

[数据挖掘]

基于数据挖掘的电能量数据异常自动预警方法

文/张江王桐张理放
内蒙古电力科学研究院

在对电能量数据异常进行预警时，由于电能量数据异常的类型复杂，导致预警的灵敏度低，为此，提出基于数据挖掘的电能量数据异常自动预警方法研究。在电能量数据异常特征提取过程中，引入了数据挖掘算法，借助聚类算法对异常电能量数据进行聚类处理，并差异化设置了不同样本数据和聚类的权值参数，根据聚类数据与聚类中心的距离，设置具体的模糊隶属度后，以聚类中心为导向计算对应的电能量数据异常特征，并按照具体的分布情况对各个异常特征的密度指标加以分析。在电能量数据异常自动预警阶段，利用小波变换的方式过滤原始电能量数据中的噪声后，结合数据特征与提取电能量数据异常特征之间的关系作出预警。在测试结果中，当电能量数据异常程度为0.1%时，设计方法可以实现准确预警，具有较高的灵敏度。

在电力企业改革不断深化的背景下，以电力系统为核心的信息化建设受到了越来越多的关注，电力企业和用户对电能量数据的质量要求逐渐提高。这些变化对于电量采集提出了更高的要求，对应的监测系统面临着新的挑战。结合现阶段电能量计量系统稳定成长的发展状态，90.0%以上的供电网实现了独立电能量计量系统的配置，由此带来的影响具有双面性。从积极的角度分析，电能量的采集结果更加精准，对后续的统计分析以及规划配置工作具有良好的辅助作用；但是从负面的角度分析，电能量数据的规模大幅提升，异常数据问题成为难以规避的问题之一。在此基础上，为了保障电力交易的公平性和公正性，确保电能量数据的有效性成为重要的工作内容，对应的工作目标就是在保证电能量数据的完整性和准确性的基础上，提高交易阶段数据的可靠性。为此，强化电能量数据异常检测的效果是十分必要的。传统与阈值为基础的数据异常预警方法虽然具有操作简单便捷的特点，但无法实

现对阈值内异常的有效识别，难以满足现阶段高精度的电能量数据质量管理要求。

本文提出一种基于数据挖掘的电能量数据异常自动预警方法，以实际电能量数据为基础，采用对比测试的方式对预警方法的应用效果加以验证分析。

电能量数据异常自动预警方法设计

1. 基于数据挖掘的电能量数据异常特征提取

要实现电能量数据异常的有效预警，就要确定异常电能量数据的特征。与其他指标参数相比，异常电能量数据特征可以更加直观地体现电能量数据的波动特性与变异特性。本文引入了数据挖掘算法，借助聚类算法对异常电能量数据进行聚类处理。

首先，假设电能量数据异常特征提取的执行数据集为 N ，其中包含的异常电能量数据样本集合数量为 M 。考虑到不同异常电能量数据样本对应特

征的分布情况存在一定的差异,因此,不同的样本设置了对应的权值参数,用 $d_i(t)$ 表示。对应地,不同权值异常电能量数据样本的聚类结果也不同,对应的权值参数需要进行差异化设置,用 $d'_i(t-1)$ 表示。在利用数据挖掘算法中的聚类算法对异常电能量数据样本进行聚类处理时,聚类中心数量的设置也是影响最终聚类效果(电能量数据异常特征提取效果)的重要因素之一,为此,结合参与特征提取异常电能量数据样本的数量,设置聚类中心的数量与其保持一致,避免出现特征提取结果不完整的情况。以异常电能量数据聚类中心为基础的聚类结果为

$$x(i) = d_i(t)y(i) \rightarrow d'_i(l) \quad (1)$$

式中, $d'_i(l)$ 为 $y(i)$ 异常电能量数据的聚类中心; $x(i)$ 为以 $d'_i(l)$ 异常电能量数据聚类中心为基础的聚类结果。

结合同一聚类中心下异常电能量数据与聚类中心之间的距离,聚类中心的模糊隶属度计算方式为

$$\varphi_{ij} = 1 - \frac{x(i) - d'_i(l)}{\lambda} \quad (2)$$

式中, φ_{ij} 为聚类数据对于聚类中心的模糊隶属度参数; λ 为距离消减系数。

结合式(2)可以看出,数据与聚类中心的距离越远,对应的隶属度越小;相反地,数据与聚类中心的距离越近,对应的隶属度越大。按照这样的方式,得到以聚类中心为导向的异常电能量数据特征为

$$a_i = \varphi_{ij}x(i)d'_i(t-1) \quad (3)$$

式中, a_i 为以 $d'_i(l)$ 聚类中心为导向的异常电能量数据特征。

整体数据的异常特征为

$$A = (a_1, a_2, \dots, a_i, \dots, a_m) \quad (4)$$

式中, A 为异常电能量数据特征集合; m 为异常电能量数据特征的维度。

各个特征在电能量数据中分布密度为

$$E_i = \sum \exp\left\{-\frac{\|a_i - a_j\|^2}{(0.5r_a)^2}\right\} \quad (5)$$

式中, E_i 为异常电能量数据特征 a_i 的密度指标参数。 r_a 为异常电能量数据特征 a_i 的聚类区间半径。

按照上述方式,实现对电能量数据异常特征的全面提取,为后续的电能量数据异常自动预警提供执行基础。

2. 电能量数据异常自动预警

结合电能量数据异常特征的提取结果,对电能量数据异常自动预警进行设计时,充分考虑了原始电能量数据中的噪声因素,为了避免出现错误预警或异常未预警的情况,首先对电能量数据进行降噪处理。由于电能量本身具有一定的波动属性,最终会在一定程度上影响异常数据的预警效果。对此,利用小波变换的方式对电能量数据进行预处理。将数据波动的范围作为阈值函数设置的基础,阈值函数为

$$c = \begin{cases} \alpha - \frac{k^2}{\alpha e^{\frac{\alpha - k^2}{n}}}, & |\alpha| \geq k \\ 0, & |\alpha| < k \end{cases} \quad (6)$$

