

基于关联特征筛选的双层聚类区域配变电压越限特征分析及预测

郭少东¹, 赵晓莉¹, 孙改平¹, 杨秀¹, 杨帆², 刘俊²

(1. 上海电力大学电气工程学院, 上海 200090; 2. 国家电网上海电力科学研究院, 上海 200437)

摘要: 针对区域配网变压器(简称“配变”)数量多, 大量新型负荷、分布式光伏等接入, 配变电压随机性波动增强, 台区用户电压质量面临挑战。为更好地对区域配变电压进行越限特征分析及预测, 提出了基于关联特征筛选的双层聚类区域配变电压预测方法。首先, 将区域配变的越限天数作为第一层聚类特征, 获得电压性质正常以及越上限的配变。其次, 针对电压越限配变提出结合 Pearson 相关系数和欧氏距离(Euclidean distance)的最优度量矩阵, 提取原有数据的内含信息, 作为 K 均值聚类(K-means)的输入, 实现对区域配变双层聚类。在此基础上, 选取该集群中代表配变表征该类配变, 利用卷积双向长短期记忆网络-注意力机制(convolutional neural network-bidirectional long and short-term memory-attention, CNN-BiLSTM-Attention)模型对配变电压进行预测, 该模型能够提取输入数据的双向信息特征, 并对重要特征加权, 从多时间尺度上获得双向特征信息用于预测。最后, 在上海市某区域配变验证了该方法的有效性。
关键词: 区域配变; 最优度量矩阵; 双层聚类; 降维; 电压预测

Analysis and Prediction of Over-Limit Characteristics for Regional Distribution Transformer Voltage Using Bilayer Clustering by Correlation Feature Screening

GUO Shaodong¹, ZHAO Xiaoli¹, SUN Gaiping¹, YANG Xiu¹, YANG Fan², LIU Jun²

(1. School of Electrical Engineering, Shanghai University of Electric Power, Shanghai 200090, China; 2. State Grid Shanghai Electric Power Research Institute, Shanghai 200437, China)

Abstract: Aiming at the large number of regional distribution transformer, a large number of new loads, distributed photovoltaics, etc., and the enhancement of the random voltage fluctuation of distribution transformer. The voltage quality of the substation users is facing challenge. In order to better analyze and predict the over-limit characteristics of regional distribution transformer voltage, a bilayer clustering regional distribution transformer voltage prediction method based on correlation feature screening is proposed. Firstly, the number of overrun days of regional distribution transformers are taken as the first layer clustering feature, and the distribution transformers with normal and over-limit voltage properties are obtained. Secondly, for the over-limit voltage distribution transformers, an optimal metric matrix combining Pearson's correlation coefficient and Euclidean distance is proposed to extract the contained information of the original data as the input of K-means to realize the bilayer clustering of regional distribution transformer. On this basis, the representative distribution transformers in the cluster are selected to characterize the distribution transformers of this category, and the convolutional neural network-bidirectional long and short-term memory-attention (CNN-BiLSTM-Attention) model is used to predict the distribution transformer voltage, which can extract the bidirectional information features of the input data, weight the important features, and obtain the bidirectional feature information from multiple time scales for prediction. Finally, the effectiveness of the proposed method is verified in a certain area of Shanghai.

Key words: regional distribution; optimal metric matrix; bilayer clustering; dimensionality reduction; voltage prediction

基金项目: 国家自然科学基金资助项目(52207121); 上海电力人工智能工程技术研究中心项目(19DZ2252800)。

Foundation item: Supported by the National Natural Science Foundation of China (52207121); the Project of Shanghai Engineering Research Center of Electric Power Artificial Intelligence (19DZ2252800)。

0 引言

随着新能源技术的发展和非线性负荷大规模接入台区中,台区电能质量问题日益突出^[1]。台区管理作为电网“四分管理”的重要组成部分,早已加入到供电企业的管理日程中,并趋向精细化,其配变电能质量指标也备受电力公司关注^[2]。电压偏差作为电能质量评估的一部分,其对电能质量有着重要影响,且配变作为台区电源,其电压直接影响台区用户电能质量^[3-4]。因此,对电压趋势变化进行预测分析引起运维人员的重视,可以提高电网运行质量,对指导配网改造更具价值^[5-7]。

随着配网数据监测系统不断覆盖,海量监测数据被传输到数据中心,使得通过数据挖掘技术研究配网电压变化成为可能^[8]。

电压稳态质量问题影响范围广、程度深,长期以来备受关注^[9]。电压指标预测与电力负荷预测在某种程度上存在相似性,两者均通过分析历史数据的特征并总结其变化规律实现对未来时间段的预测。当前预测方法包括时间序列、随机森林、回归分析、支持向量机等^[10-11]。随着人工智能技术的发展,机器学习在时序预测方面做出了重大贡献。常用的机器学习算法有长短期记忆神经网络^[12](long short-term memory, LSTM)、反向传播神经网络^[13](back propagation, BP)等。

文献[10]分别通过线性回归模型、随机时间序列模型、灰色模型和优选组合模型对电能质量稳态月度指标进行预测,时间跨度大,不能反映短期变化趋势。文献[14]将影响电压偏差因素进行降维聚类处理,利用BP神经网络提高了电压偏差预测的准确度,但利用气象因素等对电压偏差进行拟合,预测结果不稳定。文献[15]首先通过自回归滑动平均模型(auto regressive moving average model, ARMA)对电压趋势进行拟合,再利用BP模型将ARMA模型的预测值、温度等作为输入,得到预测差值,预测值与差值共同组成最终预测结果,实现对某变电站电压进行预测,但其分段预测效率较低。文献[16]首先采取自回归积分滑动平均模型(auto regressive integrated moving average model, ARIMA)预测有功功率,考虑有功功率和电能质量稳态指标之间的相关性,采用BP神经网络对电能质量指标进行拟合,其预测精度一定程度上受限于

