引用格式:洪佳瑶,夏向阳,雷云飞,等.基于局部线性嵌入和高斯混合模型算法的低压配电网相位识别方法[J].电力科学与技术学报,2025,40(6):109-121.

Citation: HONG Jiayao, XIA Xiangyang, LEI Yunfei, et al. Phase recognition of low-voltage distribution network based on locally linear embedding and Gaussian mixture model algorithms[J]. Journal of Electric Power Science and Technology,2025,40(6):109-121.

基于局部线性嵌入和高斯混合模型算法的低压配电网相位识别方法

洪佳瑶[1],夏向阳[1],雷云飞[2],易泽剑[1],朱汉钦[2],胡小忠[3]

(1.长沙理工大学电气与信息工程学院,湖南 长沙 410114;2.国网湖南省电力有限公司超高压变电公司,湖南 长沙 410114;
3.中南大学后勤保障部能源管理中心,湖南 长沙 410114)

摘 要:针对目前低压台区用户相位识别不精确的问题,提出一种基于用户电压数据的局部线性嵌入(locally linear embedding,LLE)降维和高斯混合模型(Gaussian mixture model,GMM)聚类算法的相位识别方法。先提取台区用户智能电表中的电压数据,并采用主成分分析法(principal component analysis,PCA)实现降噪和去冗余;再采用导数动态时间弯曲(derivative dynamic time warping,DDTW)度量用户电压之间的相关性;然后,采用LLE降维提取用户的电压数据特征,采用GMM聚类算法对用户进行相位识别;最后,在实际台区进行了仿真验证,所提方法准确度高达100%。研究结果表明,所提方法能有效解决台区用户相位识别问题。

关 键 词:低压台区;相位识别;局部线性嵌入;高斯混合模型算法

DOI:10.19781/j.issn.1673-9140.2025.06.011 **中图分类号**:TM727.2 **文章编号**:1673-9140(2025)06-0109-13

Phase recognition of low-voltage distribution network based on locally linear embedding and Gaussian mixture model algorithms

HONG Jiayao[1], XIA Xiangyang[1], LEI Yunfei[2], YI Zejian[1], ZHU Hanqin[2], HU Xiaozhong[3]

(1. School of Electrical & Information Engineering, Changsha University of Science & Technology, Changsha 410114, China;
2. Ultra-High Voltage Substation Company, State Grid Hunan Electric Power Co., Ltd., Changsha 410114, China;
3. Energy Management Center, Logistics Department, Central South University, Changsha 410114, China)

Abstract: In order to solve the problem of inaccurate user phase recognition in low-voltage station areas, a phase recognition method based on local linear embedding (LLE) dimensionality reduction and Gaussian mixture model (GMM) clustering algorithm based on user voltage data is proposed. In this method, the voltage data in the smart meter of the user in the station area is extracted firstly, and the principal component analysis (PCA) method is used to achieve noise reduction and redundancy. Then, the derivative dynamic time warping (DDTW) method is employed to measure the correlation between user voltages. The LLE dimensionality reduction is used to extract the voltage data features of the user, and the GMM clustering algorithm is used to perform phase recognition on the user. Finally, the simulation is verified in the actual station area, and the accuracy is as high as 100%, indicating that the proposed method can effectively solve the phase recognition problem of the user in the station area.

Key words: low-voltage station area; phase recognition; local linear embedding; Gaussian mixture model algorithm

收稿日期:2024-07-10;**修回日期**:2024-08-25

基金项目:国家自然科学基金(51977014)

通信作者:夏向阳(1968—),男,博士,教授,博士生导师,主要从事电网储能安全运行与优化控制等方面的研究;E-mail:307351045@qq.com

随着智能电网的飞速发展，低压配电线路的改造也在不断升级，智能电表开始广泛应用于低压配电网中。这导致台区户变关系和用户相位信息变动频繁[1-3]。低压台区的拓扑结构不仅明确了台区内各用户与变压器之间的连接关系，还是台区进行线损管理等工作的重要依据，对提高供电可靠性和推进智能电网建设具有重要意义[4-5]。

目前，常见的低压台区相位识别方法主要分为基于电力线载波通信和基于电压数据特征分析两种。但电力线载波通信的信道环境复杂，存在各种电磁干扰、跨相耦合、抗干扰能力较差等缺点[6-7]。基于电压数据特征分析的方法先采集用户端电压数据；再将提取的特征与低压台区用户电压之间的相关性结合，构建模型；最后，对其模型进行训练和测试，实现用户相位识别。该方法在理论研究和实际应用方面均获得了迅速发展[8-9]。目前，基于电压数据特征分析的低压台区相位识别方法主要包括基于数据的聚类分析和基于电压特征的相关性分析。部分学者针对电压数据的挖掘展开了一系列研究[10-12]。文献[13-15]先采用降维算法对用户原始电压数据进行特征降维，再采用聚类算法对用户进行相位识别。文献[16]先通过关系校验组建拓扑关系识别方程，再结合差值向量的取值范围与改进数据驱动模型，求得二次规划模型的矩阵。文献[17]采用多元线性回归法求解优化模型，给出了户变和相变关系识别流程算法。但这些研究均未考虑低压台区用户电压之间的相关性，当台区用户信息统计不完整时，这些方法难以保证识别的准确性。因此，一些学者对用户电压的相关性展开了研究[18-20]。文献[21]先采用改进动态时间规整(dynamic time warping，DTW)算法衡量用户之间的相似性，再基于自组织特征映射算法与K-means算法相结合的两阶段聚类方法，结合时序数据相似度实现基于形状的快速聚类。文献[22]提出了基于K-medoids聚类算法的相序识别方法。该方法先在确定最佳聚类数时引入戴维森堡丁指数作为聚类适应度；再通过遍历计算多组聚类结果的适应度选取最佳聚类数；最后，也采用欧氏距离确定每个聚类簇的相序。文献[23]通过计算相邻配电终端采集电气量的相似系数与相角差，统计同一次故障的动作或告警信息的对应关系来判断其保护区内线路以何种方式换相以及换相次数，实现了相序识别。文献[24]采用关联分析法，利用电表电压的时序变化与变压器电压曲线的相关性，实现了对用户电表的相序识别。该法步骤简单，准确率高，但其电压曲线相关性阈值须针对不同的台区重新进行设置，这导致其泛化性较差。这些研究对用户电压数据挖掘不够充分。电力系统中的电压数据易受到外界因素的影响，导致电压数据易出现缺失、突变等异常情况。此时，只考虑电压相关性是没有意义的。而这些研究对这些突变情况没有做出相应的分析。基于此，为消除电压数据缺失或异常所带来的影响，一些学者开始展开了研究[25-27]。文献[28]提出一种基于奇异值阈值算法(singular value thresholding，SVT)的二次补全算法，并在此基础上提出了一种基于二次矩阵补全的低压配电网相序识别方法。文献[29]提出了采用自适应K近邻(adaptive K nearest neighbor，AKNN)和自适应密度峰值(adaptive density peaks clustering，ADPC)聚类的异常检验算法，实现了对台区内的拓扑识别的方法。文献[30]提出了一种基于逐步回归法的(low-voltage distribution network，LVDN)相位辨识方法，该方法是一种通过定义显著性校正因子来更新逐步的回归(stepwise regression，SR)算法，提高了相位识别的多元线性回归模型的准确性。然而，这些方法计算量大且复杂。目前，亟待一种既考虑电压相关性，又基于数据分析的聚类算法，实现对用户相位的识别。

为解决这些问题，本文提出一种基于用户电压数据的局部线性嵌入(locally linear embedding，LLE)降维和高斯混合模型(Gaussian mixture model，GMM)的聚类算法的相位识别方法。该方法先对数据进行处理；再利用导数动态时间弯曲算法(derivative dynamic time warping，DDTW)，度量用户电压之间的相关性；然后，采用LLE降维提取用户电压数据的重要特征；最后，采用GMM算法对用户进行相位识别，并采用评价指标比较不同识别方法的识别效果。算例表明，本文所提方法可实现对低压台区用户的相位识别，具有较好的可行性与较高的准确性。对比以往方法，本文所提方法可根据电压相关性实现用户相位识别，在处理具有非线性边界的数据集和具有缺失值的数据集时仍然表现出了良好识别能力。

1 低压台区相位识别原理

1.1 用户相位识别原理

低压配电网对应的拓扑结构如图1所示。同一台区线路会流通潮流，故同一台区线路上的用户电

压和线路电流会存在一定的联系。文献[19]根据用户和变压器连接关系的电气原理推导出这二者的关系式，发现同一台区下的变压器与其所属用户电压曲线之间存在较高的相似度。文献[31]通过理论推导证明了若同一线路的2个同相电表的电气距离短、阻抗小、传送功率较低，则这二者的电压变化具有相似性。文献[32]根据实例论证了同一台区的用户电压变化相似性会高于不同台区的用户电压变化相似性。因此，本文根据用户之间的电压相似性来进行低压台区的相位识别。