式中, c 为小波系数; α 为电能量数据的波动; k 为数据波动的范围。

需要注意的是, n 值的取值结果为常量,对应的范围为正整数。在此基础上,利用阈值函数对原始的电能量进行小波变换处理,计算式为

$$f(h) = \frac{1}{\sqrt{h}} \int_{-\infty}^{+\infty} h(t)cdt \quad (7)$$

式中, $f(h)$ 为小波变换处理后的电能量信息; $h(t)$ 为原始的电能量参数。

按照这样的方式,实现对原始的电能量数据中噪声的过滤,为后续预警结果的准确性提供保障。

预警主要是根据过滤后电能量数据特征与提取电能量数据异常特征之间的关系进行的,分析方式为

$$\chi = f(h) * E_i \quad (8)$$

式中, χ 为电能量数据异常预警判定参数。

当 χ 的计算结果大于以 $d'_i(l)$ 聚类中心为导向的异常电能量数据特征值时,电能量数据已经具备 a_i 异常特征,则自动预警;当 χ 的计算结果小于以 $d'_i(l)$ 聚类中心为导向的异常电能量数据特征值时,电能量数据不具备 a_i 异常特征,则不进行预警;当 χ 的计算结果等于以 $d'_i(l)$ 聚类中心为导向的异常电能量数据特征值时,电能量数据呈现出 a_i 异常特征趋势,则自动预警。通过将电能量数据与提取到的异常电能量数据特征集合 A 中各个参数的关系进行比较,即可实现对整体异常情况的全面预警,为电

量数据的管理提供保障。

应用测试

1. 测试数据准备

以实际电力系统中的某电力计量装置的数据作为测试数据。该装置某月连续20天的正向有功电能数据见表1。

表1 正向有功电能数据

时间	电量/ (kW·h)	时间	电量/ (kW·h)
6月1日	15.22	6月11日	15.22
6月2日	16.37	6月12日	15.76
6月3日	15.15	6月13日	16.03
6月4日	15.86	6月14日	16.20
6月5日	15.36	6月15日	15.47
6月6日	16.30	6月16日	14.93
6月7日	16.44	6月17日	14.99
6月8日	16.12	6月18日	15.46
6月9日	14.96	6月19日	15.68
6月10日	14.95	6月20日	15.38

以表1数据为基础，分别在其中嵌入异常数据，对应的异常程度分别为5.0%、3.0%、2.0%、1.0%、0.5%、0.3%及0.1%，分析对应的预警效果。

在电能数据异常预警性能测试的过程中，为了能够更加直观地分析本文设计方法的测试结果，分别采用非参数核密度估计为基础的电能数据异常预警性方法和以奇异值分解（Singular Value Decomposition, SVD）为基础的电能数据异常预警性方法在相同的环境下进行预警测试。通过对比三种方法对不同程度电能数据异常的预警情况，对本文设计方法的性能进行评价。

2. 测试结果

不同预警方法测试结果对比见表2。

表2 不同预警方法测试结果对比

异常程度 (%)	非参数核密度估计预警方法	SVD预警方法	本文设计预警方法
5.0	预警	预警	预警
3.0	预警	预警	预警
2.0	预警	预警	预警
1.0	未预警	预警	预警
0.5	未预警	未预警	预警
0.3	未预警	未预警	预警
0.1	未预警	未预警	预警

由表2可知：当电能数据异常程度≥2.0%

时，三种方法均实现了有效预警；当电能数据异常程度≤1.0%时，非参数核密度估计预警方法未能预警，表明该预警方法对于小幅度异常的灵敏度较低；当电能数据异常程度≤0.5%时，SVD预警方法也未能预警，表明该方法虽然比非参数核密度估计预警方法在异常灵敏度方面有一定程度的提升，但仍难以实现对微小异常的有效预警。当电能数据异常程度仅为0.1%时，本文设计的预警方法仍实现了准确预警。因此，本文设计的基于数据挖掘的电能数据异常自动预警方法可以实现对不同程度异常的有效预警，具有较高的灵敏度，能够满足高精度的电能数据管理要求。

结束语

故障因素或人为因素导致电能采集设备采集到的电能数据存在不同程度的异常，对异常数据进行有效的预警是保障电能交易顺利进行的重要基础，同时也为开展相关校正工作提供重要的指导。本文提出基于数据挖掘的电能数据异常自动预警方法，实现了对不同程度数据异常的精准预警，对于实际的数据管理和电力系统管理具有重要的现实意义。希望通过本方法为异常的用电行为、偷电行为管控及电力相关设备故障维修提供辅助。**EA**

(收稿日期：2023-05-24)

参考文献

[1] 刘丰硕，刘然，董子慧．基于k-means聚类的智能电表异常动态数据检测方法[J]．电子设计工程，2023，31(5)：84-88.

[2] 妙红英，周凤华，王艳芹，等．基于SVD的电能计量装置异常数据快速辨识方法[J]．电子设计工程，2023，31(3)：81-84，89.

[3] 叶伟，邓刚，叶攀，等．基于关联规则算法的电池剩余电量数据监测方法[J]．电子设计工程，2022，30(13)：105-108，113.

[4] 严明辉，潘舒宸，吴滇宁，等．基于非参数核密度估计的电力市场用户电量异常数据辨识与修正方法[J]．现代电力，2022，39(1)：80-87.

[5] 彭孝强．地区电网基于OMS主站系统电量智能化计算统计的应用[J]．集成电路应用，2021，38(11)：50-51.