有功功率的预测结果。

鉴于区域配变台数众多,为了便于快速分析越限配变电压特征,并针对越限配变做出电压预测,本文通过双层K-means聚类提取越上限配变的电压特征,并利用CNN-BiLSTM-Attention方法对越上限类别下的配变电压预测,并与未聚类的双向长短期记忆网络(bidirectional long and short-term memory, BiLSTM)以及CNN-BiLSTM-Attention进行对比,验证了该方法在电压预测的精度。

1 预测模型系统框架

预测模型系统框架如图1所示。本文主要分为两部分,首先第一部分为对配变电压进行特征提取,采用K-means对配变电压特征双层聚类,区分出越上限配变月电压运行趋势。第二部分采用CNN-BiLSTM-Attention方法对越上限类别配变进行电压预测。

区域配变数量多,电压运行状态复杂,配变电压在日区间段内常表现为连续一段时间越上限。因此,在一阶段聚类中选取越限天数等指标作为K-means聚类指标,以区分出正常配变与越上限配变。基于越上限配变月电压运行特征,考虑月电压最大值、最小值、平均值、中位数、不合格率等指标表征配变该月整体电压运行情况,结合Pearson相似矩阵与欧氏距离矩阵,求解出最优度量矩阵,利用多维尺度分析(multi-dimensional scaling, MDS)对最优度量矩阵进行降维,将降维后的数据输入到K-means中进行第二层聚类。第二部分结合CNN提取特征、BiLSTM输入历史和未来信息以及Attention提升模型的表达能力和推理能力的优点,在第二层聚类获取越上限配变不同的电压运行特征后,从各越上限类别中选取典型配变采取CNN-BiLSTM-Attention方法进行电压预测。

1.1 配变运行特征筛选

单台配变电压数据采集数量大,数据维度高,配变运行情况复杂,采取特征选取的方法以表征各台配变整体运行趋势,具体特征选取如下。

1) 越限天数的确定

为区分配变越限情况,一阶段聚类以配变正常天数、长时越上限天数以及短时越上限天数为聚类依据。

正常天数:配变电压正常运行范围为(-10%,

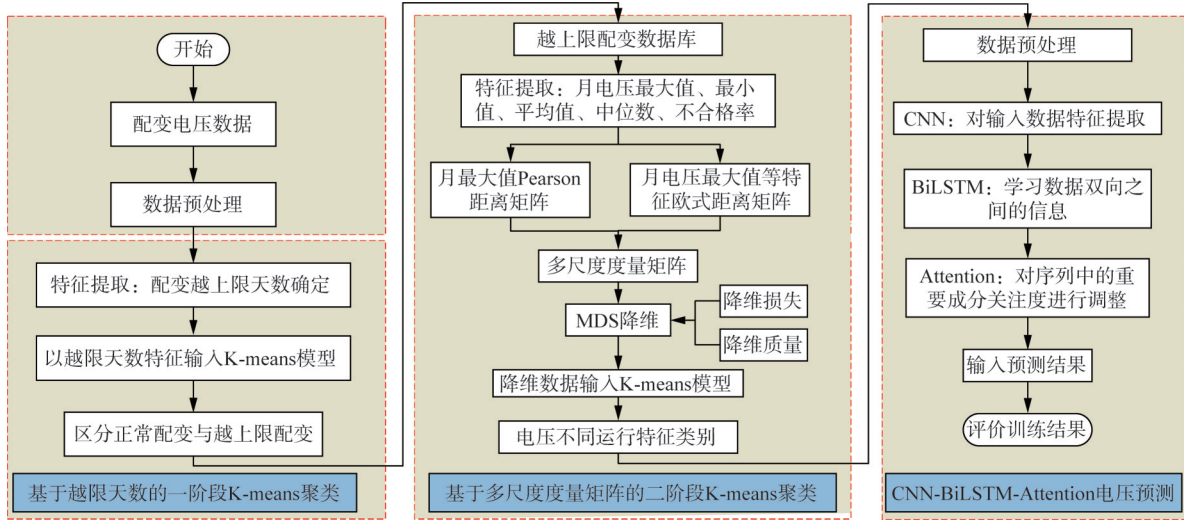


图1 预测模型系统框架

Fig. 1 System framework of predictive model

7 %), 若当日电压波动均在此范围内, 则该日为正常, 并统计天数。

长时越上限天数: 若当日电压数据连续采集超 6 h 均高于 +7 %, 则该日为长时越上限, 并统计天数。

短时越上限天数: 若当日电压数据出现越上限但未出现连续超 6 h 均高于 +7 %, 则该日为短时越上限, 并统计天数。

2) 月电压运行特征提取

对于越上限配变, 按月时间尺度统计电压最大值、电压最小值、电压平均值、中位数以及电压不合格率, 以实现单台配变高维数据的表征。

月电压最大值、最小值、平均值以及中位数均为统计学概念, 可以表征该配变月电压的整体波动范围和分布情况。电压不合格率反映该配变的电压稳定性与电压质量, 其中电压不合格率计算如下。

$$R_k = \frac{c_k}{C_k} \quad (1)$$

式中: R_k 为第 k 个月电压不合格率; c_k 为第 k 个月越限采集数; C_k 为第 k 个月采集总数。

1.2 基于最优度量矩阵输入的 K-means 聚类方法

基于首层聚类结果, 针对越上限配变更关注月电压最大值, 因此, 构建 Pearson 相关系数矩阵与欧氏距离矩阵相结合最优度量矩阵, 提取原数据所包含的潜在信息进行 K-means 聚类, 既可以提取月电压最大值运行趋势相似性, 又保留了各配变间月电压运行的欧氏距离特征。