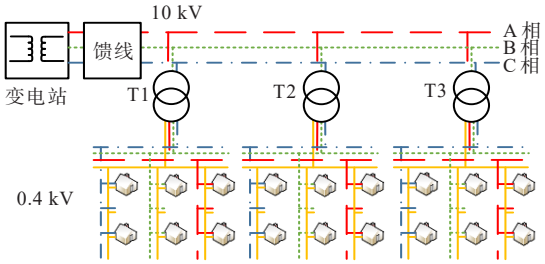


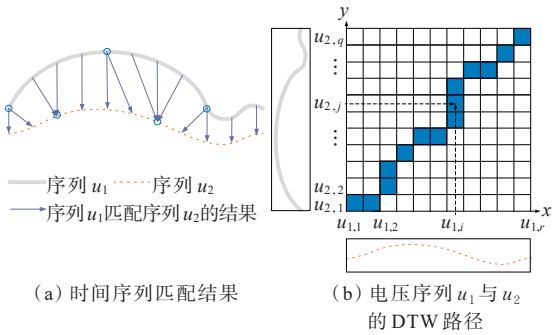
图 1　低压配电网台区示意

Figure 1　Schematic diagram of station area of low-voltage distribution network

1.2　基于 DDTW 的相似性度量

用户电压时间序列的总体组成形状大致相同，但智能电表在采集过程中可能出现的电压数据异常、缺失等情况会导致这些形状在 x 轴上对不齐。常见的相关系数方法如欧氏距离、皮尔逊相关系数等只能对等时间间隔与长度相等（对齐）的电压序列进行计算[33]，故其无法用于电压序列的相似性度量。DTW[29]算法适用于比较不同长度和变化速度的时间序列，其通过扭曲时间轴自动对齐不同长度的序列，计算电压时间序列之间的相似性。假设两个长度分别为 r 与 q 的用户的电压时间序列分别为 $u_1=\{u_{1,1},u_{1,2},\cdots,u_{1,i},\cdots,u_{1,r}\}$ 和 $u_2=\{u_{2,1},u_{2,2},\cdots,u_{2,j},\cdots,u_{2,q}\}$。时间序列 u_1 和 u_2 的匹配结果及DTW路径如图2所示。

从图2中可以看出，假定图2(a)中两条时间序列时间序列共有11个点须进行匹配，即 $r=q=11$。依据DTW算法采用一对多的匹配方式，通过扭曲 x 轴来解释 y 轴变量或者通过扭曲 y 轴来解释 x 轴变量，从图2(b)中可以更直观地看到这两条时间序列的匹配程度。通过图2(b)可以看出，采用DTW算法可以在一定程度上弥补欧氏距离点对点的匹配缺陷。如果能找到一条弯曲路径，使该路径上所有匹配点对的距离和最小，则该最小距离就是对应的



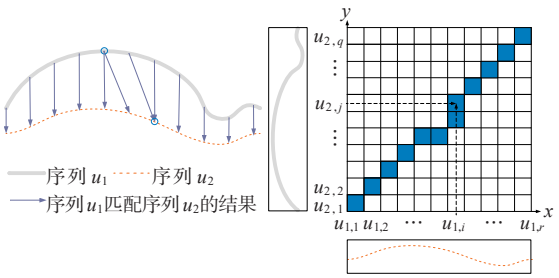
(a) 时间序列匹配结果　　(b) 电压序列 u_1 与 u_2 的DTW路径

图 2　时间序列 u_1 和 u_2 的匹配结果及DTW路径

Figure 2　Matching results and DTW paths of time series u_1 and u_2

DTW距离，这条路径也被称为最优弯曲路径。DTW算法实现了不同长度时间序列的相似性度量，匹配精度较高。但电压时间序列往往都是高维的，且电压数据中的噪声和冗余信息会影响DTW对高维时间序列的对齐效果与相似性的度量。同时，DTW在处理时间序列问题上可能会出现奇点问题，导致其对齐效果不佳[34]。

DDTW的中心思想是在DTW的基础之上引入导数，更精确地测量时间序列之间的相似性。DDTW算法是一种计算时间序列一阶导数之间相似距离的度量方法，其先对序列元素处理；再求出元素的斜率，并将其标准化；最后，提取出电压时间序列的形状特征和变化趋势。可采用DTW算法，调整两两电压时间序列之间不同元素的对应关系，找到最优弯曲路径，使该路径的两个序列距离之和最小[34]。图2所示的时间序列 u_1 和 u_2 的匹配结果及其DDTW路径如图3所示。



(a) 时间序列匹配结果　　(b) 电压序列 u_1 与 u_2 的DDTW路径

图 3　时间序列 u_1 和 u_2 的匹配结果及其DDTW路径

Figure 3　Matching results and DDTW paths of time series u_1 and u_2

从图2(a)和图3(a)的序列 u_1 与 u_2 的匹配结果中可以看出，DDTW算法的奇点情况好于DTW算法的奇点情况。图2(b)和图3(b)是DTW与DDTW

分别对序列 u_1 与 u_2 寻找的最优路径的直观表达。从中亦可很明显看出，DDTW 下的 2 个序列的路径距离要小于 DTW 的。这是因为 DDTW 算法以时间序列的估计导数插值的平方作为序列之间的距离，得到的结果更精确，能更准确地判断序列之间的相似性。

DTW 算法根据欧式距离来计算 2 个时间序列对应点之间的距离，DDTW 算法则通过 $u_{1,i}$ 和 $u_{2,j}$ 之间估计导数插值的平方来计算距离。该距离计算步骤可分为 3 步。

1）记 u_1 的导数序列为 u_1'，u_2 的导数序列为 u_2'，则有

$$\begin{cases} u'_{1,i}=\dfrac{(u_{1,i}-u_{1,i-1})+\dfrac{1}{2}(u_{1,i+1}-u_{1,i-1})}{2}, 2\leqslant i<r \\ u'_{2,j}=\dfrac{(u_{2,j}-u_{2,j-1})+\dfrac{1}{2}(u_{2,j+1}-u_{2,j-1})}{2}, 2\leqslant j<q \end{cases} \tag{1}$$

式中，$u_{1,i}$ 为序列 u_1 中第 i 个元素；$u_{2,j}$ 为序列 u_2 中第 j 个元素；$u'_{1,i}$ 为序列 u_1 中第 i 个元素的导数值；$u'_{2,j}$ 为序列 u_2 中第 j 个元素的导数值。

对时间序列求导数采用的是差分法。对于离散时间序列，一阶导数可以通过相邻两项的差值来近似。对 u_1' 和 u_2' 进行 Z-Score 标准化处理，即

$$\begin{cases} u^*_{1,i}{}'=\dfrac{u'_{1,i}-\bar{\mu}'_1}{\sigma_1} \\ u^*_{2,j}{}'=\dfrac{u'_{2,j}-\bar{\mu}'_2}{\sigma_2} \end{cases} \tag{2}$$

式中，$u^*_{1,i}{}'$ 与 $u^*_{2,j}{}'$ 分别为 $u'_{1,i}$ 与 $u'_{2,j}$ 的标准值；$\bar{\mu}'_1$ 与 $\bar{\mu}'_2$ 分别为 u_1' 与 u_2' 的期望；σ_1 与 σ_2 分别为 u_1' 与 u_2' 的标准差。

2）构建距离 $d(t_x,t_y)_{r\times q}$。其中 $d(t_x,t_y)$ 指的是 t 时刻的 $u_{1,x}$ 和 $u_{2,y}$ 的距离，即为 $u^*_{1,x}{}'$ 与 $u^*_{2,y}{}'$ 差的平方。令 u_1 与 u_2 之间的 DDTW 距离路径集合为 L，即所有路径 l 共同组成路径集合 L。l_k 为路径集 L 中横坐标为 x，纵坐标为 y 的路径，$d(l_k)$ 为路径 L 中横坐标为 x，纵坐标为 y 的 DDTW 距离。由于电压时间序列往往是高维的，则有

$$d(l_k)=d(t_x,t_y)=\left|u^*_{1,x}{}'-u^*_{2,y}{}'\right|^2 \tag{3}$$

式中，$1\leqslant x\leqslant r$，$1\leqslant y\leqslant q$。

经过预处理之后，电压序列 $u_1^*{}'$ 和 $u_2^*{}'$ 的导数动态时间弯曲路径如图 4 所示。在图 4 中，竖线正方格路径和横线正方格路径是这 2 个电压序列众多路径中的其中 2 条。

