

基于数据驱动的空间异常电价信号分区域识别方法

白云霄¹, 张云勇¹, 徐云¹, 朱芝润²

(1. 广东电力交易中心有限责任公司, 广州 510180; 2. 输变电装备技术全国重点实验室(重庆大学), 重庆 400044)

摘要: 电力市场节点电价受市场供需、线路检修、发电商报价等多方面因素影响, 易出现节点价差异常、部分区域电价信号偏离大等异常空间分布特征。当前开展现货市场运行的交易中心需要逐日分析电价的空间特征异常, 维持电力市场的安全稳定运行。电价的空间关联特性致使其出现明显分区特征, 在电价分区内部进行异常识别可以有效区分出电价的区域异常特性, 从而提升识别精度。为此提出了将电价信号分区进而辨识空间异常特征的技术思路。首先, 对电价数据进行降维处理, 提取电价信号关键分区特征, 实现了电价信号的聚类分区。进而针对待测电价信号数据的疏密特征, 运用孤立森林算法实现了空间异常电价的分层识别, 可有效辨识电价异常点和电价异常分区。最后, 基于 2022 年某省级电网交易中心实际电价数据对不同的空间异常电价识别方法的识别结果进行对比, 所提方法有效性得到了验证。

关键词: 电力市场; 空间异常电价; 异常识别; 电价分区; 数据驱动

Partition Identification Method of Spatial Abnormal Electricity Price Based on Data-Driven Technology

BAI Yunxiao¹, ZHANG Yunyong¹, XU Yun¹, ZHU Zhirun²

(1. Guangdong Power Exchange Center Co., Ltd., Guangzhou 510180, China; 2. State Key Laboratory of Power Transmission Equipment Technology, Chongqing University, Chongqing 400044, China)

Abstract: The electricity price is influenced by various factors such as market supply and demand, line maintenance, and power generation company quotations, which can easily lead to abnormal spatial distribution characteristics such as abnormal node price differences and large deviation in some regions. The trading centers currently operating in the spot market need to analyze the spatial characteristics on a daily basis, in order to maintain the safe and stable operation of the electricity market. The spatial correlation of electricity prices leads to obvious regional division. Abnormal identification within the electricity price region can effectively distinguish the regional abnormal characteristics, thereby improving identification accuracy. Therefore, this paper proposes a technical approach to partition electricity price signals and identify spatial abnormal characteristics. Firstly, this paper performs dimensionality reduction processing on electricity price data to extract key partitioning characteristics, and achieves clustering partitioning of electricity price signals. Furthermore, in response to the density characteristics of the electricity price signal data to be tested, this paper uses the isolated forest algorithm to achieve hierarchical recognition of spatial abnormal electricity prices, which can effectively identify electricity price abnormal points and regions. Finally, the effectiveness of the data-driven spatial abnormal electricity price partition identification method proposed in this paper is verified based on the actual electricity price data of a provincial power grid trading center in 2022.

Key words: power market; spatial abnormal electricity price; abnormal recognition; electricity price district; data-driven

基金项目: 国家自然科学基金资助项目(52177072); 广东电力交易中心科技项目(GDKJXM20222591)。

Foundation item: Supported by the National Natural Science Foundation of China (52177072); the Science and Technology Project of Guangdong Power Exchange Center Co., Ltd. (GDKJXM20222591) .

0 引言

中央全面深化改革委员会第二十二次会议审议通过了《关于加快建设全国统一电力市场体系的指导意见》^[1], 提出要完善统一电力市场体系的功能, 积极稳妥推进电力现货市场建设, 引导现货市场更好发现电力实时价格。同时要加强电力统筹规划和科学监管, 完善现代电力市场监管体制, 提升对电力市场的科学监管能力, 加强监测预警。电价信号是电力商品属性的直接体现, 能够准确反映电力的市场真实价值^[2], 科学、合理、准确的电价可为电力市场参与者提供良好的经济信号^[3], 提高资源配置效率^[4], 维持系统安全与经济运行^[5]。然而, 由于电价受市场供需、线路检修、发电商报价等多方面因素影响, 极易出现异常电价, 包括均值异常(电价均值超过合理阈值)、时间异常(节点电价前后时段出现剧烈变化)、空间异常(部分节点电价偏离周围电价形成明显区别)等等^[6]。异常的价格信号会为利益相关者带来了巨大的风险, 影响用户经济效益以及电力市场建设的公信力^[7]。一方面, 异常电价的形成必定表明在出清过程中存在异常行为或者存在异常现象, 而其中例如发电商异常报价、发电机组异常状态等异常现象会危害电力系统正常运行, 进而损害参与者利益, 比如 2000 年加州电力危机^[8]; 另一方面, 电价的突变会增大电力市场的风险性, 阻碍市场主体参与市场的积极性^[9]。对异常电价的识别可以从海量电价数据中选择需要重点关注的对象, 以进行后续进一步详细的分析, 找出异常电价形成原因, 再针对异常电价成因, 对电力现货市场出清环节进行针对性整改。因此, 电力市场异常电价的准确识别对于电力市场的安全稳定有着至关重要的作用^[10]。

目前, 国内外已经对异常电价识别做了大量研究。文献[11-12]通过支持向量机方法, 在电价数据已知的基础上, 预测电价价格钉(price spike)这一异常信号的出现区间及大小。其中, 文献[12]融合多种异常电价识别机制对多种不同类型的异常电价进行识别, 例如: 利用 DBSCAN 算法识别出清电价中价格钉或类价格钉的异常电价^[13]; 利用箱型图分析方法识别一段时间内的电价不合理波动^[14]; 利用 Box-Cox 变换法对一段时间的电价平均异常进行识别^[15]。文献[16]则在异常电价的识别过

程中考虑了负荷、温度等其他维度数据信息来进行统一判断。上述方法识别的对象都是电价时间异常, 本质上是同一节点电价信号 t 时刻和 $t-1$ 时刻的纵向对比, 而忽略了电价信号在空间上的关联特性, 无法识别空间异常电价。通过某省级电网交易中心的实际节点电价数据可以发现, 除了上述的电价尖峰、均值异常、波动异常等异常电价类型以外, 电价偏离异常(单个节点的电价与其周围节点的相差过大)、电价区域异常(部分区域电价与系统整体电价水平差异过大)等空间异常情况也时有发生, 但针对该类异常的识别方法, 现有研究鲜有涉及。

空间异常电价信号本质上仍是一类离群异常信号, 电力和工业领域其他方向的异常检测方法具有借鉴意义。文献[17]基于改进的 K-means 算法对风电机组异常数据进行监测。文献[18]设计了基于 DBSCAN 算法的热能发电数据异常监测模型。文献[19]提出了基于 BIRCH 算法的高维传感器数据异常监测方法。上述研究均采用聚类的方法来对异常值进行监测, 其识别效果对超参数的选择具有较强的依赖性, 而系统节点电价数据的变化范围较广, 固定的超参数设置无法适应每个时段的识别需求。另外一类方法是通过分类算法来对异常进行识别。文献[20]采用基于粒子群算法的改进支持向量机方法对异常信号进行检测。文献[21]采用决策树模型对异常电费数据进行识别。文献[22]研究了孤立森林算法在发电机组异常监测中的应用。由节点电价理论可知, 各节点电价受到系统网络参数、线路约束等的影响, 会形成明显的电价分区^[23]。在网络拓扑信息未知的情况下, 如若采用上述方法对系统节点电价整体进行异常信号识别, 将无法判断出电价分区内部的细微异常。

为了弥补现有异常电价识别方法在检测空间异常电价信号方面的不足, 本文提出了基于数据驱动的空间异常电价信号分区域识别方法, 在仅有历史电价数据的情况下, 实现对空间异常电价信号的准确识别。首先, 基于历史电价数据, 运用降维和聚类技术对节点电价进行分区, 从而在后续电价异常信号分区域识别的过程中, 忽略各电价分区之间主要联络线的影响, 发现更加细微的异常信号。进一步, 基于电价数据的分布疏离特征, 运用孤立森林算法对电价分区内部异常信号以及异常区域进行识