1.2.1 最优度量矩阵构造

n 台配变数据集矩阵 S_n 中 $X_i (i=1, 2, \dots, n)$ 为每一台配变的月电压特征数据, 数据集矩阵 S_n 如下。

$$S_n = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = [X_1, X_2, \dots, X_n]^T \quad (2)$$

式中 p 为配变月电压特征维度。

对于性质不同的配变关注的特征不同, 对重点关注特征计算 Pearson 相关系数矩阵 R , 其中元素 r_{ij} 为:

$$r_{ij} = 1 - \frac{\left| \sum_{l=1}^p (x_{il} - \bar{X}_i)(x_{jl} - \bar{X}_j) \right|}{\sqrt{\sum_{l=1}^p (x_{il} - \bar{X}_i)^2 \sum_{l=1}^p (x_{jl} - \bar{X}_j)^2}} \quad (3)$$

式中: $i, j \in (0, n]$; $l \in (0, p]$; r_{ij} 越小, 则 X_i 与 X_j 越相关。

对于配变的关联特征则计算欧式距离矩阵 D , 其中元素 d_{ij} 为:

$$d_{ij} = \sqrt{\sum_{l=1}^p (x_{il} - x_{jl})^2} \quad (4)$$

欧式距离矩阵 D 与 Pearson 相关系数矩阵 R 在量纲上存在不同, Pearson 相关系数矩阵 R 的取值范围在 $[0, 1]$ 。因此, 对 D 中元素进行归一化, 得到归一化后矩阵 D' 的元素 d'_{ij} 为:

$$d'_{ij} = \frac{d_{ij} - d_{\min}}{d_{\max} - d_{\min}} \quad (5)$$

式中: d'_{ij} 为 $\mathbf{D}'$ 中的元素; $d_{\max}$ 、 $d_{\min}$ 分别为 $\mathbf{D}$ 中元素的最大值与最小值。

则最优距离矩阵 $\mathbf{M}$ 可表示为:

$$\mathbf{M} = \lambda_1 \mathbf{R} + \lambda_2 \mathbf{D}', \lambda_1 + \lambda_2 = 1 \quad (6)$$

式中 λ_1 、 λ_2 分别为权重系数。

1.2.2 降维方法及选取指标

由于最优距离矩阵往往是一个 $s \times s$ 的高维数组 (s 为越上限类配变台数), 高维数组往往容易出现维度诅咒 (curse of dimensionality), 距离的概念变得模糊, 因此须对其进行降维后再聚类。

1) 多维尺度分析方法

多维尺度分析 (MDS)^[17] 是目前对高维数据处理的一种常见降维方法, 通过低维空间 (c 维) $\mathbf{R}^c$ ($c < C$) 重新表示高维空间 (C 维) $\mathbf{R}^C$ 的多个样本间的距离或者相似度, 挖掘数据间的潜在性质, 并且通过降维实现数据的可视化, 具有较好的降维质量与降维速率。

欧氏距离 (euclidean distance) 是常用的聚类相似性度量方法, 因此, 本文以欧氏距离作为 MDS 算法的相似性度量。MDS 要求低维空间数据间的距离尽可能与原始空间数据间的距离保持一致, 将高维度数据集降维至低维度数据集, 能够直观地反映出数据间的结构关系, 因此 MDS 降维算法的目标函数可采用 SMACOF 等优化方法对 MDS 进行优化求解, 目标函数 $f(i, j)$ 为:

$$f(i, j) = \min \sum_{i < j} (e_{ij} - E_{ij})^2 \quad (7)$$

式中: e_{ij} 为降维后样本 i 和样本 j 间的欧氏距离; E_{ij} 为原始空间下样本 i 和样本 j 间的欧氏距离。

2) 降维评估指标

目前, 对于降维方法的评估指标主要集中在三个方面, 分别为降维速率、降维损失以及降维质量。其中, 降维速率指的是降维方法将高维数据转换为低维数据的效率; 降维损失是指在数据降维过程中所带来的信息损失程度; 而降维质量则是指降维后数据的表现和保留的特征质量。本文参考文献 [18-19] 从压力函数 (stress) 以及局部连续性元标准 (the local continuity meta-criterion, LCMC) 分别计算降维损失、降维质量, 从而确定降维方法的选取。

压力函数表示原始数据间距离与低维映射中输出距离之间的差异, 其值越小表示降维损失越小。

压力函数 S_{stress} 可表示为:

$$S_{\text{stress}} = \sqrt{\frac{\sum_{i=1, j=1, i \neq j}^n (e_{ij} - E_{ij})^2}{\sum_{i=1, j=1, i \neq j}^n E_{ij}^2}} \quad (8)$$

LCMC 是一种广泛应用于降维质量评估的指标, 其定义为高维空间中 l 近邻与低维映射之间重叠的平均次数。通过定义两个独立于尺度的评估指标 Q_{local} 和 Q_{global} , 可以对降维算法进行更加细致的评估。其中, Q_{local} 和 Q_{global} 分别为具有最高 L_{LCMC} 值数据坐标的左侧和右侧重叠次数的均值, Q_{LG} 为 Q_{local} 和 Q_{global} 两者的平均值。

$$L_{\text{LCMC}} = \frac{1}{QN} \sum_{\alpha, \beta \in [1, k]} q_{\alpha, \beta} - \frac{Q}{N-1} \quad (9)$$

式中: Q 为领域大小; N 为样本数量; $q_{\alpha, \beta}$ 为降维前后重叠数量。

1.2.3 K-means 聚类方法

K-means 是种常见的聚类算法^[20], 式 (6) 求得的最优距离矩阵 $\mathbf{M}$ 是 $m \times m$ 高维数据 (m 为首层聚类后各类别的数量), 将最优距离矩阵 $\mathbf{M}$ 降维后输入 K-means 模型中, 得到第二层聚类结果。基本思想是通过迭代的方式, 将数据划分为 K 个不同的簇, 并使得每个数据点与其所属簇的质心之间的距离之和最小。K-means 流程示意图如图 2 所示。