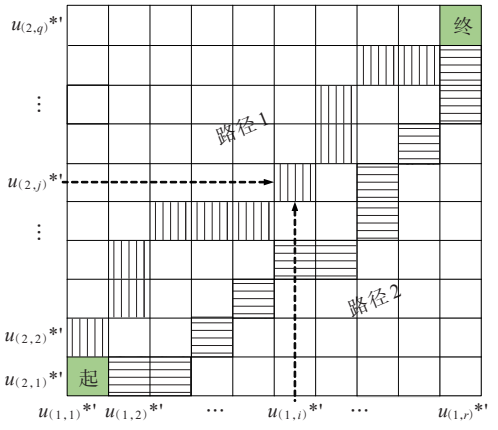


图 4 导数动态时间弯曲路径

Figure 4 DDTW path

3）定义累计距离 $D=\sum_{k=1}^{S}d(t_x,t_y)$。$\min\sum_{k=1}^{S}D(t_x,t_y)$ 表示从序列 u_1 的第 1 个元素到第 x 个元素，以及从序列 u_2 的第 1 个元素到第 y 个元素的最小累计距离。满足约束条件的路径 l 构成路径集合 S，并在路径集合 S 中找到一条最优路径使 $\sum_{k=1}^{S}d(t_x,t_y)$ 最小，其最小值即为序列 u_1 与 u_2 的 DDTW 距离，即

$$\begin{cases} \min\limits_{P}\sum\limits_{k=1}^{S}D(t_x,t_y)=d(t_x,t_y)+\min\{D(t_x,t_{y-1}), \\ \qquad\qquad D(t_{x-1},t_{y-1}),D(t_{x-1},t_y)\} \\ d_{\mathrm{DDTW}}(u_1,u_2)=\min\limits_{P}\sum\limits_{k=1}^{S}D(t_x,t_y) \end{cases} \tag{4}$$

式中，$D(t_x,t_{y-1})$ 为从序列 u_1 的第 1 个元素到第 x 个元素以及从序列 u_2 的第 1 个元素到第 $y-1$ 个元素的累计距离；$D(t_{x-1},t_{y-1})$ 为从序列 u_1 的第 1 个元素到第 $x-1$ 个元素以及从序列 u_2 的第 1 个元素到第 $y-1$ 个元素的累计距离；$D(t_{x-1},t_y)$ 为从序列 u_1 的第 1 个元素到第 $x-1$ 个元素以及从序列 u_2 的第 1 个元素到第 y 个元素的累计距离。

若时间序列 u_2 有 h 个采集时刻的电压数据缺失，$h\in(1,q)$，则删去 h 个采集时刻对应的位置，将剩下的 $q-h$ 个数据组成新的电压时间序列 $u_3=\{u_{3,1},u_{3,2},\cdots,u_{3,j},\cdots,u_{3,q-h}\}$，则时间序列 u_1 和 u_3 的导数动态时间弯曲距离为

$$d_{\mathrm{DDTW}}(u_1,u_3)=\min\sum_{k=1}^{S-h}D(t_r,t_q) \tag{5}$$

因此，通过计算电压时间序列之间的DDTW距离可以有效衡量用户之间的相似性。电压序列间的DDTW越小，用户之间的相似性越高；反之，电压序列间的DDTW越大，用户之间的相似性越低。基于此原理，可实现低压台区用户相位识别。

2　基于LLE特征降维与GMM聚类算法的低压台区相位识别

2.1　数据预处理

低压台区用户的电压数据往往是高维的，高维数据往往含有噪声和冗余信息。因此，在进行相位识别之前，应先对采集的数据进行预处理。主成分分析法[35]（principal component analysis，PCA）也称主分量分析，被广泛用于数据的降维处理。先通过PCA实现数据降维，再对其进行重构。这样的操作既去除了原始电压数据噪声与冗余，又保留了其本质特征。

对数据向量 $\boldsymbol{x}=\left(x^{(1)},x^{(2)},\cdots,x^{(m)}\right)^{\mathrm{T}}$ 进行中心化后，计算其协方差矩阵 $\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}$，即

$$\begin{cases}\tilde{x}^{(i)}=x^{(i)}-\dfrac{1}{m}\displaystyle\sum_{j=1}^{m}x^{(j)}\\ V(\boldsymbol{x})=E\left[\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^{\mathrm{T}}\right]\end{cases}\tag{6}$$

式中，$i=1,2,\cdots,m$，$j=1,2,3,\cdots,m$。$\tilde{x}^{(i)}$ 为中心化后的数据；$V(\boldsymbol{x})$ 为数据集 $\boldsymbol{x}$ 的协方差矩阵；$\boldsymbol{\mu}=E(\boldsymbol{x})$，为 $\boldsymbol{x}$ 的期望向量；

计算协方差矩阵的特征值 $z^{(i)}$ 和对应的特征向量 $\boldsymbol{\alpha}$，将特征值 $z^{(1)},z^{(2)},\cdots,z^{(n)}$ 进行排序，并获取对应的特征向量 $\boldsymbol{\alpha}^{(1)},\boldsymbol{\alpha}^{(2)},\cdots,\boldsymbol{\alpha}^{(n)}$。选择主成分的数量 n'。

最后，使用选定的主成分将数据投影到低维空间进行重构数据，这样重构后的数据即为降噪后的数据。

2.2　基于LLE的电压特征提取方法

PCA通过找出数据方差最大的方向，并将数据投影到这些方向上，但其只考虑了数据的方差，没有考虑数据的相关性。t分布-随机邻近嵌入（t-distributed stochastic neighbor embedding，t-SNE）通过概率分布转换相似性来保留局部结构，其缺点是不适用于大规模数据且对参数的选择较为敏感。奇异值分解（singular value decomposition，SVD）通过分解原始数据矩阵来揭示数据的本质结构，其缺点是仅适用于线性数据。因此，亟需一种可用于大规模数据且考虑数据相关性的非线性降维算法，对用户电压数据进行特征降维。

LLE[36]是一种非线性降维算法，其可在保留数据拓扑关系的同时，将高维数据映射到低维空间，不仅保持数据的准确性，还能有效地降低维数。对某电压时间序列 $U=\{u_1,u_2,\cdots,u_n\}$ 的特征进行降维，维数由 n 降到 d。

计算含 n 个点的时间序列 U 中任意2个点的距离，按照DDTW距离大小进行升序排序，将这 $n-1$ 个点与 u_i 距离最小的 k 个点选取出来并构成该样本点 u_i 的 k 近邻集合 $N_i=\{u_{i1},u_{i2},\cdots,u_{ik}\}$。

计算权重矩阵 $\boldsymbol{W}=\left[\boldsymbol{W}_1\ \boldsymbol{W}_2\ \cdots\ \boldsymbol{W}_i\right]^{\mathrm{T}}$ 构建局部重建模型。其中，$\boldsymbol{W}_i=\left[W_{i1}\ W_{i2}\ \cdots\ W_{ik}\right]^{\mathrm{T}}$。高维空间的样本点 u_i 可由近邻的样本点线性重构得到，即

$$u_i=W_{ij}u_j+W_{i2}u_j+\cdots+W_{ik}u_k\tag{7}$$

式中，W_{ij} 为 u_j 的重构系数。

定义高维空间下的损失函数 $\varepsilon(\boldsymbol{W})$，即

$$\varepsilon(\boldsymbol{W})=\sum_{i=1}^{k}\left(u_i-\sum_{N_i}W_{ij}u_j\right)^2=\sum_{i=1}^{k}\boldsymbol{W}_i^{\mathrm{T}}\left(u_i-u_j\right)^2\boldsymbol{W}_i=\sum_{k=1}^{k}\boldsymbol{W}_i^{\mathrm{T}}Z_i\boldsymbol{W}_i\tag{8}$$

$$Z_i=\left(u_i-u_j\right)^2\tag{9}$$

对 $\varepsilon(\boldsymbol{W})$ 做拉格朗日对偶求极值，即

$$L(\boldsymbol{W})=\sum_{i=1}^{k}\boldsymbol{W}_i^{\mathrm{T}}Z_i\boldsymbol{W}_i+\lambda\left(\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{I}_k-1\right)\tag{10}$$

式中，$\boldsymbol{I}_k=(1\ 1\ \cdots\ 1)^{\mathrm{T}}$，$\lambda$ 为拉格朗日乘子。

令 $\tau=\dfrac{-\lambda}{2}$，根据约束条件 $\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{I}_k=1$，对式(10)进行归一化，即可求出 τ。将其带入式(10)即可得到高维权重向量 $\boldsymbol{W}_i$，即

$$\boldsymbol{W}_i=\frac{Z_i^{-1}\boldsymbol{I}_k}{\boldsymbol{I}_k^{\mathrm{T}}Z_i^{-1}\boldsymbol{I}_k}\tag{11}$$

