rate of 10%.

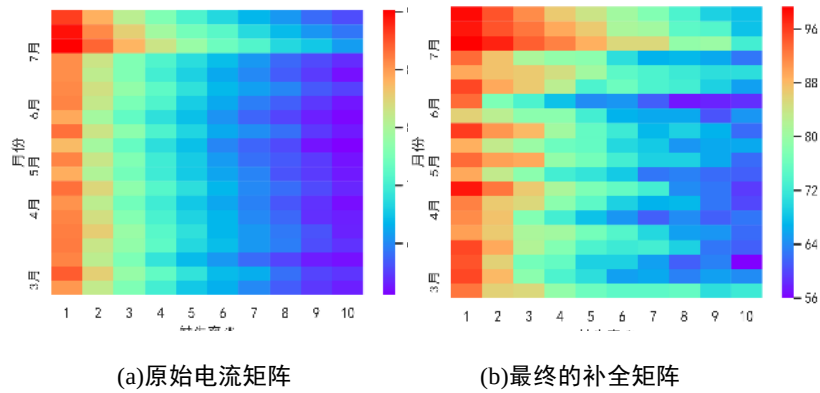


图 A4 不同数据缺失率下，采用本文算法前、后的相序识别准确率热力图

Fig.A4 Heat map of phase sequence identification accuracy rate before and after adopting proposed algorithm under different data missing rates

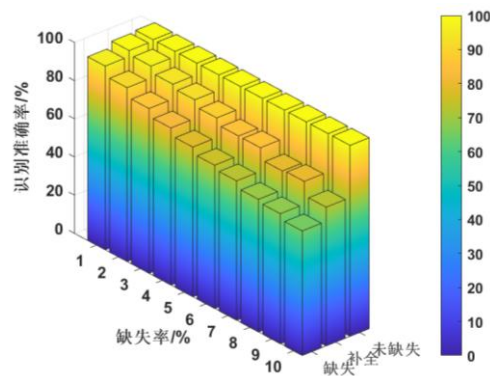


图 A5 不同数据缺失率下，3 种情况的相序识别准确率对比图

Fig.A5 Comparison of phase sequence identification accuracy rate among recognition accuracy among three conditions under different data missing rates

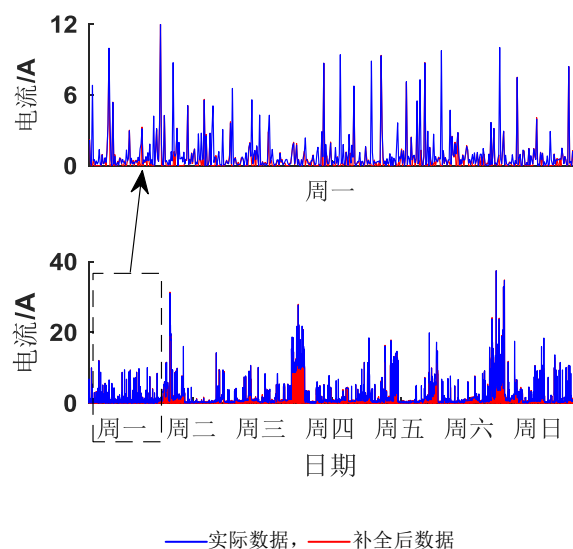
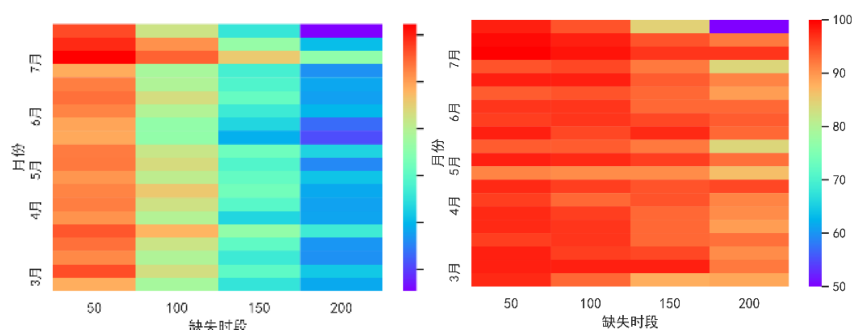


图 A6 连续缺失 200 个时刻的电流数据时，实际数据和补全后数据的曲线

Fig.A6 Actual data and completed data curves for current data missing of 200 consecutive moments



(a)原始电流矩阵

(b)最终的补全矩阵

图 A7 当某个时段的数据全部缺失时，采用本文算法前、后的相序识别准确率热力图

Fig.A7 Heat map of phase sequence identification accuracy rate before and after adopting proposed algorithm