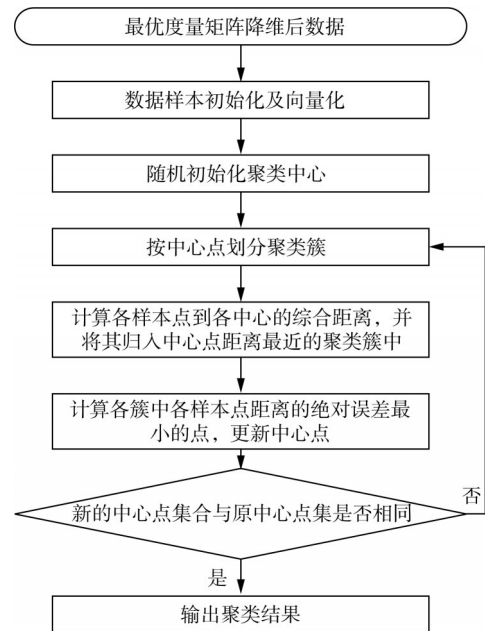


图 2 K-means 聚类算法流程图

Fig. 2 Flow chart of K-means clustering algorithm

2 CNN-BiLSTM-Attention预测模型

CNN具有局部连接、权重共享等特点,优化了输入BiLSTM模型的电压序列,使得模型可以更好地理解电压序列中的波动性和尖峰特点。然而,由于长序列存在依赖关系,仍然难以解决其预测问题。为此,本文提出了一种基于注意力机制的CNN-BiLSTM电压预测模型,该模型能够捕捉长序列中的依赖关系,并且能够突出对电压变化产生影响的因素,以提高预测性能。

在CNN-BiLSTM-Attention串联模型中,CNN的作用是对输入电压数据进行特征提取和降维,以减少BiLSTM模型的参数数量,提高训练速度和模型的泛化能力,BiLSTM掌握序列之间的规律,负责对时序电压进行预测,Attention的作用是对影响电压因素的关注意度进行调整,使模型能够更加关注序列中的重要部分,并减少对无关部分的注意力,从而提高模型的预测精度。模型流程如图3所示。

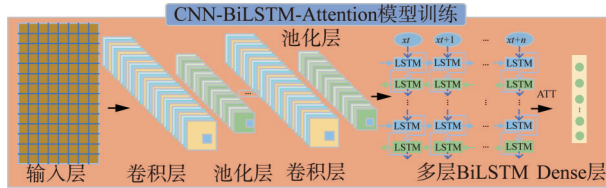


图3 CNN-BiLSTM-Attention组合预测模型

Fig. 3 Combined predictive model of CNN-BiLSTM-Attention

2.1 卷积神经网络

CNN是一种经典的深度学习模型,其基本结构由输入层、卷积层、池化层以及输出层构成。卷积层是CNN的核心,通过应用卷积核来提取输入电压时间序列的局部特征^[21]。池化层通过减少卷积层之间的参数量和电压时间序列进行降维及二次特征提取,进一步增强模型的鲁棒性。卷积层和池化层构成了卷积单元,可通过堆叠多个卷积单元提高模型的深度和表达能力。最后,通过全连接层将所有卷积单元提取的局部特征汇总,实现对电压时间序列的综合特征表达。

2.2 双向长短期记忆递归神经网络

BiLSTM源于LSTM,其核心思想是利用正向和反向的LSTM模型处理输入数据,并将两个方向的输出连接起来^[22]。因此,BiLSTM模型的输出涵盖了输入电压数据过去和未来的双向特征信息,如

图4所示。BiLSTM擅长发现电压数据长时间序列的内在规律,并且在训练过程中采用了权值共享的策略,不仅可以保证模型具有更好的泛化能力,而且在不增加数据量的情况下提高了模型的鲁棒性^[23]。LSTM方法具体过程参考文献[24]。

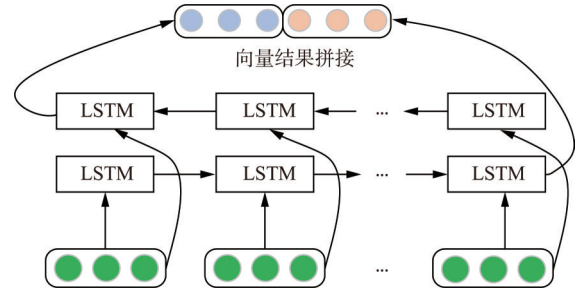


图4 BiLSTM示意图

Fig. 4 Schematic diagram of BiLSTM

2.3 注意力机制

注意力机制^[25-26]有助于提升模型的表达能力和推理能力。通过学习输入序列中不同位置的关联性和重要性,注意力机制能够捕捉到序列中的长程依赖和上下文信息,从而更好地理解和表示输入数据。这使得模型能够进行更精确的推断和预测,提高了其在各种任务中的性能。本文针对BiLSTM层在处理多维数据组成的长时间序列提取最终矢量出现偏差的情况,引入了注意力机制。方法是基于注意力机制为BiLSTM层的输出结果合理分配注意力权重,更新其输出结果以减小偏差。

2.4 评价指标

本文选取均方根误差((root mean squared error, RMSE)、平均绝对百分误差(mean absolute percentage error, MAPE),确定系数 R^2 ,各指标计算公式如下所示。

$$e_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^t (y_i - \hat{y}_i)^2}{t}} \quad (10)$$

$$e_{\text{MAPE}} = \frac{1}{t} \sum_{i=1}^t \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^t (y_i - \hat{y}_i)^2}{\sum_{i=1}^t (\bar{y} - \hat{y}_i)^2} \quad (12)$$