将求出的 n 维权重矩阵 $\boldsymbol{W}$ 映射到 d 维，求低维矩阵 $\boldsymbol{Y}=\left[y_1\ y_2\ \cdots\ y_d\right]^{\mathrm{T}}$。构建低维空间映射模型 $\min\displaystyle\sum_{i=1}^{k}\left(y_i-\sum_{j=1}^{k}W_{ij}y_i\right)^2$。$\displaystyle\sum_{j=1}^{k}W_{ij}y_i=\boldsymbol{Y}\boldsymbol{W}_i$。经过整理可以得到低维空间下的损失函数 $\varepsilon(\boldsymbol{Y})$，即

$$\varepsilon(\boldsymbol{Y})=\min\left(y_i-\sum_{j=1}^{k}W_{ij}y_i\right)^2=\min\operatorname{tr}\left(\boldsymbol{Y}(\boldsymbol{I}-\boldsymbol{W})(\boldsymbol{I}-\boldsymbol{W})^{\mathrm{T}}\boldsymbol{Y}^{\mathrm{T}}\right)\tag{12}$$

约束条件为

$$\sum_{i=1}^{n} y_i = 1 \tag{13}$$

式中，令 $\boldsymbol{M}=(\boldsymbol{I}-\boldsymbol{W})(\boldsymbol{I}-\boldsymbol{W})^{\mathrm{T}}$。

求式(12)的拉格朗日对偶式，并对其求极值，即

$$L(\boldsymbol{Y})=\mathrm{tr}\left(\boldsymbol{Y}\boldsymbol{M}\boldsymbol{Y}^{\mathrm{T}}+\lambda\left(\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{T}}-m\boldsymbol{I}\right)\right) \tag{14}$$

对式(14)求偏导数，即可得到 $\boldsymbol{Y}$，即

$$\boldsymbol{M}\boldsymbol{Y}^{\mathrm{T}}=-\lambda\boldsymbol{Y}^{\mathrm{T}} \tag{15}$$

将 $\boldsymbol{M}$ 的特征值按由小到大的升序排列，并取前 d 个非零特征值对应的特征向量组合，即可构成目标低维矩阵 $\boldsymbol{Y}$。但由于 $(\boldsymbol{I}-\boldsymbol{W})(\boldsymbol{I}-\boldsymbol{W})^{\mathrm{T}}\boldsymbol{E}=0$，即 $\boldsymbol{M}$ 的最小特征值为 0，故应提取 $\boldsymbol{M}$ 从小到大的第 2 个到第 $d+1$ 个特征值对应的特征向量，构成 $\boldsymbol{M}=\left(\boldsymbol{\beta}_2\ \ \boldsymbol{\beta}_3\ \ \cdots\ \ \boldsymbol{\beta}_{d+1}\right)^{\mathrm{T}}$，得到最终的低维矩阵。

2.3 基于GMM的相位识别方法

根据文2.1、2.2所述方法，得到去噪后的低维数据矩阵 $\boldsymbol{Y}$ 后，进一步采用聚类算法对低压台区内用户进行相位识别。

高斯混合聚类(Gaussian mixture model，GMM)算法[37]是单一高斯概率密度函数的扩展，其能够平滑地近似任意形状的密度分布。GMM算法的思想是假设用户数据点是由一个或多个高斯分布生成的，通过最大似然估计来估计每个簇的高斯分布的参数。假设某原始数据集为 $X=\{x_1,x_2,\cdots,x_n\}$，现对数据集 X 进行GMM聚类分析。

确定聚类簇数 $k_{n'}$，初始化模型参数。本文设定聚类簇数 $k_{n'}=m$，则数据集 X 中某一点 x_i 的高斯混合分布的概率函数与约束条件分别为

$$p(x_i)=\sum_{j=1}^{m}\alpha_j f(x) \tag{16}$$

$$\mathrm{s.t}\sum_{j=1}^{m}\alpha_j=1 \tag{17}$$

式中，$p(x_i)$ 为 x_i 的概率分布密度函数，在GMM模型中，这些数据的概率分布函数可以通过加权函数表示；模型参数 α_j 为第 j 簇的高斯混合系数，也称权值因子，代表选择第 j 个混合成份的概率；$f(x)$ 为正态分态的概率密度函数。

$$f(x_i)=\frac{1}{\sqrt{(2\pi)^m\left|\delta_i^2\right|}}\left[\frac{-(x_i-u_i)^2}{2\delta_i^2}\right] \tag{18}$$

式中，u_i 与 δ_i^2 分别为 x_i 的期望与方差。

在聚类开始时，初始化参数模型，再使用期望最大化(expectation maximization，EM)算法来求解GMM的模型参数。

对于给定的模型参数 θ，计算每个数据点 x_i 对每个簇类的隶属度 $E(\alpha_{jm})$ 和属于每个高斯分量 j 的后验概率 $p(x_i)$，即

$$E(\alpha_{jm})=\frac{\alpha_j f(x_i)}{\sum_{j=1}^{m}\alpha_j f(x_i)} \tag{19}$$

$$p(x_i)=\alpha_j f(x_i) \tag{20}$$

隶属度越大，数据样本由该分模型生成的概率越大。故有

$$\alpha_j^{\mathrm{new}}=\frac{1}{n}\sum_{i=1}^{n}p(x_i) \tag{21}$$

$$\mu_j^{\mathrm{new}}=\frac{\sum_{i=1}^{n}p(x_i)x_i}{\sum_{i=1}^{n}p(\boldsymbol{x}_i)} \tag{22}$$

$$\delta_j^{\mathrm{new}}=\frac{\sum_{i=1}^{n}p(x_i)\left(x_i-\mu_j^{\mathrm{new}}\right)^2}{\sum_{i=1}^{n}p(x_i)} \tag{23}$$

式中，α_j^{new} 为更新后的第 j 簇的高斯混合系数；μ_j^{new} 为更新后的均值；δ_j^{new} 为更新后的方差。

根据式(21)～(23)计算出的后验概率，更新每个高斯分布的参数。

重复EM算法，迭代更新模型参数直到模型收敛，最终得到数据样本点对各个类别的隶属度。根据高斯混合分布把样本集划分为 $k_{n'}$ 个簇 $C=\{C_1,C_2,\cdots,C_{n'}\}$，数据集 $X=\{x_1,x_2,\cdots,x_n\}$ 所属的簇为

$$\gamma_i=\arg\max_j\ p(x_i) \tag{24}$$

2.4 相位识别算法的评价指标

低压台区相位识别流程如图5所示。

为衡量相位识别的准确性，使用调整兰德系数(adjusted rand index，ARI)I_{ARI} 作为评价指标[38]。通过比较样本点对位于同一类簇和不同类簇的数目来度量2个聚类结果之间的相似度。I_{ARI} 越接近1，聚类结果与真实结果越一致。I_{ARI} 的计算式为

$$I_{\mathrm{ARI}}=\frac{2(ad-bc)}{(a+b)(b+d)+(a+c)(c+d)} \tag{25}$$

式中，a 为试验情况与真实情况都属于一个类簇的数目；b 为真实情况属于同一簇但试验情况不属于同一簇的数目；c 为真实情况不属于同一簇但试验

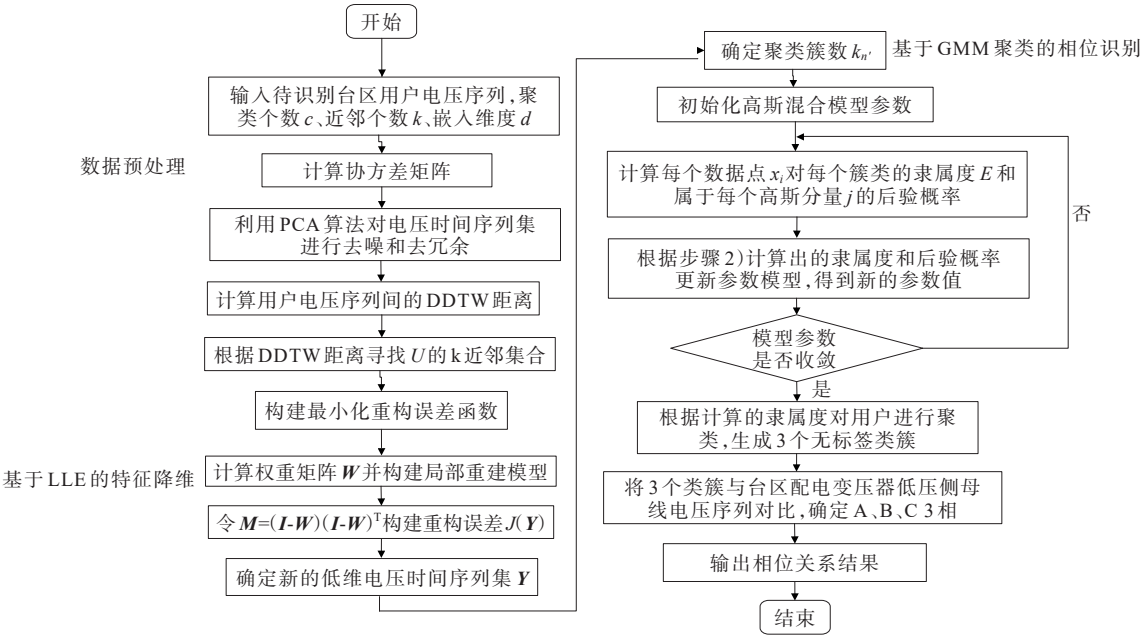


图 5　低压台区相位识别流程

Figure 5　Flow chart of phase recognition in low-voltage station area

情况属于同一簇的数目；d 为试验情况与真实情况都不属于一个类簇的数目。

3　算例分析

3.1　真实台区相位识别结果

本文选取中国河南省某地的实际低压台区进行算例分析。该台区内智能电表电压数据采样频率为 15 min，采集时长为 1 d，构成 96 点电压序列。经实际验证，该台区内一共有 117 个单相用户。部分台区用户日电压曲线如图 6 所示。