别,从节点和区域两个维度判断电价是否存在异常。最后,通过分析对比不同异常信号检测方法,证明本文所提基于数据驱动的空间异常电价信号分区域识别方法的有效性。

1 基于数据驱动的空间异常电价信号分区域识别方法概述

1.1 异常电价信号特征

在一个统一出清的电力市场中,如果没有出现极端阻塞、发电商异常报价、线路检修等异常情况,理想的价格信号应该是所有节点的电价在合理的阈值范围内相似并平稳变化。而异常电价可以根据其与正常电价的差异被分为三类,即均值异常(违反阈值要求)、时间异常(违反平稳变化要求)和空间异常(违反电价相似要求)。

1) 均值异常:某一时段电价的平均值超过了阈值或达到规定的异常值。 t 时段的电价均值 $\bar{\lambda}_t$ 可由式(1)计算得到。

$$\bar{\lambda}_t = \frac{\sum_{i=1}^{N_{\text{node}}} \lambda_{i,t}}{N_{\text{node}}} \quad (1)$$

式中: $\lambda_{i,t}$ 为第 i 个节点在 t 时段的节点电价; N_{node} 为系统内的节点数量。

这一类型的异常表明整体电价水平相较于正常情况出现了异常,当 $\bar{\lambda}_t$ 小于零时为负电价,当 $\bar{\lambda}_t$ 等于零时为零电价,当 $\bar{\lambda}_t$ 超过设定的阈值时,为高电价。

2) 时间异常:某一节点前后时段的电价出现剧烈变化。这一类型的异常电价是电力市场中最为普遍的异常类型,可以通过电价变化率来进行判断,第 i 个节点在 t 时段相较于前一时段电价的变化率 $\eta_{i,t}$ 可以通过式(2)计算得到。

$$\eta_{i,t} = \frac{\lambda_{i,t} - \lambda_{i,t-1}}{\lambda_{i,t-1}} \times 100\% \quad (2)$$

这一类型的异常主要关注的是电价信号的时序特征,当 $\eta_{i,t}$ 为正且超过阈值时,为电价陡升;当 $\eta_{i,t}$ 为负且超过阈值时,为电价陡降;当 $\eta_{i,t}$ 与 $\eta_{i,t+1}$ 均超过阈值且符号相反时,为价格钉;当一段时间内 $\eta_{i,t}$ 频繁超过阈值且符号频繁变化时,为剧烈波动。由引言内容可知作为典型的时序信号,针对其异常信号的识别,目前国内外已经进行了大量研究。

3) 空间异常:某一节点或区域的电价与周围节点和区域形成明显区别。可以通过不同节点之间的电价差值判断电价的偏离程度。 t 时段第 i 个节点和第 j 个节点之间的电价差值 $\Delta\lambda_{i-j}^t$ 可以由式(3)计算得到。

$$\Delta\lambda_{i-j}^t = \lambda_{i,t} - \lambda_{j,t} \quad (3)$$

这一类型异常电价主要关注电价的空间特征,电价的排序对异常的识别没有影响。当节点 i 与周围节点的 $\Delta\lambda_{i-j}^t$ 均超过阈值时, t 时段第 i 个节点的电价为电价偏离异常;当存在多个节点电价偏离异常,且彼此之间的 $\Delta\lambda_{i-j}^t$ 相似时,则这些相似节点形成电价区域异常。这一类异常电价信号的识别方法,目前鲜有研究。

1.2 空间异常电价信号识别方法理论基础

在空间异常电价的识别过程中,由于缺乏完整的系统拓扑信息,只能通过电价分区一定程度上揭示系统的拓扑信息用于电价空间异常检测。

根据节点电价理论可知节点电价的计算公式如式(4)所示。

$$\lambda_t = \mathbf{1}\mu_t + \mathbf{T}^T(\underline{\sigma}_t - \bar{\sigma}_t) \quad (4)$$

式中: λ_t 为 t 时段所有节点的电价向量; μ_t 为 t 时段节点电价的电能量分量; $\bar{\sigma}_t$ 和 $\underline{\sigma}_t$ 分别为支路潮流上下限约束的对偶乘子向量,只有当支路潮流约束为起作用约束时, $\bar{\sigma}_t$ 和 $\underline{\sigma}_t$ 中对应的分量才会呈现为非零量,否则全部为0; $\mathbf{T}$ 为功率转移分布因子矩阵,其作用是构建节点和支路之间功率分配关系,同样也可以用来将线路阻塞引起的成本分摊至各节点; $\mathbf{1}$ 为分量均为1的列向量。

单个节点的电价表达式如式(5)所示。

$$\lambda_{i,t} = \mu_t + \mathbf{T}_i^T(\underline{\sigma}_t - \bar{\sigma}_t) \quad (5)$$

式中 $\mathbf{T}_i$ 为第 i 个节点所对应的功率转移分布因子向量,即功率转移分布因子矩阵 $\mathbf{T}$ 中第 i 列向量。

功率转移分布因子矩阵 $\mathbf{T}$ 是由支路节点导纳矩阵 $\mathbf{B}^{\text{line}}$ 和节点导纳矩阵 $\mathbf{B}^{\text{x}}$ 经过式(6)计算得到的。

$$\mathbf{T} = \mathbf{B}^{\text{line}}(\mathbf{B}^{\text{x}})^{-1} \quad (6)$$

支路节点导纳矩阵 $\mathbf{B}^{\text{line}}$ 中的任意一个元素 $B_{l,n}^{\text{line}}$ 表示的第 l 条与第 n 个节点的关系,如果第 l 条与第 n 个节点相连,则 $B_{l,n}^{\text{line}}$ 的绝对值为第 l 条支路导纳,其符号由支路的规定潮流方向确定,如果第 l 条与第 n 个节点不相连,则 $B_{l,n}^{\text{line}}$ 为0。由此可见, $\mathbf{B}^{\text{line}}$ 是一个稀疏矩阵。而节点导纳矩阵 $\mathbf{B}^{\text{x}}$ 中的任意一个

元素 $B_{i,j}^x$ 表示第 i 个节点与第 j 个节点之间的关系, 如果第 i 个节点与第 j 个节点相连, 则 $B_{i,j}^x$ 的值为第 i 个节点与第 j 个节点之间的线路导纳的负数, 否则为 0。特别地, 当 $i=j$ 时, $B_{i,j}^x$ 的值为所有与第 i 个节点相连的支路导纳之和。

根据支路节点导纳矩阵 B^{line} 和节点导纳矩阵 B^x 的分量特征可以发现在拓扑结构上相对集中、地位相似的节点, 其对应的功率转移分布因子向量 T_i 越相似。而在同一出清时段, 所有节点电价电能量分量和支路潮流上下限约束的对偶乘子向量相等, 节点电价之间的相似程度仅由功率转移分布向量决定。因此可以通过对相似节点电价的聚类, 一定程度上揭示系统的拓扑信息。

DBSCAN 算法正是基于数据的密度进行聚类, 能够将数据相似且分布集中的电价进行聚合, 这符合“通过节点电价相似性揭示系统拓扑信息”的要求, 同时 DBSCAN 算法无需事先指定聚类数量, 这也符合电价分区的实际要求。鉴于以上原因, 本文选择 DBSCAN 算法对电价进行分区聚类。

而在空间异常电价的识别过程中, 孤立森林算法是直接刻画数据之间的疏离程度, 数据之间的排序对监测结果没有影响, 算法简单高效, 在工业中有着广泛应用, 也十分契合空间电价异常识别过程中的节点电价不存在序列特征这一特点。故选择孤立森林算法来进行空间异常电价的识别。

