

张娟, 朱建璇, 刘书明, 等. 基于数据场聚类的 DMA 小区漏损分级预警[J]. 净水技术, 2025, 44(11): 174-183.

ZHANG J, ZHU J X, LIU S M, et al. DMA community leakage classification warning based on data field clustering [J]. Water Purification Technology, 2025, 44(11): 174-183.

基于数据场聚类的 DMA 小区漏损分级预警

张娟^{1,2}, 朱建璇¹, 刘书明^{2,*}, 牛豫海¹, 王晓婷³

(1. 河北建投水务投资有限公司, 河北石家庄 050051; 2. 清华大学环境学院, 北京 100084; 3. 京东集团, 北京 100176)

摘要 【目的】针对传统漏损监测与处理方法在多独立计量区域(DMA)小区漏损预警应用中存在处理优先级难以明确的问题,提出一种基于数据场聚类的 DMA 小区漏损分级预警方法,旨在提升多个 DMA 小区漏损处理的时效性与准确性,为水务企业多 DMA 小区漏损情况下的处理优先级方案。【方法】首先,通过 Z-score 与主成分分析(PCA)对漏损率异常的 DMA 小区的日均异常漏损率、日均异常漏损量、最小夜间流量、最小夜间流量/日流量共 4 个指标进行数据标准化和降维处理,提取前 2 项主成分作为聚类特征;其次,结合数据场理论构建高斯势函数,利用遗传算法优化辐射因子,并基于势熵最小化原则确定最优聚类参数;最后,根据势值分布划分漏损预警等级,并引入持续时长机制实现动态预警。【结果】基于 L 市 30 个 DMA 小区的漏损数据试验表明:该方法聚类效果显著,成功将漏损事件划分为 5 个聚类预警级别,其中 1 级预警占比为 73.3%(22 个小区),5 级预警仅为 3.3%(1 个小区),符合现实情况;动态预警机制与实际工作环境相融合,提高了漏损处理的效率。【结论】本文方法通过数据场聚类与动态机制融合,能够精准划分漏损等级并实现实时动态预警。实际应用中可为水务企业提供漏损处理的优先级决策依据,推动供水管网管理向数字化、精细化方向升级。

关键词 数据场聚类 独立计量区域(DMA) 漏损预警 预警优先级 动态预警

中图分类号: TU991 **文献标志码:** A **文章编号:** 1009-0177(2025)11-0174-10

DOI: 10.15890/j.cnki.jsjs.2025.11.020

DMA Community Leakage Classification Warning Based on Data Field Clustering

ZHANG Juan^{1,2}, ZHU Jianxuan¹, LIU Shuming^{2,*}, NIU Yuhai¹, WANG Xiaoting³

(1. HCIG Water Investment Co., Ltd., Shijiazhuang 050051, China;

2. School of Environment, Tsinghua University, Beijing 100084, China;

3. JD.com, Inc, Beijing 100176, China)

Abstract [Objective] To solve the problem that the priority of traditional leakage monitoring and processing method in the application of multi district metering area (DMA) community leakage early warning is difficult to determine, a DMA community leakage hierarchical early warning method based on data field clustering is proposed, aiming to improve the timeliness and accuracy of multiple DMA community leakage processing, and to provide a priority scheme for water enterprises in the case of multi DMA community leakage. [Methods] Firstly, through Z-score and principal component analysis (PCA), data standardization and dimensionality reduction were performed on the four indicators of DMA communities with abnormal leakage rate, abnormal leakage amount, minimum night flow, and minimum night flow/daily flow, and the first two principal components were extracted as clustering features; Secondly, the Gaussian potential function was constructed based on the data field theory, the radiation factor was optimized using genetic algorithm, and the optimal clustering parameters were determined based on the principle of minimizing the potential entropy; Finally, according to the potential distribution, the leakage warning level was divided, and the duration mechanism was introduced to achieve dynamic warning. [Results] Based on the leakage data of 30 DMA communities in L City, the experiment

[收稿日期] 2025-02-12

[基金项目] 河北省重大科技成果转化专项(23283601Z)

[作者简介] 张娟(1986—),女,正高级工程师,E-mail:13739737722@163.com。

[通信作者] 刘书明(1976—),男,教授,研究方向为供水管网漏损管理、供水管网优化设计管理、管网水质保障、水能关系等,E-mail:shumingliu@tsinghua.edu.cn.

showed that the clustering effect of this method was remarkable, and the leakage events were successfully divided into five cluster warning levels, of which the first level warning accounts for 73.3% (22 communities), and the fifth level warning only accounts for 3.3% (1 community), which was in line with the reality; The dynamic early warning mechanism was integrated with the actual working environment to improve the efficiency of leakage treatment. [Conclusion] Through data field clustering and dynamic mechanism fusion, this method can accurately classify leakage levels and achieve real-time dynamic early warning. In practical application, it can provide priority decision-making basis for leakage treatment for water enterprises, and promote the upgrading of water supply pipe network management in the direction of digitalization and refinement.