式中: t 为预测数; y_i 为电压实际值; $\bar{y}$ 为电压实际值的平均值; $\hat{y}_i$ 为电压预测值。

3 算例分析

本文选取上海市某区域内 690 台配电变压器作为分析样本, 低压侧电压采集数据从 2022 年 1 月 1 日至 2022 年 12 月 31 日, 共 325 d 有效采样天数, 采样间隔为 15 min。因 1 月以及 2 月采集缺失天数多, 在二阶段聚类时, 仅以 3—12 月的 306 d 数据进行分析。气象数据来自于上海市虹桥机场历史天气, 采样间隔为 30 min, 为了与电压数据采集频率相同, 采用插值法对气象数据进行插补。

3.1 基于配变运行特征的双层聚类研究

3.1.1 基于越限天数的 K-means 首层聚类

对 690 台配变利用 K-means 聚类, 并用 Davies-Bouldin 指数确定最佳聚类个数为 2, 类别 A 主要为配变正常运行类别, 后续对其不作关注, 而类别 B 主要为越上限配变, 聚类中心如表 1 所示。

表 1 基于电压越限天数的 K-means 聚类

Tab. 1 K-means clustering based on days of over-limit-voltage

聚类中心	长时越限天数/d	短时越限天数/d	正常天数/d	配变台数
类 A	3. 79	5. 90	315. 31	280
类 B	265. 45	32. 50	27. 05	410

3.1.2 基于最优测量矩阵的 K-means 聚类

在首层聚类后, 类 B 为电压越上限配变, 对月电压最大值 Pearson 度量矩阵施加更大权重。

针对最优距离矩阵的高维度特征, 采用 MDS、PCA、t-SNE 等降维方法, 通过降维质量和降维损失等评估指标, 对降维效果进行综合评价, 以确定选取哪种降维方法。降维评价如表 2 所示, MDS 相较于 PCA 以及 t-SNE 具有更低的降维损失以及更高降维质量。

表 2 降维方法评价指标

Tab. 2 Evaluation indexes of dimensionality reduction methods

降维方法	降维损失	降维质量		
		Q_{local}	Q_{global}	Q_{LG}
t-SNE	0. 335 9	0. 718 2	0. 810 1	0. 764 2
PCA	0. 933 0	0. 585 1	0. 892 6	0. 738 9
MDS	0. 001 0	0. 875 9	0. 956 8	0. 916 4

在对最优度量矩阵利用 MDS 降维后, 将降维后的数据输入到 K-means 模型中, 得到聚类结果。

基于本文关联特征筛选及最优测量矩阵输入的双层聚类方法, 可以较好地分类出月电压最大值在

数值以及变化趋势的不同。类别 B 的聚类个数为 6, 类 B 的聚类效果图如图 5 所示。从电压月最大值分布趋势看, 图 5(b) 和图 5(d) 月电压最大值幅值相近, 但运行趋势不同, 图 5(b) 月电压最大值在 12 月时有明显上升; 图 5(a) 与图 5(b) 在月电压最大值运行趋势上相近, 但月电压最大值的幅值不同, 图 5(b) 的月电压最大值整体高于图 5(a)。

从月电压最大值运行趋势上来看, 图 5(a)、图 5(b)、图 5(d) 运行状态都较为平稳, 但图 5(a)、图 5(b) 均呈现出夏季、秋季电压运行平稳, 在 12 月份时电压出现上升。图 5(c)、图 5(e)、图 5(f) 电压运行状态波动性大, 图 5(c) 呈现出“V”字形走势, 夏季高温天气时电压出现下降; 图 5(f) 呈现出“M”形, 春季电压升高, 夏季电压最低, 秋、冬季电压先升高后降低; 图 5(g) 电压在 11 月出现大的波动, 其他月份相对平稳。

3.2 配变低压侧电压预测

根据双层聚类算法对区域配变进行集群聚类, 挑选出不同运行性质的配变, 类别 B 电压长期偏高, 合格率一般, 对电网安全运行不利, 因此, 仅针对越上限配变进行电压预测。根据图 5 聚类种类划分, 选择各聚类类别下的典型配变, 共 6 台配变作为电压预测的代表配变。

3.2.1 数据预处理

将数据集划分为训练集、验证集和测试集。其中训练集和验证集被用于模型的训练和验证, 而测试集则用于最终的预测任务。

3.2.2 输入特征选择

电压的波动往往受气象因素变化所影响。通过计算 Spearman 相关系数评估配变电压同气温、大气压强、湿度、风向、风速、露点温度的 Spearman 相关性, 如表 3 所示。

为提高模型预测的准确性, 将与配变电压相关性系数小于 0.1 的气象因素剔除。温度与湿度对配变电压变化影响较高, 具体而言, 温度与配变电压呈负相关, 而湿度与配变电压呈现出正相关性。电压具有周期效应和较强的自相关性, 时间尺度可以反映电压的变化规律。采用滑动窗口的方式构建短时预测模型的输入和输出, 输入数据包括各历史时刻的温度、湿度、露点温度、大气压强、时间以及电压数据, 输出数据为预测时刻的电压数据。