1.3 空间异常电价信号识别方法概述

在电力系统拓扑信息缺失的情况下, 仅能通过电价历史数据来对电价进行分区, 需要遵循以下标准: 1) 所有节点都尽可能被划分为某一电价分区; 2) 在同一时刻, 分区内部所有节点的电价尽可能相似, 分布尽可能集中; 3) 电价分区数量不宜过多。在这一过程, 需要使用到长时间、高维度的历史电价数据, 如果直接进行分区则会影响分区效果, 所以需要原始数据进行降维, 再对降维后的数据执行分区操作。其后, 基于电价数据的疏离程度和分布特征在各电价分区内部识别电价异常点, 在各分区之间识别异常电价分区, 从而实现对空间异常电价信号的准确识别。

最终本文所提的方法的流程图如图 1 所示。主要分为两个步骤, 一是对电价历史数据进行降维和分区, 二是基于分区信息对待测数据进行异常电价识别, 各步骤的具体内容将在后续章节中进行详细

说明。

2 基于降维与聚类的电价分区算法介绍

面对市场中众多的节点, 提取每个节点一年每天 96 时段共 35 040 时段的电价数据作为聚类数据集。由于该数据集维数过多, 直接聚类效果不尽人意, 首先对特征数据集采用 t-SNE 进行高维数据降维。完成降维后, 通过基于密度的 DBSCAN 分类方法进行稀疏区域的划分, 得到待测电力市场的电价分区。

2.1 基于 t-SNE 算法的电价数据降维

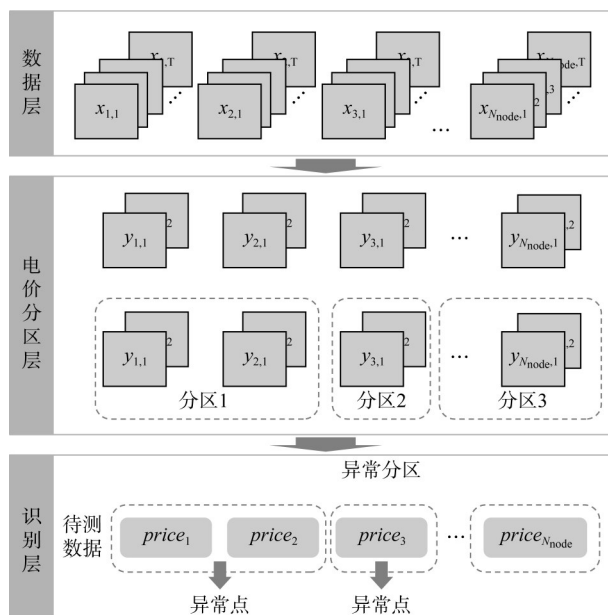


图 1 基于数据驱动的空间异常电价信号分区域识别方法流程图

Fig. 1 Flowchart of portion identification method of spatial abnormal electricity prices based on data-driven technology

t-SNE 是一种非线性无监督降维算法。其主要思想是通过仿射变换将高维数据点映射到概率分布上, 再在低维空间中构建新的概率分布, 不断优化低维空间数据, 使得高、低维数据空间的两个概率分布尽可能保持一致。相对于其他非线性降维方法, t-SNE 算法能够更好地保留数据的局部特征, 在降维后的低维空间中, 相似的数据点会聚集在一起, 形成明显的类簇, 以便后续更好地对电价进行分区。

在利用 t-SNE 算法进行电价数据降维时, 首先将系统各节点电价数据组成高维原始数据集 $X =$

$\{x_1, x_2, \dots, x_{N_{\text{node}}}\}$, N_{node} 为系统内节点数量, 任意一个节点的电价数据 $x_n = (x_{n,1}, x_{n,2}, \dots, x_{n,T})$, T 为数据集的维度, 即电价的时段数量。 x_i 和 x_j 为数据集 $\mathbf{X}$ 中的任意两个数据点, 以 x_i 为中心构建方差为 σ_i 的高斯分布, 则可以得到如式(7)所示的 x_j 对 x_i 的相似条件概率 p_{ji} , 其表示 x_j 在 x_i 领域内的概率大小, 其值越大, 表明 x_j 离 x_i 越近。

$$p_{ji} = \frac{\exp(-\|x_j - x_i\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_k - x_i\|^2 / 2\sigma_i^2)} \quad (7)$$

式中: x_k 为数据集 $\mathbf{X}$ 中的任意一个数据点, $k \neq i$; σ_i 为以 x_i 为中心的高斯方差, 其决定了以 x_i 为中心构建的高斯分布的形状。

为了保证条件概率的对称性以简化后续梯度计算, 加快优化过程, 采用更加通用的联合概率代替原始的条件概率, 如式(8)所示。

$$p_{ij} = p_{ji} = \frac{p_{ij} + p_{ji}}{2N_{\text{node}}} \quad (8)$$

式中 p_{ij} 和 p_{ji} 均为 x_i 和 x_j 之间的联合概率。

由式(7)—(8)可知, 若要获得准确的联合概率 p_{ij} , 需要确定高斯方差 σ_i 和 σ_j 。t-SNE算法中使用困惑度(Perplexity)的概念来寻找合适的高斯方差。

以数据点 x_i 的高斯方差 σ_i 为例, 数据点 x_i 的困惑度 Perplexity 可以解释为 x_i 附近的有效近邻点个数, 其可以表征为关于 $H(P_i)$ 的指数函数, 如式(9)所示。

$$P_{\text{erplexity}}(P_i) = 2^{H(P_i)} \quad (9)$$

式中: P_i 为 x_i 与其他所有数据点之间的联合概率 p_{ji} 构成的联合概率分布; $H(P_i)$ 为联合概率分布 P_i 的熵, 其计算方法如式(10)所示。

$$H(P_i) = - \sum_{j=1}^{N_{\text{node}}} p_{ji} \log_2(p_{ji}) \quad (10)$$

由式(9)—(10)可知, 通过调节概率分布的困惑度 Perplexity(P_i) 大小可以进而调整对应的高斯方差 σ_i 。为了保证 t-SNE 算法保持较高的鲁棒性, 困惑度大小通常选择 5~50 之间。确定分布的困惑度以后, 再利用二分搜索的方式来寻找数据点 x_i 所对应的最佳高斯方差 σ_i 。

完成原始高维数据空间的联合概率分布计算以后, 同样需要对降维后目标数据集的联合概率分布进行假设。假设 $\mathbf{Y} = \{y_1, y_2, \dots, y_{N_{\text{node}}}\}$ 为降维后的

目标数据集, 其中任意一个节点电价的低维映射点 $y_n = (y_{n,1}, y_{n,2})$ 。 y_i 和 y_j 为数据集 $\mathbf{Y}$ 中的任意两个数据点。

为了解决数据降维过程中出现的拥挤问题(crowding problem)^[24], 可以利用自由度为 1 的 t 分布^[25]来表征低维数据空间的概率分布。低维数据空间中数据点的联合概率 q_{ij} 如式(11)所示。

$$q_{ij} = q_{ji} = \frac{(1 + \|y_j - y_i\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_k - y_i\|^2)^{-1}} \quad (11)$$

式中: q_{ij} 和 q_{ji} 均为 y_i 和 y_j 之间的联合概率; y_k 为数据集 $\mathbf{Y}$ 中的任意两个数据点, $k \neq i$; y_i 与其他所有数据点之间的联合概率 q_{ji} 构成的联合概率分布 Q_i 。

理想情况下, 若降维前后数据的联合概率分布 P_i 和 Q_i 均相等, 则表明了低维数据集 $\mathbf{Y}$ 准确反映了高维数据集 $\mathbf{X}$ 中各数据点之间的相似关系。t-SNE 算法通过不断地迭代计算, 调整低维数据集 $\mathbf{Y}$, 来达到联合概率分布 P_i 和 Q_i 均相等的目的。这一过程中通过 KL 散度(Kullback-Leibler divergence)来衡量两个联合概率分布之间差异性, 如式(12)所示。

$$C = \sum_{i=1}^{N_{\text{node}}} KL(P_i \| Q_i) = \sum_{i=1}^{N_{\text{node}}} \sum_{j=1}^{N_{\text{node}}} p_{ij} \ln(p_{ij}/q_{ij}) \quad (12)$$