Keywords data field clustering district metering area (DMA) leakage warning warning priority dynamic warning

在都市化浪潮奔涌向前的当下,供水管网漏损问题已然成为水务行业亟待攻克的难关。相关数据^[1]揭示,全国范围内,每年因管网漏损而浪费的水资源高达数十亿 m³,这无疑加剧了水资源的短缺困境,同时致使供水企业的运营成本攀升至高位。在此背景下,独立计量区域(DMA)技术凭借精细化管理的显著优势,被广泛应用于漏损监测与控制领域^[2]。然而,传统漏损检测方法在 DMA 小区的应用过程中,暴露出诸多问题,如诊断存在延迟、预警效果欠佳等,这使得探索更为高效的漏损预警技术迫在眉睫。

当前,学术界针对 DMA 漏损检测展开了深入研究,提出了多种方法。基于最小夜间流量(minimum night flow, MNF)的定额法与比例法,虽然能快速锁定存量漏损^[3-5],但对于新增漏损的判定,却需依赖流量分布法的辅助,导致实时性不足^[6];组合流量法通过融合全天流量与 MNF 数据,将漏损诊断准确率提升至 91.28%^[1],但其阈值矩阵的构建过度依赖人工经验,难以契合动态数据变化的节奏。随着科技的进步,为突破传统方法的局限,数据挖掘与机器学习技术逐渐被引入该领域^[7]。基于粒子群优化的最小二乘支持向量机模型,借助预测误差分布实现漏损检测,但其模型训练过程对大量标注数据存在过度依赖,难以迅速应对突发性漏损^[8];遗传算法在优化反向传播(back propagation, BP)神经网络参数方面表现卓越,能够显著提升漏失定位的收敛速度与精度^[9],但此类监督学习方法对数据完备性的要求过高。基于概率分布模型的漏损分级预警方法指标相对单一,难以考虑综合情况^[10]。聚类和自编码器等无监督方法,近年来也被应用于分级预警和异常检测^[11-13]。然而,传统聚类算法在处理高维、非线性流量数据时,易受噪声干扰,致使聚类效果不尽如人意。尽管 K 均值(KS)聚

类算法通过引入形状距离度量,在一定程度上提升了聚类精度^[14],但其参数选择缺乏自适应性,限制了其在实际应用中的推广范围。主成分分析(PCA)作为经典的降维技术,在提取流量数据关键特征方面具有显著优势^[15],但现有研究多聚焦于静态数据分析,对于动态预警机制的探索相对匮乏。

鉴于上述问题,本文创新性地提出一种基于数据场聚类的 DMA 小区漏损分级预警方法。首先,运用 Z-score 标准化与 PCA 对多维度漏损指标进行降维处理,精准提取关键特征;其次,结合数据场理论构建势函数,借助遗传算法优化辐射因子,并依据势熵最小化原则确定最优聚类参数;最后,基于势值分布划分漏损预警等级,并引入持续时长机制,实现动态预警。

1 研究方法

1.1 数据预处理

将漏损率异常的 DMA 小区挑选出来,用于漏损分级预警。DMA 小区管网漏损和多个指标有关,其中日均 MNF、日均 MNF/日均流量、日均漏损水量、日均漏损率是最常用的 4 个指标。

1.1.1 Z-score 标准化处理

首先,每个指标之间数值相差较大,本文用 Z-score 方法进行标准化处理,如式(1)~式(2)。

$$z = \frac{x_i - \mu}{s} \quad (1)$$

$$s = \sqrt{\frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})^2} \quad (2)$$

其中:z——标准分数;

x_i ——样本值;

\bar{x} ——样本平均值;

μ ——数据集的均值;

s——数据集的标准差;

m ——样本个数。

标准化可将输入的数据转换为均值为 0、标准差为 1 的数据,以减少异常值对漏损异常值分类结果的影响。

1.1.2 PCA

本文利用 PCA 对日均 MNF、日均 MNF/日均流量、日均漏损水量、日均漏损率这 4 个指标进行特征提取。PCA 是一种经典的降维技术,其主要目标是通过线性变换将高维数据投影到低维空间,提取数据中的主要特征,减少变量之间的相关性和数据冗余,同时保留数据的主要信息。

将输入数据构建为矩阵,如式(3), m 为 30。每个样本有 4 种变量,即 30 个 DMA 小区的 MNF、MNF/日均流量、日均漏损水量以及日均漏损率数据。

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & x_{m4} \end{bmatrix} \quad (3)$$

经标准化处理后,方差矩阵表示如式(4)。

$$C = \frac{1}{4} X^T X \quad (4)$$

为了确定协方差矩阵的主成分,采用雅可比行列式法如式(5),计算协方差矩阵的特征值(k_i)和特征向量(λ_i)。通过迭代过程将协方差矩阵对角化,直至满足条件 $|R - \lambda I| = 0$ 。

$$|R - \lambda I| = 0 \quad (5)$$

其中: R ——旋转矩阵;

I ——单位矩阵;