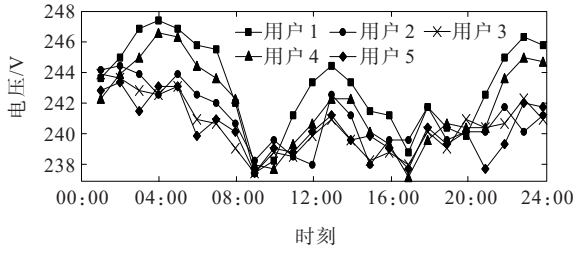
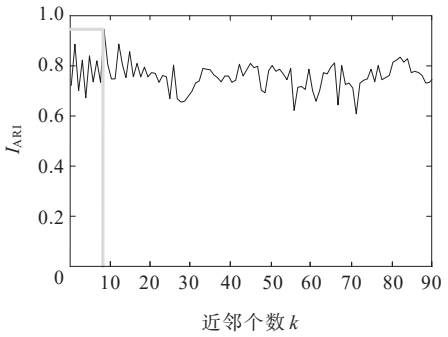


图 6　部分台区用户日电压曲线

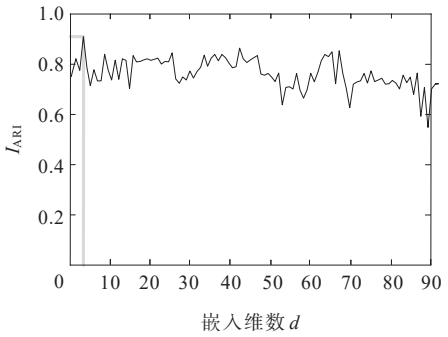
Figure 6　Daily voltage curves of selected users in station area

I_{ARI} 与近邻个数、嵌入维数的关系如图 7 所示。

经过试验，设 $k=8$，$d=2$。利用文 2.2 的方法，获得每个用户跟其他所有的 116 个用户的 DDTW 距离，找到该用户的 K 近邻集合。117 个用户的 DDTW 距离见表 1。在表 1 中，U_i 为第 i 个用户。计算每个用户的重构误差函数并生成权重



(a) I_{ARI} 随着近邻个数 k 的变化



(b) I_{ARI} 随着嵌入维数 d 的变化

图 7　I_{ARI} 与近邻个数、嵌入维数的关系

Figure 7　Variation of I_{ARI} with number of nearest neighbors and number of embedding dimensions

矩阵 $\boldsymbol{W}=[\boldsymbol{W}_1\ \ \boldsymbol{W}_2\ \ \cdots\ \ \boldsymbol{W}_{117}]^{\mathrm{T}}$。求出低维嵌入矩阵 $\boldsymbol{Y}=[1.146\,1\ \ 0.193\,2\ \ \cdots\ \ 0.128\,3]^{\mathrm{T}}$。

对用户电压进行降维之后，进一步采用 GMM 聚类算法对用户进行相位识别。先利用文 2.3 中的

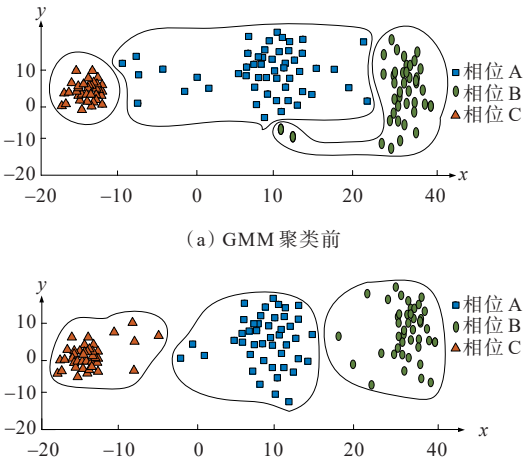
表 1 117 个用户的 DDTW 距离

Table 1 DDTW distance for 117 users

DDTW	U_1	U_2	…	U_{98}	…	U_{116}	U_{117}
U_1	0.00	293.68	…	214.97	…	196.73	12.31
U_2	293.68	0.00	…	21.89	…	41.38	230.54
…	…	…	0.00	…	…	…	…
U_{98}	214.97	21.89	…	0.00	…	3.08	184.66
…	…	…	…	…	0.00	…	…
U_{116}	196.73	41.38	…	3.08	…	0.00	178.73
U_{117}	12.31	230.54	…	184.66	…	178.73	0.00

方法确定聚类簇数 k_n=3，初始化模型参数在采用 EM 算法对模型参数不断进行迭代更新，直到更新之后的模型参数收敛；最后，进行收敛性检查，根据收敛之后的模型参数求出用户 3 个相位的隶属度，并根据式(24)的类簇依据实现对这 117 个用户的相位识别。

本文采用 LLE 方法将用户电压数据降维至 2 维，降维后的电压数据为 $n\times 2$ 维。GMM 聚类算法的相位识别结果如图 8 所示。



(a) GMM 聚类前

(b) GMM 聚类后

图 8 GMM 聚类算法的相位识别结果

Figure 8 Phase recognition results of GMM clustering algorithm

从图 8 中可以看出，该台区一共有 11 位用户存在相位错误的情况，经过 GMM 聚类识别之后这些用户均识别出正确的相位。人工排查确认该低压台区用户相位识别的准确率为 100%。

3.2 算法对比分析

3.2.1 数据缺失情况下不同相似性度量的仿真对比分析

多种外界因素的干扰会导致相位结果不准确。其中，智能电表采集的数据的异常缺失与节点电压波动是最常见的。为比较 DDTW 与 DTW 算法在低压用户的相位识别中的优劣，以中国某地真实台区数据为例，从数据缺失、误差及线路分支数量这 3 个方面对这 2 种算法进行仿真分析。

实际中，不同用户的智能电表可能会存在配置不同与老化问题，这些均可能使采集数据缺失或者采集的用户电压序列数据不等长。为验证本文所提方法在数据缺失条件下相位识别结果的有效性，本文设置不同缺失率下的电压数据，将采用的 DDTW 与 DTW 距离、分段线性法、近邻法、等效线性方法、剪辑近邻法进行比较。不同缺失率下相似性算法的 I_{ARI} 对比见表 2。

表 2 不同缺失率下相似性度量算法的 I_{ARI} 对比

Table 2 Comparison of similarity measurement algorithms in different cases

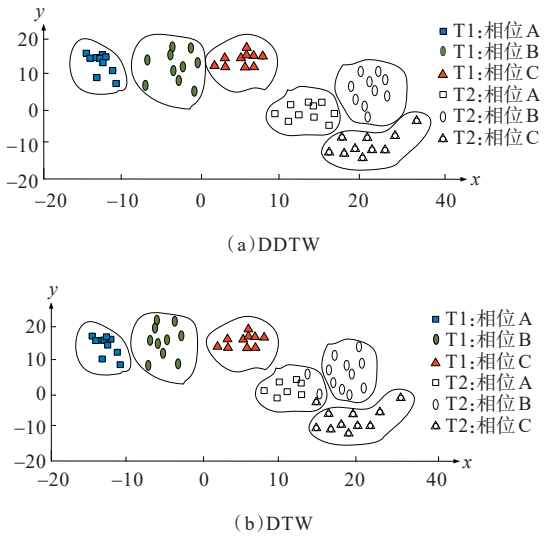
算法	不同缺失率下的 I_{ARI}					
	0%	5%	10%	15%	20%	25%
DDTW	1.00	1.00	0.98	0.97	0.96	0.87
DTW	1.00	0.99	0.96	0.94	0.84	0.8
分段线性法	1.00	0.96	0.91	0.82	0.72	0.63
近邻法	1.00	0.84	0.80	0.82	0.72	0.63
等效线性	1.00	0.95	0.88	0.77	0.62	0.53
剪辑近邻	1.00	0.89	0.78	0.73	0.52	0.46