式中: C 为迭代优化的目标函数, 是数据集内所有数据点概率分布的 KL 散度之和; $KL(\cdot)$ 为计算两分布之间的 KL 散度。

最小化目标函数 C , 保证降维前后数据的联合概率分布 P_i 和 Q_i 尽可能相似, 采用梯度下降法进行优化, 目标函数 C 关于降维后任意数据点 y_j 的梯度如式(13)所示, 迭代求解计算公式如式(14)所示。

$$\frac{\partial C}{\partial y_i} = 4 \sum_{j=1}^{N_{\text{node}}} (p_{ij} - q_{ij})(y_i - y_j) \left(1 + \|y_i - y_j\|^2\right)^{-1} \quad (13)$$

$$\mathbf{Y}^{(t+1)} = \mathbf{Y}^{(t)} - \gamma \frac{\partial C}{\partial \mathbf{Y}^{(t)}} \quad (14)$$

式中: $\mathbf{Y}^{(t)}$ 和 $\mathbf{Y}^{(t+1)}$ 分别为第 t 次和第 $t+1$ 次的迭代得到的结果; γ 为梯度下降的学习率; $\partial C / \partial \mathbf{Y}^{(t)}$ 为目标函数关于第 t 次迭代结果的梯度, 可由式(13)求得。

2.2 基于 DBSCAN 算法的降维后节点电价分区

DBSCAN 算法基于数据的密度进行聚类, 可以

实现不规则空间的聚类,同时无需事先指定聚类数量,符合电价分区的要求。数据密度则是通过规定范围内的数据点个数确定的,2.1节中得到低维电价数据集 $\mathbf{Y}$ 中任意两点之间的距离可以通过式(15)确定。

$$d_{\text{ist}}(y_j, y_i) = \sqrt{(y_{j,1} - y_{i,1})^2 + (y_{j,2} - y_{i,2})^2} \quad (15)$$

式中 $d_{\text{ist}}(y_j, y_i)$ 为数据点 y_i 和 y_j 之间的距离。

在设定的距离参数 ε 下,数据集 $\mathbf{Y}$ 内距离数据点 y_i 的距离小于 ε 的数据点的集合可以形成一个最小邻域 $N_\varepsilon(y_i)$,如式(16)所示。

$$N_\varepsilon(y_i) = \{y_j | y_j \in \mathbf{Y}, d_{\text{ist}}(y_j, y_i) \leq \varepsilon\} \quad (16)$$

$$i = 1, 2, \dots, N_{\text{node}}$$

基于数据点之间的距离以及不同数据点之间的从属关系,数据集 $\mathbf{Y}$ 中的数据点被分为三类,分别为核心点、边界点和噪声点。

核心点是指该数据点周围半径 ε 的临域范围内包含至少 P_{tmin} 个数据点(包含该核心点本身)。数据集 $\mathbf{Y}$ 内所有核心点可以构成集合 $C_{\text{ore}}(\mathbf{Y})$,如式(17)所示。

$$C_{\text{ore}}(\mathbf{Y}) = \{y_i | y_i \in \mathbf{Y}, S_{\text{ize}}(N_\varepsilon(y_i)) \geq P_{\text{tmin}}\} \quad (17)$$

式中: P_{tmin} 为设定的密度阈值参数,其表示单个核心点所在的数据集合中数据点的最小数量; $S_{\text{ize}}(N_\varepsilon(y_i))$ 为数据集 $N_\varepsilon(y_i)$ 内的数据点的数量。

边界点是指该数据点是非核心点的同时,属于任意一个核心点所在的最小临域。数据集 $\mathbf{Y}$ 内所有边界点可以构成集合 $B_{\text{ord}}(\mathbf{Y})$,如式(18)所示。

$$B_{\text{ord}}(\mathbf{Y}) = \{y_n | y_n \notin C_{\text{ore}}(\mathbf{Y}), y_n \in N_\varepsilon(y_i), \forall y_i \in C_{\text{ore}}(\mathbf{Y})\} \quad (18)$$

噪声点是指该数据点既不是核心点也不是边界点。数据集 $\mathbf{Y}$ 内所有噪声点可以构成集合 $N_{\text{oise}}(\mathbf{Y})$,如式(19)所示。

$$N_{\text{oise}}(\mathbf{Y}) = \{y_m | y_m \in \mathbf{Y}, y_m \notin C_{\text{ore}}(\mathbf{Y}), y_m \notin B_{\text{ord}}(\mathbf{Y})\} \quad (19)$$

利用DBSCAN算法对降维后的电价数据 $\mathbf{Y}$ 进行聚类的步骤如图2所示。

1) 导入降维后的电价数据集 $\mathbf{Y}$,并初始化簇类标志 $C_{\text{ate}}=0$ 。

2) 从数据集 $\mathbf{Y}$ 中任取一数据点 y_i 并判断其是否已经被标记为核心点、边界点或者噪声点,若已

经标记,则重复步骤2),进行下一个数据点的判断;若未被标记,则跳转至步骤3)。

3) 利用式(17)判断数据点 y_i 是否为核心点,若是则标记为核心点,跳转至步骤4);否则标记为噪声点,跳转至步骤2)。

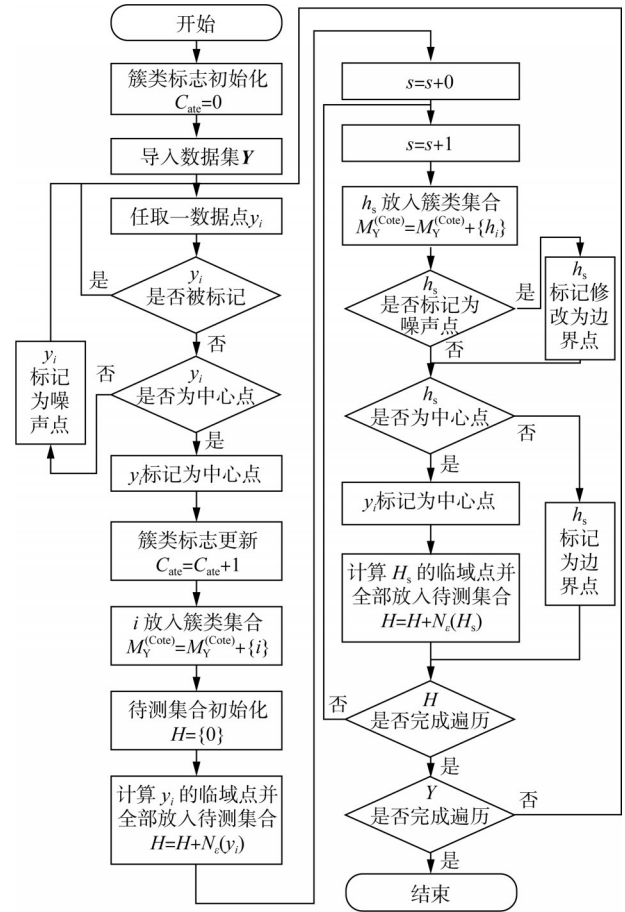


图2 利用DBSCAN算法进行电价聚类的流程图

Fig. 2 Flowchart of electricity price clustering using DBSCAN algorithm

4) 簇类标志 C_{ate} 更新,并将数据点 y_i 放入该簇类集合 $M_Y^{(C_{\text{ate}})}$;初始化待测试集合 H ,通过式(16)计算据点 y_i 半径 ε 范围内所有的临近点,并将其全部加入到待测集合 H 。

5) 按顺序选取待测集合 H 中的数据点 h_s ,并将其归入该簇类集合 $M_Y^{(C_{\text{ate}})}$;判断其是否被标记为噪声点。若 h_s 已经被标记为噪声点,则修改标记为边界点,再继续步骤6);否则直接继续步骤6)。