When all data in a certain period is missing,

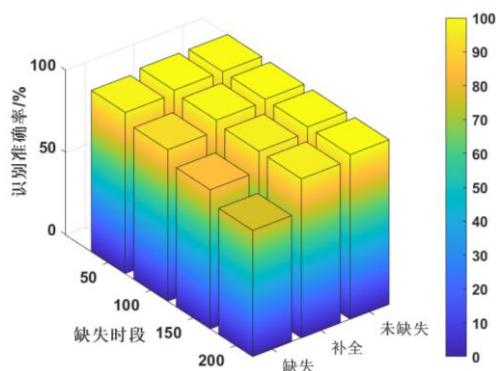


图 A8 不同的数据缺失时段长度下，3 种情况的相序识别准确率对比图

Fig.A8 Comparison of phase sequence identification accuracy rate among recognition accuracy among three conditions under different data missing period lengths

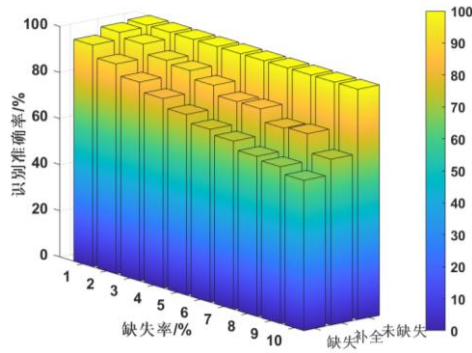


图 A9 通信延时下的相序识别准确率对比图

Fig.A9 Comparison of phase sequence identification accuracy rate under communication delay

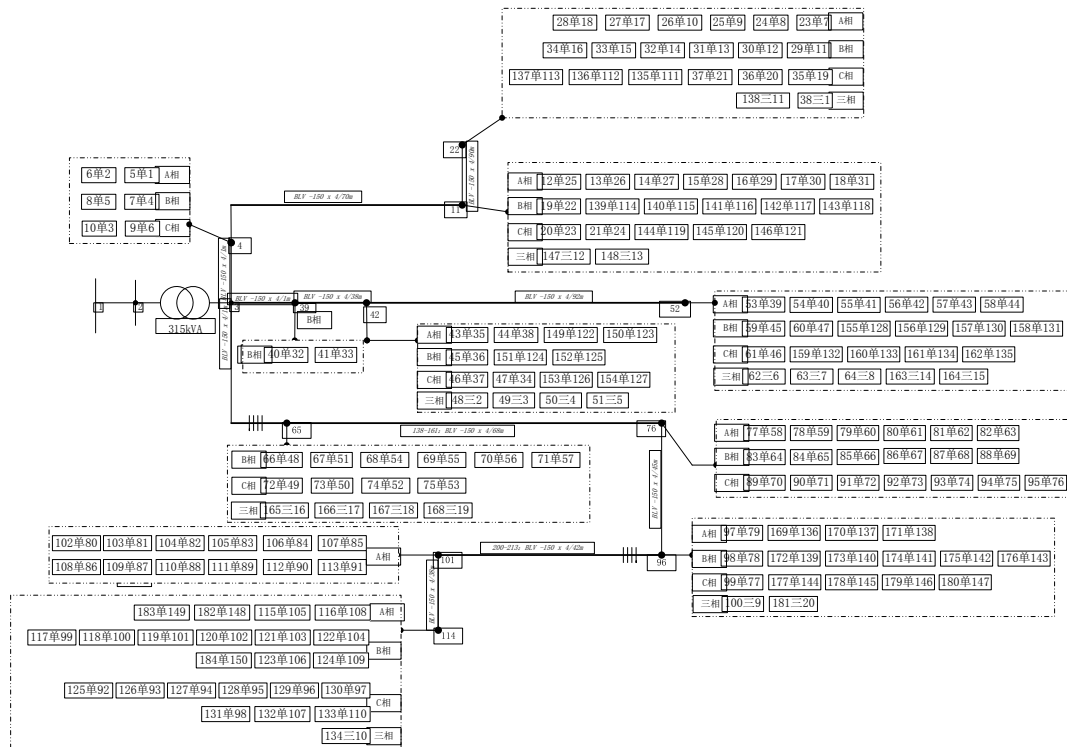


图 A10 台区 2 的拓扑连接图

Fig.A10 Topology connectivity diagram of Platform Area 2

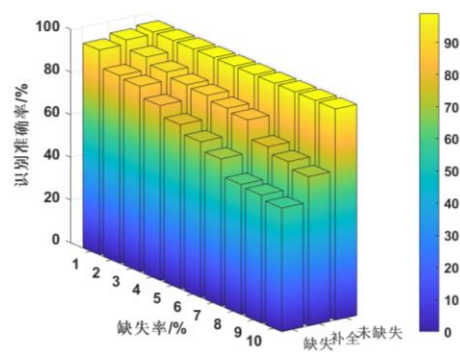


图 A11 识别准确率对比图

Fig.A11 Comparison chart of recognition accuracy