由表 2 可知，当缺失率为 0% 时，所有距离度量的方法准确率均为 100%；随着缺失率增加，所有算法的准确率均有不同程度的下降；当缺失率为 25% 时，除了 DDTW 和 DTW 这 2 种距离度量方法 I_{ARI} 值高于 0.50 外，其他算法的 I_{ARI} 值均低于 0.70。当智能电表采集的数据一旦缺失，就应采用插值法进行数据补全，但插值方法的不同选择会导致不同程度的误差。由此可见，高维电压时间序列的数据缺失对 DDTW 和 DTW 这 2 种距离度量影响不大。因此，相较于欧氏距离与皮尔逊相似性，DDTW 和 DTW 距离度量算法均具有更高的鲁棒性与准确性。

3.2.2 DDTW 与 DTW 对不同台区用户相位识别的仿真对比分析

为衡量用户之间的接近程度，将 DDTW 与 DTW 算法进行仿真对比分析。为更好地对比这 2 种算法，分别取 2 个不同台区(T1、T2)下的 3 个不同相位的用户进行分析。每个台区每个相位各取 10 个用户，共计 60 个用户。这 60 个用户分别用先 DDTW 和 DTW 算法进行距离度量，再采用 GMM

聚类分析。DDTW 与 DTW 算法的相位识别仿真结果如图 9 所示。



(a)DDTW

(b)DTW

图 9　DDTW 与 DTW 度量算法对不同台区用户的相位识别仿真结果

Figure 9　Simulation results of phase recognition for users in different station areas using DDTW and DTW metric algorithms

从图 9 中可以看出，DDTW 算法的用户相位识别结果与实际用户相位一致，准确率为 100%；在 DTW 算法的用户相位识别结果中，有 3 个用户与实际用户相位不一致，其准确率为 95%。这是因为同一台区下的用户尽管相位不一致，但其相似性仍要远高于不同台区下的用户的相似性。与 DTW 算法相比，DDTW 算法能更好地捕捉复杂动态特征的时间序列特征。因此，在处理复杂的高维电压时间序列上，DDTW 算法的准确性比 DTW 算法的准确性更好。

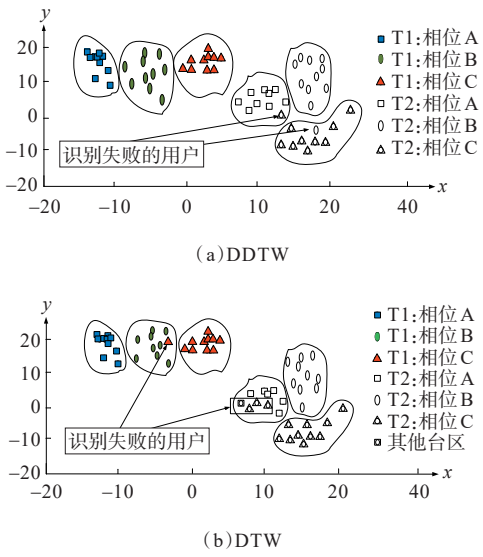
1) 数据缺失。

图 10～12 为不同缺失率下的这 2 种算法的相位识别结果。DDTW 与 DTW 算法在不同缺失率下的相位识别正确率如图 13 所示。

从图 13 中可以看出，当数据正常时，DDTW 算法对这 2 个台区的用户相位识别的准确率高达 100%；当数据缺失率增加时，DDTW 算法对这 2 个台区的用户准确率也远高于 DTW 算法的。

2) 误差及线路分支数量。

对用户电压数据分别施加 ±1.0%、±1.5%、±2.0% 的服从正态分布的扰动。同时，在台区拓扑结构的基础上，增设 3 条分支线路，分别对台区进行若干次仿真试验，得到 DDTW 与 DTW 在不同情况

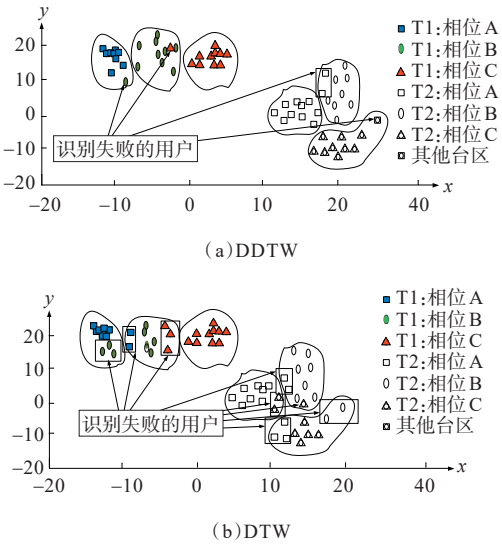


(a)DDTW

(b)DTW

图 10　10% 缺失率下 DDTW 与 DTW 对 2 个台区用户的识别结果

Figure 10　Recognition results of DDTW and DTW for users in two station areas under 10% missing rate



(a)DDTW

(b)DTW

图 11　20% 缺失率下 DDTW 与 DTW 对 2 个台区用户的识别结果

Figure 11　Recognition results of DDTW and DTW for users in two station areas under 20% missing rate

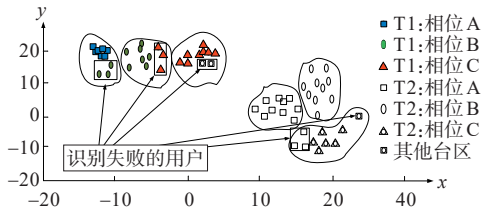
下的相位识别准确率，结果如图 14 所示。

从图 14 中可以看出，当分支线路相同时，对电压施以相同大小的正态扰动时，以 DDTW 算法为基础的相位识别准确率要高于以 DTW 算法为基础的相位识别准确率

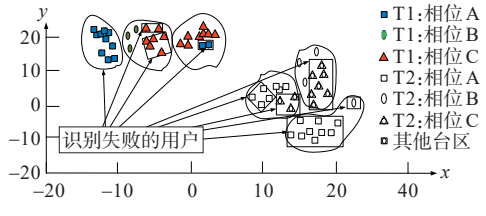
综上所述，DDTW 比 DTW 更适用于低压台区的相位识别。

3.2.3　数据特征降维对比分析

为证明本文所提特征降维方法在相位识别中



(a)DDTW



(b)DTW

图 12 30% 缺失率下 DDTW 与 DTW 对 2 个台区用户的识别结果

Figure 12 Recognition results of DDTW and DTW for users in two station areas under 30% missing rate

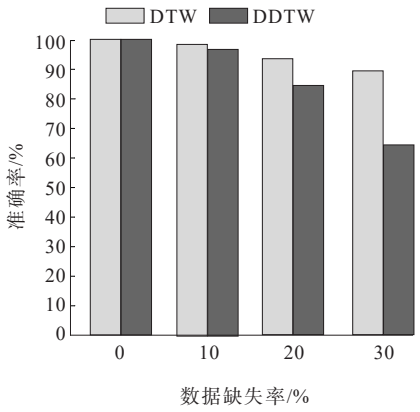


图 13 DDTW 与 DWT 算法在不同缺失率下的相位识别正确率

Figure 13 Phase recognition accuracy of DDTW and DTW algorithms under different missing rates

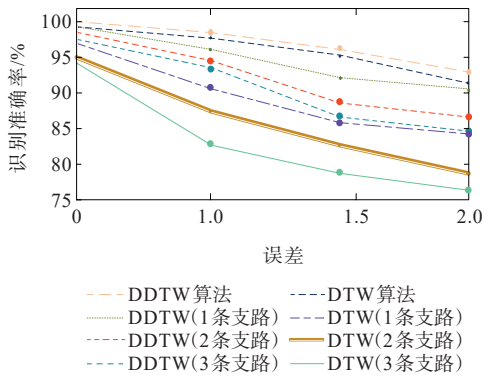


图 14 DDTW 与 DTW 在不同情况下的相位识别平均准确率

Figure 14 Average phase recognition accuracy of DDTW versus DTW under different conditions

的准确性，将本文方法与 PAA（piecewise aggregate approximation）降维、PCA 降维和 t-SNE 降维作比较，并采用 GMM 聚类结合算法（各个方法的聚类簇数均为 3）处理降维后的数据，分别计算每个算法的 I_{ARI}。不同特征降维方法的 I_{ARI} 值与相位识别准确率见表 3。

表 3 不同特征降维方法的 I_{ARI} 值与相位识别准确率

Table 3 ARI values and phase recognition accuracy for different feature dimensionality reduction methods