6) 利用式(17)判断数据点 h_s 是否为核心点,若是则标记为核心点,并利用式(16)计算据点 h_s 半径 ε 范围内所有的临近点,将其全部加入到待测集

合 H ,然后再继续步骤7);否则,依据式(18)标记为边界点,在继续步骤7)。

7) 判断待测集合 H 中所有数据点是否均完成标记,如是继续步骤8);若否则跳转至步骤5)。

8) 判断数据集 Y 中所有数据点是否均完成标记,如是则完成聚类;若否则跳转至步骤2)。

通过上述方法对电价数据进行了聚类,共得到 N_{Cate} 个电价分区,在第3节中,本文将基于电价分区来进行异常节点电价的识别。

3 基于数据分布疏密特征的电价异常识别

在进行空间异常电价的识别过程中,分两步进行,首先对各个电价分区内部的异常点进行识别,剔除异常数据后,得到每个区域内部的正常电价数据点,再依据剩余电价数据的数据特征进行电价分区间的异常识别,检测是否存在异常电价区域。

3.1 电价分区内部异常电价识别流程

基于电价分区信息,利用孤立森林算法对分区后的电价数据异常进行识别,主要分为两个环节,第一个环节是孤立森林(iForest)的训练;第二个环节是用训练后的孤立森林来计算每个数据点的异常分数,并对异常数据点进行识别。

其中,以任意第 C_{ate} 个电价分区为例,第一个环节对孤立森林进行训练的流程步骤如图3所示。

1) 设定超参数: (1) iForest内孤立树(iTree)的数量 N_{Tree} , (2) iForest内单个iTree最大高度 N_{Sub} 。

2) 依据电价分区结果 $M_Y^{(C_{\text{ate}})}$,从待测数据集中导入第 C_{ate} 个电价分区数据 $M_X^{(C_{\text{ate}})}$,并对数据进行预处理,如若数据集平均值小于10,则将数据四舍五入精确到个位,否则按5的倍数取整,以此避免DBSCAN算法使用过程中极微小差别对识别结果的影响。

3) 判断数据集 $M_X^{(C_{\text{ate}})}$ 的数据数量是否大于单个iTree最大高度 N_{Sub} ,并利用式(20)对每个电价分区对应的iTree最大高度 $N_{\text{Sub}}^{(C_{\text{ate}})}$ 进行赋值。

$$N_{\text{Sub}}^{(C_{\text{ate}})} = \begin{cases} N_{\text{Sub}}, & S_{\text{ize}}(M_X^{(C_{\text{ate}})}) \geq N_{\text{Sub}} \\ S_{\text{ize}}(M_X^{(C_{\text{ate}})}), & S_{\text{ize}}(M_X^{(C_{\text{ate}})}) < N_{\text{Sub}} \end{cases} \quad (20)$$

4) iTree标志 q 初始化。

5) iTree标志 q 更新。

6) 清空所有的子空间 $K_{s,k}$,并将iTree的层数标志 s 以及iTree每层子空间的标志 k 初始化。

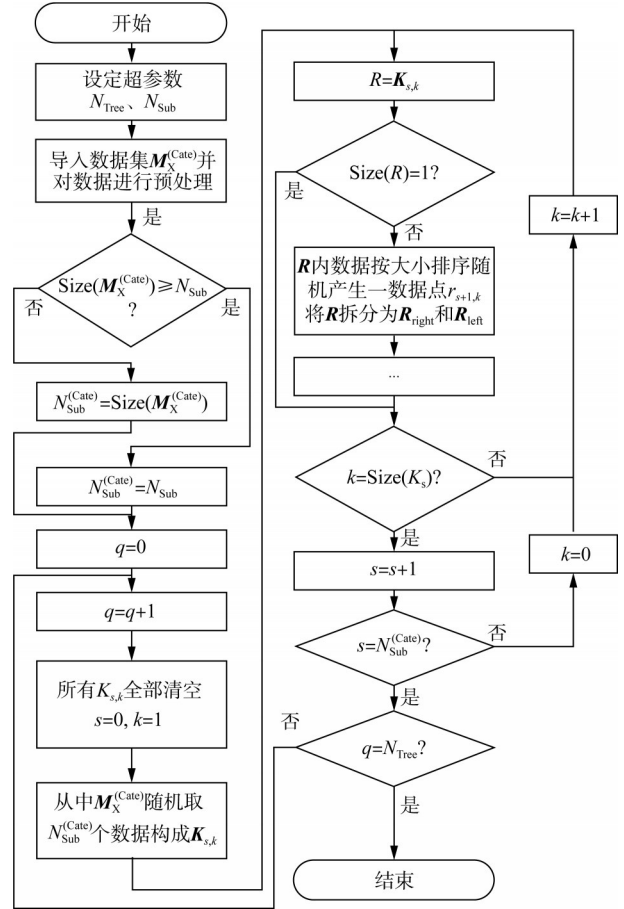


图3 孤立森林树训练流程图

Fig. 3 Flowchart of iForest training

7) 从数据集 $M_X^{(C_{\text{ate}})}$ 中随机抽取 $N_{\text{Sub}}^{(C_{\text{ate}})}$ 个数构成子空间 $K_{s,k}$,该子空间被称为根节点。

8) 将子空间 $K_{s,k}$ 内的数据全部载入待分割区间 R 。

9) 判断待分割区间 R 内是否仅包含一个数据节点,如是则将对应的子空间 $K_{s,k}$ 标记为叶节点并跳转至步骤11),否则继续。

10) 对待分割区间 R 内数据进行排序,随机生成一数据点 $r_{s+1,k}$ 将区间 R 内数据分割为 R_{right} 和 R_{left} 两个集合,将新生成的两个数据推入下一层的子空间集合中,即 $K_{s+1} = \{K_{s+1}, R_{\text{right}}, R_{\text{left}}\}$ 。

11) 判断当前层的子空间是否均完成了分割,若是则跳转至步骤13);否则继续步骤12)。

12) 子空间的标志 k 更新,跳转至步骤8)。

13) 层数标志 s 更新。

14) 判断当前 iTree 是否已达到最大层数, 若是则将目前层的所有子空间 $K_{s,k}$ 均标记为叶节点, 并继续步骤 15); 否则跳转至步骤 12)。

15) 判断是否完成了所有 iTree 的训练, 若是则完成该电价分区 iForest 训练过程, 结束流程; 否则跳转至步骤 5)。

依照上述步骤完成所有电价分区 iForest 的训练以后, 依次对各个电价分区内部的待测数据进行异常信号识别。首先判断第 C_{ate} 个电价分区待测数据 $M_X^{(C_{ate})}$ 极差与平均值的比值是否大于 3%, 若是, 则视该电价分区内所有节点电价正常, 无需进行异常识别; 否则需要利用式 (21) 计算该电价分区内每个节点电价的异常分数, 具体如式 (21) 所示。

$$S(C_{ate}, x_i, N_{sub}^{(C_{ate})}) = 2^{\left(- \sum_{q=1}^{q=N_{tree}} f_q(x_i) / g(N_{sub}^{(C_{ate})}) \right)}, \forall x_i \in M_X^{(C_{ate})} \quad (21)$$

式中: 函数 $f_q(x_i)$ 为计算用第 q 个 iTree 去分割出数据 x_i 所需的路径长度, 如式 (22) 所示。

$$f_q(x_i) = S_{plit x_i}^q + g(S_{ame x_i}^q), q = 1, 2, \dots, N_{tree} \quad (22)$$

式中: $S_{plit x_i}^q$ 为第 q 个 iTree 上从根节点到数据 x_i 所在的叶节点经历的层数; $S_{ame x_i}^q$ 为第 q 个 iTree 上和数据点 x_i 在的同一个叶节点的数据数量, 若数据点 x_i 单独存在于一个叶节点, 则 $S_{ame x_i}^q = 0$; 函数 $g(x)$ 为计算用第 q 个 iTree 去分割出数据 x_i 所需的平均路径长度, 如式 (23) 所示。

$$g(x) = 2(\ln(x-1) + \gamma) - \frac{2(x-1)}{x} \quad (23)$$

式中: γ 为欧拉常数, 其值约等于 0.577 215 664 9。

由式 (21) 可知, 节点电价的异常分数是一个以 2 为底的指数函数, 其指数的取值范围为 $(-\infty, 0)$, 故异常分数 $S(C_{ate}, x_i, N_{sub}^{(C_{ate})})$ 的取值范围为 $(0, 1)$ 。异常分数越接近 0, 则表明将待测数据点从数据集中分离出去所需的路径越长, 该数据点越不可能是异常点, 反之异常分数越接近 1, 则该待测数据点越容易被从数据集中分离, 该数据点越可能是异常点。依据各电价分区内节点电价异常分数的类别数量确定合理的阈值, 完成对每个电价分区内部的异常点识别。