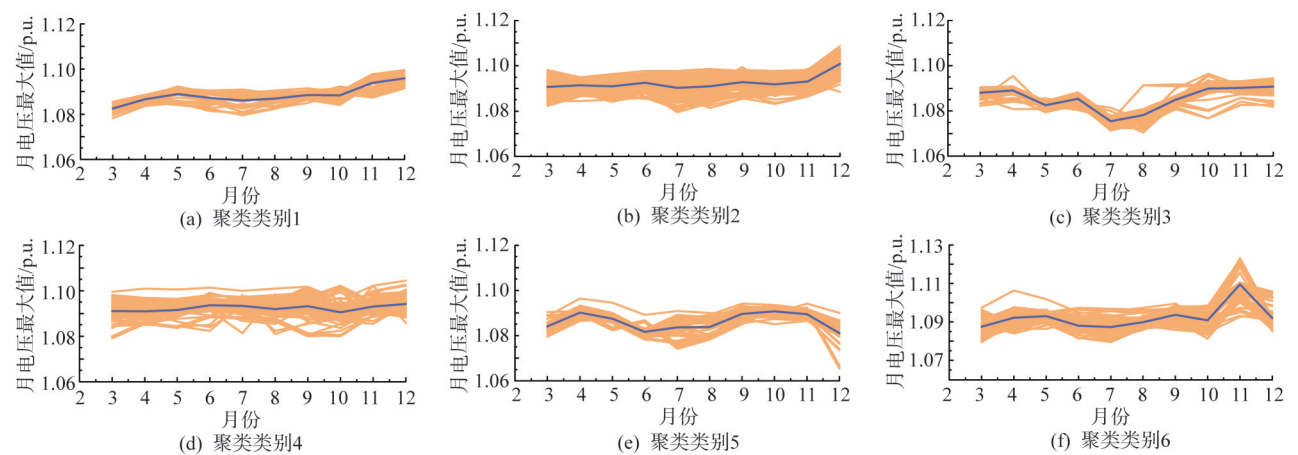


图5 越上限电压配变类B聚类结果效果图

Fig. 5 Effect of cluster results of over-upper limit voltage distribution transformers matching variant B

表 3 配变电压与气象因素 Spearman 相关性系数

Tab. 3 Correlation coefficients between distribution transformers voltages and Spearman meteorological factors

气象因素	越上限配变
温度	-0.351 8
大气压强	0.214 1
湿度	0.264 2
风向	0.039 6
风速	-0.219 1
露点温度	-0.249 6

表 4 类别 B 多个模型预测结果对比

Tab. 4 Comparison of prediction results of multiple models in category B

模型		评价指标		
		RMSE 值	R^2	MAPE 值/%
未聚类	BiLSTM	0.003 09	0.917 76	0.200 82
	CNN-BiLSTM-Attention	0.002 93	0.927 80	0.196 58
双层聚类	本文算法	0.002 53	0.942 46	0.163 64

3.3 实验结果对比分析

为验证本文方法的预测结果具有更高的准确性,采用基于注意力机制的CNN-BiLSTM模型对6个类别中的典型配变进行电压预测。预测集为2022年11月1日至2022年12月31日,预测点时间间隔为15 min。根据预测结果,计算各预测模型的RMSE值、 R^2 、以及MAPE值。

考虑到神经网络每次预测结果可能会有一定的差异,本文对越上限类别下的6台配变分别构建预测模型,采取了多次预测并对评价指标取平均值的方法,该评估策略可以提高预测结果的稳定性和可靠性,并更准确地比较不同模型的预测性能。各预测算法对越上限配变的预测指标如表4所示。

通过对评价指标的计算结果进行分析,可以观察到本文的预测方法在各个指标上都优于未聚类的BiLSTM和CNN-BiLSTM-Attention预测模型,预测精度更高。本文算法相较于未聚类的BiLSTM算法在RMSE值、 R^2 上分别提升了18.12%、2.69%。测试集中,越上限配变6种聚类类别的日RMSE值

对比曲线如图6所示,对比各日的RMSE值,本文方法取得了更好的预测结果。由于电压波动较为剧烈,曲线呈现出明显的锯齿状特征,基于注意力机制的CNN-BiLSTM预测模型能够更准确地捕捉到电压波动的信息,从而提高了预测的准确性。

4 结论

本文提出一种基于关联特征筛选及最优度量矩阵双层聚类的电压预测方法,以上海市某区域配变电压数据为例,对预处理后的大规模电压数据利用月电压特征表征该配变,采用双层聚类算法有效地实现了配变电压运行特征模式的分类效果。算例结果表明:

1) 基于关联特征筛选的双层聚类可以更加有效地区分越限配变与正常配变,并对越限配变不同运行特征进行分类,有利于电力公司对不同配变电压运行类型采取不同的治理手段。

2) 基于双层聚类后的配变电压预测模型与未聚类的BiLSTM和基于注意力机制的CNN-BiLSTM预测模型相比,具有更好的预测效果。

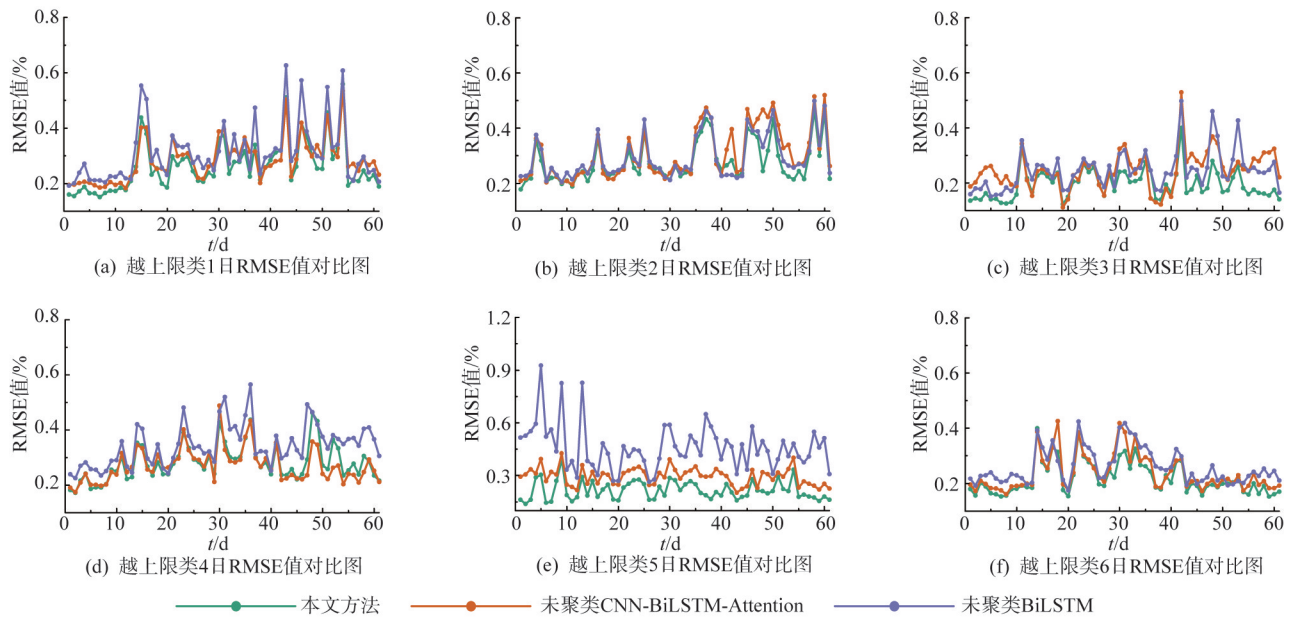


图6 越上限电压各类配变单日预测对比图

Fig. 6 Comparison charts of single-day prediction of various types of distribution transformer beyond the upper limit voltage

参考文献

- [1] 刘科研, 盛万兴, 秦梦雅, 等. 计及多种扰动源的有源配电网电能质量分析方法[J]. 电网技术, 2020, 44(6): 2303–2309. LIU Keyan, SHENG Wanxing, QIN Mengya, et al. Power quality analysis for active distribution network considering multiple disturbance sources [J]. Power System Technology, 2020, 44(6): 2303–2309.
- [2] 张敏. 基于物联网技术的配电台区拓扑识别研究[D]. 南京: 南京师范大学, 2021.
- [3] 赵长伟, 骈睿珺, 杜天硕, 等. 基于LSTM的重要用户电能质量趋势预测分析模型[J]. 电力系统及其自动化学报, 2022, 34(7): 26–33. ZHAO Changwei, JIAO Ruijun, DU Tianshuo, et al. Trend prediction and analysis model for power quality of important users based on LSTM [J]. Proceedings of the CSU-EPSA, 2022, 34(7): 26–33.
- [4] 周剑桥, 张建文, 席东民, 等. 基于SOP有功-无功协同的低压配电网末端电压越限治理[J]. 电力系统自动化, 2023, 47(6): 110–122. ZHOU Jianqiao, ZHANG Jianwen, XI Dongmin, et al. Terminal voltage overlimit mitigation for low-voltage distribution network based on coordinated active and reactive power control of soft open point [J]. Automation of Electric Power Systems, 2023, 47(6): 110–122.
- [5] 林顺富, 汤继开, 汤波, 等. 典型电能质量稳态指标预测模型研究[J]. 电网技术, 2018, 42(2): 614–620. LIN Shunfu, TANG Jikai, TANG Bo, et al. Study on forecasting model of typical power quality steady state indices [J]. Power System Technology, 2018, 42(2): 614–620.
- [6] 钟庆, 梁家豪, 王钢, 等. 基于趋势跨度指数的稳态电能质量趋势识别及预警方法[J]. 电网技术, 2023, 47(5): 2139–2148. ZHONG Qing, LIANG Jiahao, WANG Gang, et al. Trend identification and early warning method for steady-state power quality based on trend-span index [J]. Power System Technology, 2023, 47(5): 2139–2148.
- [7] CAO Y, QIANG L I, TAN Y, et al. A comprehensive review of energy internet: basic concept, operation and planning methods, and research prospects [J]. Journal of Modern Power Systems and Clean Energy, 2018, 6(3): 399–411.
- [8] 苗新, 张东霞, 孙德栋. 在配电网中应用大数据的机遇与挑战[J]. 电网技术, 2015, 39(11): 3122–3127. MIAO Xin, ZHANG Dongxia, SUN Dedong. The opportunity and challenge of big data's application in power distribution networks [J]. Power System Technology, 2015, 39(11): 3122–3127.
- [9] SALARVAND A, MOALLEM M, MIRZAEIAN B. Obtaining a quantitative index for power quality evaluation in competitive electricity market [J]. IET Generation, Transmission & Distribution, 2010, 4(7): 810–823.
- [10] 丁泽俊, 刘平, 欧阳森, 等. 电能质量预测与预警机制及其应用[J]. 电力系统及其自动化学报, 2015, 27(10): 87–92. DING Zejun, LIU Ping, OUYANG Sen, et al. Mechanism of power quality forecast and early warning and their application [J]. Proceedings of the CSU-EPSA, 2015, 27(10): 87–92.
- [11] 张素香, 赵丙镇, 王风雨, 等. 海量数据下的电力负荷短期预测[J]. 中国电机工程学报, 2015, 35(1): 37–42. ZHANG Suxiang, ZHAO Bingzhen, WANG Fengyu, et al. Short-term power load forecasting based on big data [J]. Proceedings of the CSEE, 2015, 35(1): 37–42.
- [12] 张未, 余成波, 王士彬, 等. 基于VMD-LSTM-LightGBM的多特征短期电力负荷预测[J]. 南方电网技术, 2023, 17(2): 74–81. ZHANG Wei, YU Chengbo, WANG Shibin, et al. Multi-featured short-term power load forecasting based on VMD-LSTM-LightGBM [J]. Southern Power System Technology, 2023, 17(2): 74–81.
- [13] 张沛, 朱驻军, 刘翌, 等. 基于RReliefF-BP网络的新型电力