方法	I_{ARI}	相位识别准确率/%
本文所提方法	1.00	100
PCA	0.80	75
PAA	0.86	90
t-SNE	0.91	93

由表 3 可知，在这 4 种降维算法中，本文所提方法的 I_{ARI} 与准确率最高。其原因可能在于：PCA 是建立在数据是线性假设上的，但用户电压数据是非线性的离散序列，这就导致 PCA 降维会丢失一些重要的电压特征；PAA 算法基于平均值或聚合统计量进行降维，如果电压序列长度较大，单个数点的波动会被平均化，重要特征可能被掩盖；t-SNE 是一种非线性的降维方法，但其对参数的设置非常敏感，这样使得参数对结果影响颇大；本文所采用的基于 DDTW 的 LLE 算法不仅解决了欧氏距离无法适用于高维数据处理的问题，还保留了 PAA 没有的数据的局部结构，其参数也比 t-SNE 的更简单。这些都使得 LLE 的非线性降维效果好于 PCA 和 PAA。

3.2.4 用户相位识别对比分析

为进一步证明本文所提聚类算法在相位识别中的准确性，将本文方法与具有噪声的基于密度的聚类方法（density-based spatial clustering of applications with noise，DASCAN）、近邻传播算法（affinity propagation，AP）和综合层次聚类算法（blanced iterative reducing and clustering，BIRCH）这 3 种算法进行对比分析。分别采用这 4 种聚类算法，对 9 d 的用户电压数据进行试验，计算其相位识别率。这 4 类聚类算法在 9 d 用户数据中的相位识别准确率如图 15 所示。为更直观地评估这 4 种聚类算法，分别计算这 4 种聚类算法的 I_{ARI} 评价指标，结果如图 16 所示。

从图 15、16 中可以看出，本文所提出的聚类算法相较于 DASCAN 算法、BIRCH 算法和 AP 算法。DASCAN 算法对参数的选择较为敏感且难以处理

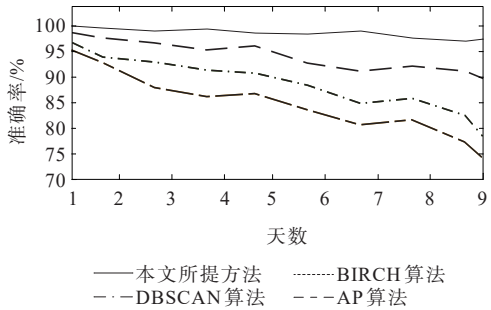


图 15　4 类聚类算法在 9 d 用户数据中的相位识别准确率

Figure 15　Phase recognition accuracy of four clustering algorithms in 9-day user data

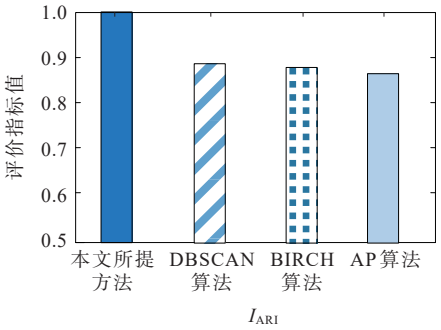


图 16　这 4 种聚类算法下的评价指标对比

Figure 16　Comparison of evaluation metrics across these four clustering algorithms

密度差异较大的数据集，GMM 聚类算法在处理复杂数据分布和生成新数据样本方面更具优势；BIRCH 算法的稳定性受数据集初始读入顺序的影响且须调整多个参数；AP 算法对初始信息选择的敏感；GMM 则不受这些的影响。因此，GMM 在用户相位识别上的自适应性和灵活性方面比 DASCAN、AP、BIRCH 算法更具优势。

4　结语

针对台区用户相位识别准确率低的情况，本文提出一种基于概率分布的聚类算法，并采用中国河南省某地区电网数据进行仿真算例分析，得到以下结论：

1）所提出 DDTW 算法能有效捕捉局部电压数据的形态变化和趋势，通过引入电压时间序列导数，有效解决 DTW 中的奇点问题，精确地确定同一台区与不同台区的用户之间的相似性。同时，该算法可以对高维数据进行降噪和冗余处理，有效提升数据的有效性，保证后续算法的准确性。

2）LLE 特征降维与 GMM 聚类算法不仅能保留数据特征，还能精确识别用户电压形态变化，根据概率对用户进行相位识别。仿真分析结果表明，本文所提方法准确率高于其他方法的。

3）本文所提出的算法仍具有一定的局限性，LLE 特征降维结果受参数如近邻个数 k 的影响，参数须人工设定，这也是今后研究的方向。

参考文献：

[1] 陈招安，黄纯，张志丹，等．基于 T 型灰色关联度和 KNN 算法的低压配电网台区拓扑识别方法[J]．电力系统保护与控制，2021，49(1)：163-169.
CHEN Zhaoan, HUANG Chun, ZHANG Zhidan, et al. Topology identification method of a low voltage distribution network based on T-type grey correlation degree and KNN algorithm[J]. Power System Protection and Control, 2021, 49(1): 163-169.

[2] 戴雪婷，李辉，何柏娜，等．配电网拓扑分析与线路参数辨识研究[J]．电力科学与技术学报，2023，38(4)：46-56.
DAI Xueting, LI Hui, HE Baina, et al. Research on topology analysis and line parameter identification of distribution network[J]. Journal of Electric Power Science and Technology, 2023, 38(4): 46-56.

[3] 刘文彬．基于量测数据的低压配电网拓扑识别[D]．济南：齐鲁工业大学，2023.
LIU Wenbin. Topology identification for low voltage distribution network based on measurement data[D]. Jinan: Qilu University of Technology, 2023.

[4] PENGWAH A B, FANG L X, RAZZAGHI R, et al. Topology identification of radial distribution networks using smart meter data[J]. IEEE Systems Journal, 2022, 16(4): 5708-5719.

[5] HOOGSTEYN A, VANIN M, KOIRALA A, et al. Low voltage customer phase identification methods based on smart meter data[J]. Electric Power Systems Research, 2022, 212: 108524.

[6] DUAN Y L, LIU Z, LIU Y Y, et al. Signal injection-based topology identification for low-voltage distribution networks considering missing data[J]. Energies, 2024, 17(9): 2060.

[7] 黄毕尧，张明，李建岐，等．联合高低频电力线通信的中压配电网拓扑自动识别方法[J]．高电压技术，2021，47(7)：2350-2358.
HUANG Biyao, ZHANG Ming, LI Jianqi, et al. Automatic identification of medium-voltage power distribution network topology based on high and low frequency power line communication[J]. High Voltage Engineering, 2021, 47(7): 2350-2358.

[8] BLAKELY L, RENO M J, FENG W C. Spectral clustering for customer phase identification using AMI voltage timeseries[C]//2019 IEEE Power and Energy Conference at Illinois (PECI). Champaign, IL, USA. IEEE, 2019: 1-7.

[9] 李自强，李睿，孙抗．基于时序拓扑数据分析的电力电缆

局部放电模式识别[J].电子科技大学学报,2024,53(3):440-446.
LI Ziqiang, LI Rui, SUN Kang. Power cable partial discharge pattern recognition based on topological data analysis for time series[J]. Journal of University of Electronic Science and Technology of China,2024,53(3):440-446.

[10] 李方硕,刘丽娜,申杰,等.低压配电网用户相位识别方法研究[J].中国测试,2024,50(3):52-61.
LI Fangshuo, LIU Lina, SHEN Jie, et al. Research on phase identification method of low-voltage distribution network users[J]. China Measurement & Testing Technology,2024,50(3):52-61.

[11] 李克明,江亚群,黄世付,等.基于DTW距离和聚类分析的配电台区低压拓扑结构辨识方法[J].电力系统保护与控制,2021,49(14):29-36.
LI Keming,JIANG Yaqun,HUANG Shifu,et al.Topology identification method of a low-voltage distribution station area based on DTW distance and cluster analysis [J]. Power System Protection and Control, 2021, 49(14):29-36.

[12] 罗钧腾,章坚民,姚力,等.基于电压和功率数据的低压用户相别识别优化建模与应用[J].电力系统自动化,2021,45(7):123-131.
LUO Junteng,ZHANG Jianmin,YAO Li,et al.Modeling and application of phase identification optimization for low-voltage customer based on voltage and power data [J]. Automation of Electric Power Systems, 2021, 45(7):123-131.

[13] 连子宽,姚力,刘晟源,等.基于t-SNE降维和BIRCH聚类的单相用户相位及表箱辨识[J].电力系统自动化,2020,44(8):176-184.
LIAN Zikuan, YAO Li, LIU Shengyuan, et al. Phase and meter box identification for single-phase users based on t-SNE dimension reduction and BIRCH clustering[J]. Automation of Electric Power Systems, 2020, 44(8):176-184.