3.2 异常电价分区识别流程

完成电价分区内部异常电价识别后, 利用式

(24) 计算每个电价分区内正常电价的平均值 $\bar{x}_{C_{ate}}$ 。

$$\bar{x}_{C_{ate}} = E(M_X^{*(C_{ate})}), C_{ate} = 1, 2, \dots, N_{C_{ate}} \quad (24)$$

式中: $M_X^{*(C_{ate})}$ 为第 C_{ate} 个电价分区的电价数据集 $M_X^{(C_{ate})}$ 中剔除异常电价数据点后的正常电价数据集; 函数 $E(\cdot)$ 表示求取集合内数据的平均值。

将每个电价分区的正常电价平均值组成一个新的数据集 Z , 如式 (25) 所示。

$$Z = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{C_{ate}}, \dots, \bar{x}_{N_{C_{ate}}}\} \quad (25)$$

将数据集 Z 视为待测数据集, 重复 3.1 节中的步骤, 完成该数据集下 iForest 的训练, 再利用式 (21)~(23) 计算每个电价分区的异常分数, 从而识别电价分区内的电价是否相较于整个系统存在异常。

4 算例分析

本文所用数据来源为某省级电网交易中心 2 000+ 节点 2022 年全年每 15 min 的实际节点电价数据。在本节中, 首先测试了 DBSCAN 算法中不同超参数组合对电价分区效果的影响, 并分析得出最优的参数组合, 得到理想的电价分区结果。其后, 基于电价分区信息, 对比不同的空间异常电价识别方法识别结果, 证明本文所提基于数据驱动的空间异常电价信号分区域识别方法的有效性。

4.1 电价分区参数选择及分区结果

在实际应用过程中可以利用待测日以前的电价数据进行分区从而获得最新、最准确的分区信息, 并基于最新获得的分区信息对待测日的空间异常电价进行识别。基于此实际应用场景, 将 2022 年最后一天的 96 时段节点电价数据设置为测试集, 剩余的电价数据设置为训练集。利用 DBSCAN 算法对训练集的电价数据进行分区。需要注意的是, 虽然 DBSCAN 算法可以避免事先确定分区数量, 但算法执行过程中距离参数 ε 和密度阈值参数 $P_{t, min}$ 的设定, 会影响电价分区结果, 表 1 展示了 16 种参数组合下的电价分区结果, 分别从电价分区个数、未分区节点数量、极差、四分位差和方差五个方面来衡量不同参数组合下的分区效果。

从表 1 中可知, 当距离参数 ε 超过 10 时, 衡量数据分布集中程度的极差、四分位差和方差的数值均明显增大, 此时分区结果不满足第一章中所提出的“分区内部所有节点的电价尽可能相似, 分布尽

表 1 不同参数组合下的电价分区效果

Tab. 1 Effects of electricity price zonings under different parameter combinations

距离参 数 ε	密度阈值 参数 P_{tmin}	电价分 区数量	未分区节 点数量	极差	四分 位差	方差
3	3	43	8	16.87	4.25	4.57
	5	36	36	19.93	4.81	5.26
	10	35	44	20.32	4.94	5.38
	20	28	123	22.90	4.68	5.64
5	3	29	2	27.55	7.74	7.59
	5	25	26	27.70	5.93	6.79
	10	24	36	28.38	6.15	6.98
	22	23	63	28.65	5.50	6.76
10	3	13	2	52.43	9.63	11.02
	5	13	0	52.03	7.39	12.47
	10	12	0	53.08	7.60	9.65
	20	9	4	66.85	9.31	11.83
20	3	5	0	101.51	14.33	16.01
	5	4	0	109.04	11.24	14.26
	10	4	0	114.46	15.17	16.68
	20	2	0	205.51	21.30	24.60

可能集中”原则；而当距离参数 ε 等于 3 或 5 时，分区后的数据集中程度差距不大，但电价分区数量以及未分区节点数量存在明显差异。同时考虑到电价分区应尽可能包含所有节点，同时电价分区不宜过多，导致过度分区，距离参数 ε 设为 5，密度阈值参数 P_{tmin} 设为 3 为宜。在此组超参数设定下，算例测试所用某省级电网交易中心 2022 年节点电价大致分为 29 个电价分区。从电价分区结果上看，整体分布仍是基于行政区域划分，同时在经济发达地区，由于电网建设程度更加完善，分区也更加紧密。部分沿海地区受地理环境影响，会形成相对独立的电价分区。由此可以证明，在电网拓扑信息缺失的情况下，通过历史节点电价数据可以还原部分网络信息，所得的电价分区结果，具备一定的真实性和可信任性，可以用于后续空间异常电价信号的识别。

4.2 不同空间异常电价信号识别方法效果对比

4.2.1 空间电价异常节点识别效果

由于目前尚未有专门识别空间异常电价信号的方法，在本节中对比了本文所提方法与电力工业领域两种常用的异常数据识别方法对空间异电价信号

的识别效果，3 种方法如下所示。

M1：本文所提基于数据驱动的空间异常电价信号分区域识别方法；

M2：利用孤立森林算法进行空间异常电价信号整体识别；

M3：利用 K-means 算法进行空间异常电价信号整体识别。

需要注意的是，为了方便对比不同异常信号的识别效果，M2 中孤立森林算法所有参数设置以及数据预处理流程与 M1 中保持一致，两者区别仅为不再进行分区识别。M3 中 K-means 算法所需聚类的数量设定为 M1 中得到的最优分区数量 29，聚类后到聚类中心距离超过该簇类平均距离 3 倍以上的被视为异常点。

对测试集中某省级电网交易中心 2022 年中最后一天 96 时段的节点电价数据分别运用上述 3 种方法进行识别。以相邻节点之间电价差值相比较得到的空间异常电价识别结果为标准，对比 3 种方法的识别效果。3 种方法识别出的异常电价信号点的比例如表 2 所示，识别准确识别率如图 4 所示，识别错误率如图 5 所示（准确识别率是指识别出的异常信号中正确的数量占人工识别出的异常信号总数的比例，当人工识别异常信号数量为 0，若方法没有检测出异常信号，则此时识别准确率定义为 100%，若此时检测出异常信号，则此时识别准确率定义为 0%；错误识别率是指识别出的异常信号中错误的数量占识别出的异常信号总数的比例）。

表 2 M1—M3 的空间异常电价信号点识别比例

Tab. 2 Identification ratio of spatial abnormal electricity price signal points for M1—M3

方法	异常电价信号点识别比例/%
M1	2.06
M2	6.37
M3	0.93

从表 2 中数据可以发现，M3 方法识别出的空间异常电价信号最少，约为本文所提 M1 方法的一半，而 M2 方法识别出的空间异常电价信号最多，是本文所提方法 M1 的 3 倍以上。结合图 5 中各方法的准确识别率可以发现，相比于 M1 方法 95.09 % 的平均准确识别率和 M2 方法 93.17 % 的平均准确识别率，M3 方法由于识别出的异常信号数

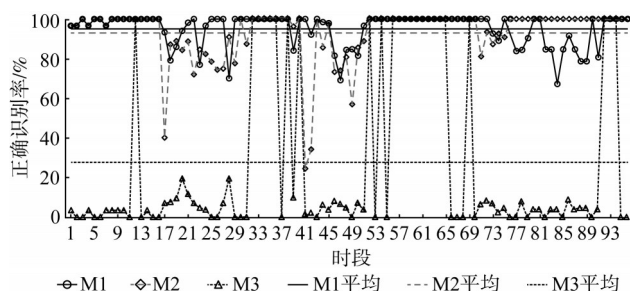


图4 M1—M3的空间异常电价信号准确识别率

Fig. 4 Accurate recognition rates of spatial abnormal electricity price signals for M1—M3