- 系统静态电压稳定在线评估方法[J]. 南方电网技术, 2023, 17(3): 65–74.
- ZHANG Pei, ZHU, Zhujun, LIU Zhao, et al. Online assessment method for static voltage stability of new power system based on RReliefF-BP network[J]. Southern Power System Technology, 2023, 17(3): 65–74.
- [14] 王知芳, 杨秀, 潘爱强, 等. 基于改进集成聚类和BP神经网络的电压偏差预测[J]. 电工电能新技术, 2018, 37(5): 73–80.
- WANG Zhifang, YANG Xiu, PAN Aiqiang, et al. Voltage deviation forecasting based on improved ensemble clustering and BP neural network[J]. Advanced Technology of Electrical Engineering and Energy, 2018, 37(5): 73–80.
- [15] 李孟特, 顾春华, 温蜜, 等. 一种基于ARMA-BP组合模型的电压偏差预测方法[J]. 智慧电力, 2020, 48(12): 14–19.
- LI Mengte, GU Chunhua, WEN Mi, et al. A voltage deviation prediction method based on ARMA-BP combined model[J]. Smart Power, 2020, 48(12): 14–19.
- [16] 苏卫卫. 基于数据挖掘技术的稳态电能质量数据预测预警研究[D]. 北京: 华北电力大学, 2014.
- [17] 杨秀, 李安, 孙改平, 等. 基于改进GMM-CNN-GRU混合的非侵入式负荷监测方法研究[J]. 电力系统保护与控制, 2022, 50(14): 65–75.
- YANG Xiu, LI An, SUN Gaiping, et al. Non-invasive load monitoring based on an improved GMM-CNN-GRU combination[J]. Power System Protection and Control, 2022, 50(14): 65–75.
- [18] 张美霞, 李丽, 杨秀, 等. 基于高斯混合模型聚类和多维尺度分析的负荷分类方法[J]. 电网技术, 2020, 44(11): 4283–4296.
- ZHANG Meixia, LI Li, YANG Xiu, et al. A load classification method based on gaussian mixture model clustering and multi-dimensional scaling analysis[J]. Power System Technology, 2020, 44(11): 4283–4296.
- [19] FANAEE-T H, THORESEN M. Performance evaluation of methods for integrative dimension reduction [J]. Information Sciences: An International Journal, 2019(493): 105–119.
- [20] 李秉晨, 于惠钧, 刘靖宇. 基于Kmeans和CEEMD-PE-LSTM的短期光伏发电功率预测[J]. 水电能源科学, 2021, 39(4): 204–208.
- LI Bingchen, YU Huijun, LIU Jingyu. Prediction of short-term photovoltaic power generation based on Kmeans and CEEMD-PE-LSTM[J]. Water Resources and Power, 2021, 39(4): 204–208.
- [21] 崔杨, 朱晗, 王议坚, 等. 基于CNN-SAEDN-Res的短期电力负荷预测方法[J]. 电力自动化设备, 2024, 44(4): 164–170.
- CUI Yang, ZHU Han, WANG Yijian, et al. Short-term power load forecasting method based on CNN-SAEDN-Res [J]. Electric Power Automation Equipment, 2024, 44(4): 164–170.
- [22] 李超然, 潘鹏程, 杨伟荣, 等. 基于改进相似日优化HBA-BiLSTM-KELM的光伏发电功率预测[J]. 太阳能学报, 2024, 45(5): 508–516.
- LI Chaoran, PAN Pengcheng, YANG Weirong, et al. Research on PV system power prediction based on improved similar day and HBA-BiLSTM-KELM neural network [J]. Acta Energaie Solaris Sinica, 2024, 45(5): 508–516.
- [23] 李滨, 蒙旭光, 白晓清. 基于BiLSTM循环神经网络的风储系统控制策略[J]. 电力系统及其自动化学报, 2023, 35(12): 20–28.
- LI Bin, MENG Xuguang, BAI Xiaoqing. Control strategy of wind storage based on BiLSTM recurrent neural network [J]. Proceedings of the CSU-EPSC, 2023, 35(12): 20–28.
- [24] 庞传军, 张波, 余建明. 基于LSTM循环神经网络的短期电力负荷预测[J]. 电力工程技术, 2021, 40(1): 175–180.
- PANG Chuanjun, ZHANG Bo, YU Jianming. Short-term power load forecasting based on LSTM recurrent neural network [J]. Electric Power Engineering Technology, 2021, 40(1): 175–180.
- [25] 乔石, 王磊, 张鹏超, 等. 基于模态分解及注意力机制长时间网络的短期负荷预测[J]. 电网技术, 2022, 46(10): 3940–3951.
- QIAO Shi, WANG Lei, ZHANG Pengchao, et al. Short-term load forecasting by long- and short-term temporal networks with attention based on modal decomposition [J]. Power System Technology, 2022, 46(10): 3940–3951.
- [26] 吴晓刚, 阎洁, 葛畅, 等. 基于改进GRU-CNN的风光水一体化超短期功率预测方法[J]. 中国电力, 2023, 56(9): 178–186, 205.
- WU Xiaogang, YAN Jie, GE Chang, et al. Ultra-short-term power forecasting method for wind-solar-hydro integration based on improved GRU-CNN [J]. Electric Power, 2023, 56(9): 178–186, 205.

收稿日期: 2023-07-10; 网络首发日期: 2024-05-23

作者简介:

郭少东(1997), 男, 硕士研究生, 研究方向为基于大数据的台区电压, 1792205336@qq.com;

赵晓莉(1979), 女, 通信作者, 讲师, 硕士, 研究方向为电力系统规划与稳定控制、新能源接入电力系统的运行分析, zhaoxiaoli@shiep.edu.cn;

孙改平(1984), 女, 副教授, 博士, 研究方向为电力系统调度优化, sunfrog2002@163.com。