[14] 闫东辉.基于密度聚类的低压台区归属关系及相位识别方法[J].南方能源建设,2023,10(5):149-156.
YAN Donghui. Identification of low-voltage distribution network attribution relationship and phase information based on density clustering[J]. Southern Energy Construction,2023,10(5):149-156.

[15] 崔雪原,刘晟源,金伟超,等.基于APAA和改进DBSCAN算法的户变关系及相位识别方法[J].电网技术,2021,45(8):3034-3043.
CUI Xueyuan, LIU Shengyuan, JIN Weichao, et al. Consumer-transformer relationship and phase identification based on APAA and improved DBSCAN algorithm[J].Power System Technology,2021,45(8):3034-3043.

[16] 周慷,温立超,杨震一,等.基于改进数据驱动模型的低压台区拓扑关系自动识别技术研究[J].能源与环保,2022,44(11):214-219.
ZHOU Kang,WEN Lichao,YANG Zhenyi,et al.Research on automatic identification of low-voltage substation topology based on improved data-driven model[J].China Energy and Environmental Protection,2022,44(11):214-219.

[17] 张波,赵永红,唐亮,等.基于AMI数据的城市低压配电网拓扑校验方法研究[J].电子测量技术,2019,42(6):67-71.
ZHANG Bo, ZHAO Yonghong, TANG Liang, et al. Research on topology verification of city LV power distribution network based on AMI[J]. Electronic Measurement Technology,2019,42(6):67-71.

[18] ZHANG M Z, LUAN W P, GUO S, et al. Topology identification method of distribution network based on smart meter measurements[C]//2018 China International Conference on Electricity Distribution (CICED). Tianjin, China.IEEE,2018:372-376.

[19] 黄旭波,黄磊,艾丹,等.基于GIS和电压波形相关性分析的低压配电网拓扑识别方法[J].哈尔滨理工大学学报,2023,28(5):42-50.
HUANG Xubo, HUANG Lei, AI Dan, et al. Topology identification method of low-voltage distribution network based on GIS and voltage waveform correlation analysis[J]. Journal of Harbin University of Science and Technology,2023,28(5):42-50.

[20] 张华,龙呈,胡思洋,等.基于层次聚类法与皮尔逊相关系数的配电网拓扑校验方法[J].电力系统保护与控制,2021,49(21):88-96.
ZHANG Hua, LONG Cheng, HU Siyang, et al. Topology verification method of a distribution network based on hierarchical clustering and the Pearson correlation coefficient[J]. Power System Protection and Control, 2021,49(21):88-96.

[21] 宋玮琼,郭帅,李冀,等.基于电压时序数据的配电台区户变关系智能识别[J].电力系统及其自动化学报,2022,34(1):48-55.
SONG Weiqiong, GUO Shuai, LI Ji, et al. Intelligent recognition of user-transformer relationship in distribution station area based on time series voltage data [J].Proceedings of the CSU-EPSA,2022,34(1):48-55.

[22] 叶保璇,张娟,刘恒,等.基于电压相似性聚类的台区相序识别方法[J].机电工程技术,2022,51(8):61-63+198.
YE Baoxuan, ZHANG Juan, LIU Heng, et al. Phase sequence recognition method of station area based on voltage similarity clustering[J]. Mechanical & Electrical Engineering Technology,2022,51(8):61-63+198.

[23] 陈恒,王鹏玮,徐丙垠,等.配电线路自适应相序识别方法[J].电力系统保护与控制,2022,50(6):53-64.
CHEN Heng, WANG Pengwei, XU Bingyin, et al. Adaptive phase sequence identification method for

distribution lines[J]. Power System Protection and Control,2022,50(6):53-64.

[24] ZHOU L, ZHANG Y J, LIU S L, et al. Consumer phase identification in low-voltage distribution network considering vacant users[J]. International Journal of Electrical Power and Energy Systems,2020,121:106079.

[25] 崔阳阳,赵洪山,曲岳晗,等.基于残差U型网络的低压台区电力缺失数据补全方法[J].电力系统自动化,2022,46(9):83-90.
CUI Yangyang,ZHAO Hongshan,QU Yuehan,et al.Residual U-Net based complementation method for missing electricity data of low-voltage stations[J]. Automation of Electric Power Systems,2022,46(9):83-90.

[26] 李开放,林湘宁,李正天,等.基于电压序列最小KL散度索引的低压台区拓扑识别[J].电力系统及其自动化学报,2024,36(10):22-32.
LI Kaifang, LIN Xiangning, LI Zhengtian, et al. Low-voltage substation area topology identification based on minimum KL divergence index of voltage series[J]. Proceedings of the CSU-EPSA,2024,36(10):22-32.

[27] 涂崎,沈冰,邹晓峰,等.5G环境下配网差动保护采样数据缺失应对策略[J].电力工程技术,2022,41(3):143-151.
TU Qi, SHEN Bing, ZOU Xiaofeng, et al. Countermeasures for current differential protection in 5G-based active distribution network under sampling data missing scenario[J]. Electric Power Engineering Technology,2022,41(3):143-151.

[28] 洪文慧,李钦豪,张勇军,等.基于二次矩阵补全的低压配电网相序识别算法[J].电力自动化设备,2022,42(9):133-138.
HONG Wenhui, LI Qinhao, ZHANG Yongjun, et al. Quadratic matrix completion based phase sequence identification algorithm for low-voltage distribution network[J]. Electric Power Automation Equipment, 2022, 42(9):133-138.

[29] 史子轶,夏向阳,刘佳斌,等.基于AKNN异常检验与ADPC聚类的低压台区拓扑识别方法[J].中国电力,2024,57(5):168-177.
SHI Ziyi, XIA Xiangyang, LIU Jiabin, et al. Low-voltage substation area topology recognition method based on AKNN anomaly detection and ADPC clustering[J]. Electric Power,2024,57(5):168-177.

[30] YI Y Q, LIU S L, ZHANG Y J, et al. Phase identification of low-voltage distribution network based on stepwise regression method[J]. Journal of Modern Power Systems and Clean Energy,2023,11(4):1224-1234.

[31] 唐捷,蔡永智,李其霖,等.基于三相表特征约束聚类的低压台区用户相序识别方法[J].电力系统自动化,2022,46(8):60-67.
TANG Jie, CAI Yongzhi, LI Qilin, et al. Phase sequence identification method for users in low-voltage distribution-station area based on feature constraint clustering of three-phase meter[J]. Automation of Electric Power Systems,2022,46(8):60-67.

[32] ZHAO Y, TU Z D, XIAO Y, et al. Electrical principles of verification of the topological structure of a low-voltage distribution network based on voltage from an Advanced Metering Infrastructure[J]. The Journal of Engineering, 2019,2019(11):8218-8224.

[33] 郑英丽,朴丽莎,王丽珍.Pearson相关系数下非对称相似度计算及其应用[J].云南民族大学学报(自然科学版),2024,33(6):736-745.
ZHENG Yingli, PIAO Lisha, WANG Lizhen. Asymmetric similarity calculation and application under Pearson correlation coefficient[J]. Journal of Yunnan Minzu University (Natural Sciences Edition),2024,33(6):736-745.

[34] 刘苏,黄纯,侯帅帅,等.基于DDTW距离与DBSCAN算法的户变关系识别方法[J].电力系统自动化,2021,45(18):71-77.
LIU Su, HUANG Chun, HOU Shuaishuai, et al. Identification method for household-transformer relationship based on derivative dynamic time warping distance and density-based spatial clustering of application with noise algorithm[J]. Automation of Electric Power Systems,2021,45(18):71-77.

[35] 张华.基于DDTW的非线形建筑数字图像形貌拼接仿真[J].计算机仿真,2020,37(6):448-451+470.
ZHANG Hua.Simulation of digital image mosaic of non-linear building based on DDTW[J].Computer Simulation, 2020,37(6):448-451+470.

[36] 杨佳沛,王宇,彭广建,等.基于DTW-GMM的光纤传感系统声纹识别方法[J].电子测量与仪器学报,2024,38(4):176-186.
YANG Jiapei, WANG Yu, PENG Guangjian, et al. Voiceprint recognition method of optical fiber sensing system based on DTW-GMM[J]. Journal of Electronic Measurement and Instrumentation,2024,38(4):176-186.

[37] 杨秋颖,翁小清.基于LLE和高斯混合模型的时间序列聚类[J].计算机技术与发展,2022,32(8):33-41.
YANG Qiuying, WENG Xiaoqing. Time series clustering based on LLE and Gaussian mixture model[J]. Computer Technology and Development,2022,32(8):33-41.

[38] QIAN P J, CHUNG F L, WANG S T, et al. Fast graph-based relaxed clustering for large data sets using minimal enclosing ball[J].IEEE Transactions on Systems, Man, and Cybernetics Part B, Cybernetics, 2012, 42(3): 672-687.