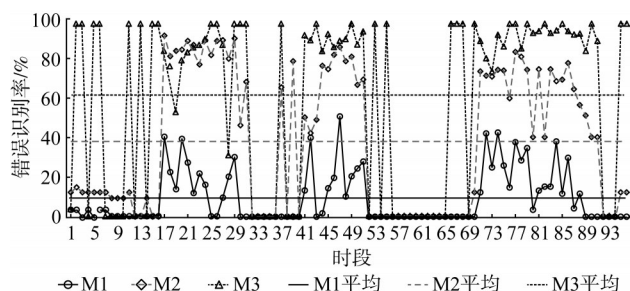


图5 M1—M3的空间异常电价信号错误识别率

Fig. 5 Error recognition rates of spatial abnormal electricity price signals for M1—M3

量过少, 导致准确识别率较低, 平均不到 30 %。同时, 虽然 M1 方法和 M2 方法有着相似的识别准确率, 但相较于 M2, M1 的识别效果更加平稳, 绝大部分时段识别准确率可以维持在 80% 以上, 没有极端情况出现。

再结合图 6 中所示的错误识别率可以发现, M3 具有最高的错误识别率, 平均达到 60 % 以上, 由此可见, K-means 作为专门的聚类算法, 异常识别只是其附属功能, 对于节点电价信号这种单维、非时序数据的异常识别, 效果并不理想。而 M2 方法的错误识别率则比本文所提 M1 方法高出 3 倍多, 达到 38.93 %, 可见 M2 方法较高的准确识别率是依靠大量检出异常信号来保证的, 虽然准确识别率较高, 但同样会错误检测出大量异常信号, 造成误报率过高。通过 3 种方法识别效果的对比, 可以发现孤立森林算法在空间异常电价信号的识别上具备一定的优势, 但在不分区识别的情况下, 会将大量原本在分区内部是正常的电价信号识别为异常, 造成后续异常电价信号成因分析的困难。综上所述, 本文所提方法能够保证空间异常电价信号的识别效果。

4.2.2 空间异常电价分区域识别效果

本文所提 M1 方法除了可以完成对异常电价信号点的识别, 还可以从市场整体角度出发辨别电价信号区域异常, 以便后续进一步分析电价信号异常成因。对 4.2.1 节中所测试的同一天节点电价数据进行区域异常电价检测, 检测结果如图 6 所示。

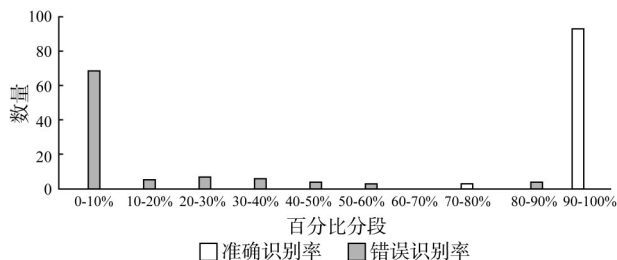


图6 本文所提方法异常电价分区域识别效果

Fig. 6 Effectiveness of M1 in identifying abnormal electricity price zones

由图 6 可知, 本文所提 M1 方法在所测试的所有时段中几乎都能够完美识别所有电价信号区域, 平均准确率达到 99.27 %。同时, 有 60 % 以上的时段错误识别率在 10% 以下, 整体错误识别率也仅为 11.72 %。虽然存在电价信号异常区域过量识别的问题, 但极端错误识别情况占比极少, 且集中区域电价分布较为平均的情况、电价数据疏离特征不明显的时段, 此时需要辅助以人工经验加以判别。

5 结语

电力市场节点电价数据存在广泛的空间相关性, 利用这一特性进行空间异常电价信号识别十分重要。本文所提出的基于数据驱动的空间异常电价信号分区域识别方法, 利用降维和聚类算法充分挖掘历史电价数据中各节点之间的耦合关系, 实现大规模电力市场电价分区。在分区信息的基础之上, 针对电价数据的疏离程度, 分两步对异常电价节点和异常电价区域进行识别。相较于目前常用的整体异常信号识别方法, 能够极大程度上改善识别效果, 弥补了现有异常电价识别方法在空间异常电价信号识别方面的不足, 同时, 也为后续异常电价成因溯源分析提供了信息基础, 对于电力市场的稳定运行具有重要意义。

参考文献

- [1] 国家发展改革委. 国家能源局关于加快建设全国统一电力市

- 场体系的指导意见(发改体改[2022]118号)[EB/OL]. (2022-01-18) [2023-8-22]. https://www.gov.cn/zhengce/zhengceku/2022-01/30/content_5671296.htm.
- [2] 王怡, 杨知方, 余娟, 等. 节点电价与对偶乘子的内在关联分析与扩展[J]. 电力系统自动化, 2021, 45(6): 82-91.
WANG Yi, YANG Zhifang, YU Juan, et al. Analysis and extension of internal relationship between locational marginal price and dual multiplier [J]. Automation of Electric Power Systems, 2021, 45(6): 82-91.
 - [3] 王剑晓, 钟海旺, 夏清, 等. 基于价值公平分配的电力市场竞争机制设计[J]. 电力系统自动化, 2019, 43(2): 7-17.
WANG Xiaojian, ZHONG Haiwang, XIA Qing, et al. Competitive mechanism design in electricity market based on fair benefit allocation [J]. Automation of Electric Power Systems, 2019, 43(2): 7-17.
 - [4] 杨强, 蔡华祥, 严明辉, 等. 激励相容与深度博弈的云南电力市场机制设计[J]. 云南电力技术, 2018, 46(2): 73-80.
YANG Qiang, CAI Huaxiang, YAN Minghui, et al. Design and application of Yunnan electricity market mechanism according to the theory of incentive compatibility and game in depth[J]. Yunnan Electric Power, 2018, 46(2): 73-80.
 - [5] 吴迪, 王正风, 高卫恒, 等. 省级电力现货市场全网安全校核[J]. 中国电力, 2023, 56(9): 57-65.
WU Di, WANG Zhengfeng, GAO Weiheng, et al. Research on power grid security check for provincial electricity spot market[J]. Electric Power, 2023, 56(9): 57-65.
 - [6] 赵筠筠, 张焰. 电力市场环境价格突变的机理分析和风险防范[J]. 华东电力, 2008(8): 6-10.
ZHAO Junjun, ZHANG Yan. Price spike analysis and risk prevention in electricity markets [J]. East China Electric Power, 2008(8): 6-10.
 - [7] 袁磊. 上网电价波动风险分析[D]. 北京: 华北电力大学, 2009.
 - [8] BUSHNELL M. California's electricity crisis: a market apart? [J]. Energy Policy, 2004, 32(9): 1045-1052.
 - [9] 王富阳, 刘友波, 胥威汀, 等. 计及市场化电价信号导向的含高比例水电输电网扩展规划[J]. 电力自动化设备, 2021, 41(11): 190-198.
WANG Fuyang, LIU Youbo, XU Weiting, et al. Expansion planning of transmission network with high proportion of hydropower considering guidance of marketized electricity price signal [J]. Electric Power Automation Equipment, 2021, 41(11): 190-198.
 - [10] 林志坚, 杜旭昕, 黄朝凯, 等. 电力企业基于准实时数据电价异常分析的设计与应用[J]. 电子设计工程, 2017, 25(22): 58-61.
LIN Zhijian, DU Xuxin, HUANG Chaokai, et al. Design and application of analysis based on quasi real time data of abnormal price [J]. Electronic Design Engineering, 2017, 25(22): 58-61.
 - [11] 李正欣, 赵林度. 基于支持向量机的电价价格钉预测[J]. 中国电力, 2007, 40(11): 42-45.
LI Zhengxin, ZHAO Lindu. An electricity price spikes forecasting model based on SVM [J]. Electric Power, 2007, 40(11): 42-45.
 - [12] 赵越, 白杨, 刘思捷, 等. 我国电力市场建设中的动态电价风险评估研究[J]. 价格理论与实践, 2021(11): 167-172, 199.
ZHAO Yue, BAI Yang, LIU Sijie, et al. Research on dynamic rate risk assessment in my country's power market construction [J]. Price: Theory & Practice, 2021(11): 167-172, 199.
 - [13] 李文杰, 闫世强, 蒋莹, 等. 自适应确定DBSCAN算法参数的算法研究[J]. 计算机工程与应用, 2019, 55(5): 1-7, 148.
LI Wenjie, YAN Shiqiang, JIANG Ying, et al. Research on method of self-adaptive determination of DBSCAN algorithm parameters [J]. Computer Engineering and Applications, 2019, 55(5): 1-7, 148.
 - [14] 杨锡运, 刘玉奇, 李建林. 基于四分位法的含储能光伏电站可靠性置信区间计算方法[J]. 电工技术学报, 2017, 32(15): 136-144.
YANG Xiyun, LIU Yuqi, LI Jianlin. Reliability confidence interval calculation method for photovoltaic power station with energy storage based on quartile method [J]. Transactions of China Electrotechnical Society, 2017, 32(15): 136-144.
 - [15] 王梓齐, 刘长良. 基于Box-Cox变换和相对熵残差分析的风电机组齿轮箱状态监测[J]. 中国电机工程学报, 2020, 40(13): 4210-4219.
WANG Ziqi, LIU Changliang. Wind turbine gearbox condition monitoring based on Box-Cox transformation and relative entropy residual analysis [J]. Proceedings of the CSEE, 2020, 40(13): 4210-4219.
 - [16] ZAMANI-DEHKORDI P, RAKAI L, ZAREIPOUR H. A data-driven method to detect the abnormal instances in an electricity market [C]//2015 IEEE 14th International Conference on Machine Learning and Applications, December 9-11, 2015, Miami, USA. New York: IEEE, 2015: 1050-1055.
 - [17] 陶永辉, 王勇. 基于改进的K-means风电机电异常数据检测[J]. 国外电子测量技术, 2023, 42(4): 141-148.
TAO Yonghui, WANG Yong. Abnormal data detection of wind turbine based on improved K-means [J]. Foreign Electronic Measurement Technology, 2023, 42(4): 141-148.
 - [18] 郭莉, 吴晨, 薛贵元. 基于DBSCAN聚类的高温发电大数据异常检测模型[J]. 工业加热, 2023, 52(1): 35-38, 48.
GUO Li, WU Chen, XUE Guiyuan. Anomaly detection model of thermal power generation big data based on DBSCAN Clustering [J]. Industrial Heating, 2023, 52(1): 35-38, 48.
 - [19] 赵娇. 基于BIRCH聚类算法的高维传感器数据异常检测[J]. 传感技术学报, 2022, 35(12): 1686-1690.
ZHAO Jiao. Anomaly detection of high dimensional sensor data based on BIRCH clustering algorithm [J]. Chinese Journal of Sensors and Actuators, 2022, 35(12): 1686-1690.
 - [20] 赖思银. 基于PCA和改进SVDD的异常检测算法[J]. 微型电脑应用, 2022, 38(8): 129-132.
LAI Siyin. Anomaly detection algorithm based on PCA and improved SVDD [J]. Microcomputer Applications, 2022, 38(8): 129-132.
 - [21] 向黎黎, 肖私宇, 钟爱, 等. 基于PCA和决策树模型的异常电费数据检测和识别研究[J]. 电力大数据, 2022, 25(4): 42-47.
XIANG Lili, XIAO Siyu, ZHONG Ai, et al. Research on the detection and identification of abnormal electricity bill data based on PCA and decision tree Model [J]. Power Systems and Big Data, 2022, 25(4): 42-47.
 - [22] 王宏伟. 基于深度学习的发电机组异常检测和预警研究[D]. 南京: 南京邮电大学, 2022.
 - [23] 宋嗣博, 郭红霞, 杨苹, 等. 基于节点边际电价的电力市场分区策略研究[J]. 电力建设, 2017, 38(9): 132-138.
SONG Sibao, GUO Hongxia, YANG Ping, et al. Electricity

market partitioning strategy based on locational marginal price [J]. Electric Power Construction, 2017, 38(9): 132 – 138.

[24] LU Y, CORANDER J, YANG Z R. Doubly stochastic neighbor embedding on spheres [J]. Pattern Recognition Letters, 2019(128): 100 – 110.

[25] KITAZONO J, GROZAVU N, ROGOVSCHI N, et al. T-distributed stochastic neighbor embedding with inhomogeneous degrees of freedom [J]. Lecture Notes in Artificial Intelligence, 2017(9949): 119 – 128.

[26] 蔡湛,李泽文,邹睿奇,等. 基于负荷转移的售电公司分时电价优化策略[J]. 南方电网技术, 2024, 18(1): 77 – 84.

CAI Zhan ,LI Zewen, ZOU Ruiqi, et al. Optimization strategy

of time-of-use electricity price for electricity selling companies based on load transfer [J]. Southern Power System Technology, 2024, 18(1): 77 – 84.

收稿日期: 2023-09-13; 网络首发日期: 2024-05-13

作者简介:

白云霄(1988), 男, 高级工程师, 硕士, 研究方向为电力市场建设与运营, 516486451@qq.com;

张云勇(1987), 男, 高级工程师, 研究方向为电力系统市场化, 121769438@qq.com;

徐云(1994), 女, 工程师, 硕士, 研究方向为电力市场建设与运营, 1192678054@qq.com.

(上接第 172 页 Continued from Page 172)

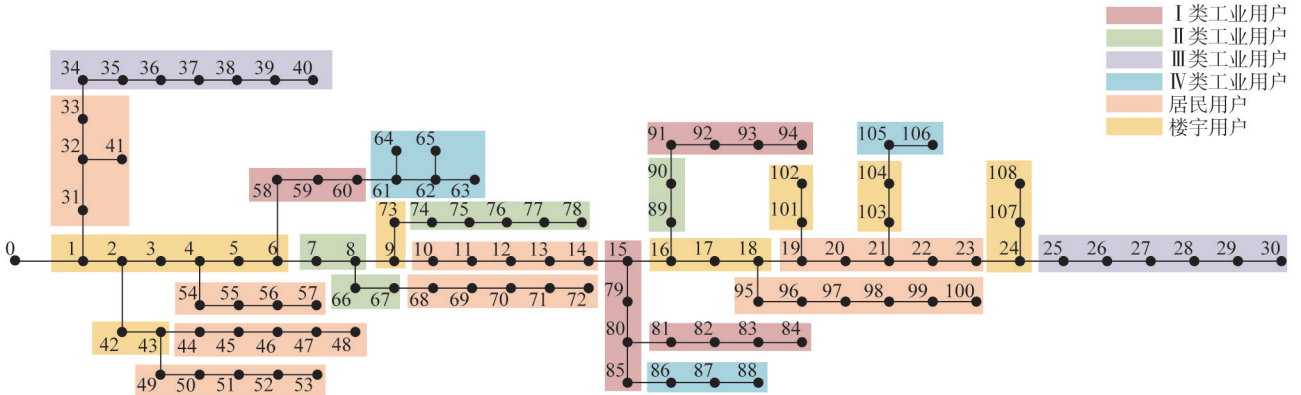


图 A3 109 节点系统负荷区域划分及对应负荷种类图

Fig. A3 109-node system load area division and corresponding load type diagram

表 A1 光伏安装位置及安装容量

Tab. A1 Photovoltaic installation position and installation capacity									
光伏安装节点	3	7	8	14	20	22	25	31	53
安装容量/kW	120	70	70	60	70	100	40	100	100
光伏安装节点	58	68	73	74	76	86	87	96	108
安装容量/kW	60	60	100	70	100	120	100	70	100