λ ——特征向量。

将特征值 k_i 排序,即 $k_1 > k_2 > k_3 > k_4$,该排序揭示了每个特征值在数据集中的重要性。计算每个特征值占全部特征值的总数的比例,以此确定每个特征值的贡献率。贡献率的计算如式(6)。

$$W_i = \frac{k_i}{\sum_{i=1}^4 k_i} \quad (6)$$

其中: $\sum_{i=1}^4 k_i$ ——累计特征值;

W_i ——贡献率。

计算得到贡献度最大的 2 种特征分量可以作为主成分,如选用第一主成分 F_1 、第二主成分 F_2 来表示所有漏损数据的信息,用来当作聚类属性。

1.2 数据场

数据场是基于物理场提出的概念,用于解决数据挖掘和模式识别中非线性复杂问题的一种方法^[16]。数据场理论的基本思想源于物理学中的场理论,将数据看作“物质点”,每个数据点根据其属性对空间施加影响力,形成一个能够表征数据分布特征的“场”。这一方法能够有效地处理高维、非线性和噪声数据,已被广泛应用于聚类分析、分类问题和异常检测等领域。

1.2.1 场强函数

为了定量描述数据场中源点对周围数据对象的影响力度,引入了场强函数的概念。该函数类比于物理学中的引力和静电力公式,用以表征数据场中源点对周围空间中其他数据对象的作用强度。考虑到高斯分布的普适性和短程场作用在表达数据聚类特性时的便利性,场强函数通常定义 $[f_y(x)]$ 为式(7)。

$$f_y(x) = \rho e^{-\frac{d^2(x,y)}{2\sigma^2}} \quad (7)$$

其中: $d(x,y)$ ——数据对象 x,y 之间的欧几里得距离;

σ ——辐射因子;

ρ ——权重变量,在数据场的假设条件下,所有数据点具有相同的作用能力时,则 ρ 被设为 1。

式(7)表明,数据点 x,y 距离越近, $d(x,y)$ 越小,场强函数值 $f_y(x)$ 越大,表明源点 y 对周围数据对象 x 的影响增强。当 2 个数据点在空间上重合时,距离为 0, $d(x,y) = 0$,场强函数值达到最大值,当 2 个数据点远离时, $d(x,y)$ 越来越大,场强函数值逐渐减小,趋近 0,表明源点的影响随距离的增加而减弱。式(7)中 σ 量化了数据点对周围数据对象的影响力度,场强函数值与 σ 成反比, σ 越小,数据点的辐射能力越强。

1.2.2 势函数

在数据挖掘领域,研究单个数据对象的场强函数对于理解整个数据集的全局特性具有局限性。因此,为了深入分析数据场中各数据对象相互作用

的总体效应,必须考察所有数据对象在空间中产生的场强值的综合变化。为此,引入势函数的概念,用以表征空间中特定点 y 处,由所有数据对象共同作用所产生的场强值累加结果。该势函数 $[\phi(y)]$ 的数学表达式为式(8)。

$$\phi(y) = \sum_{i=1}^n f_y(x_i) = \sum_{i=1}^n \rho e^{-\frac{d^2(x_i, y)}{2\sigma^2}} \quad (8)$$

其中: n ——数据对象的总数。

式(8)表明势函数是一个高斯核函数,它将所有数据对象在点 y 处的场强值进行叠加。由此定义的势函数,每个数据对象点都以自己为中心,独立地对其周围空间施加影响,而不受外界的影响,具有独立性;每个数据对象点的势值是所有数据对象在该点产生的场强值的叠加总和,具有叠加性;势值随着距离的增加急剧下降,靠近数据对象的区域具有较高的势值,而远离数据对象的区域具有较低的势值,具有衰减性。势函数提供了一种量化数据场中各点受数据对象影响强度的方法,为分析数据对象间的相互关系和数据场的整体结构提供了一个有用的工具。

1.2.3 辐射因子及其优化

辐射因子在势函数中扮演着关键角色,它直接影响势值的分布特征。为了深入理解辐射因子与势值之间的相互关系,首先考察空间中仅存在一个数据点时的场强和势场分布。图1分析了辐射因子如何影响单个数据点在不同距离上的场强值,这些曲线描绘了场强值随距离 d 变化的衰减特性,其中辐射因子决定了衰减的速率。

由图1可知,当距离达到 2.12σ 左右的时候,势值几乎为0,因此,为了方便计算,距离为 2.12σ 的数据点的影响直接忽略不计。

当 σ 较小时,表示数据点辐射的能量集中于较小的空间范围内,导致近场区域的势值较高,而远离数据点的区域势值迅速减小,如图中最左侧 $\sigma = 0.5$ 的数据曲线。

相反,当 σ 较大时,数据点辐射的能量分散在更大的空间范围内,使得势值随距离的衰减变得更平缓,即在较远的距离上仍然能观测到较高的势值,如图1中最右侧 $\sigma = 2$ 的数据曲线。

如果每个数据对象在空间中的势值大致相同,说明数据分布具有最大的不确定性,即熵达到最大

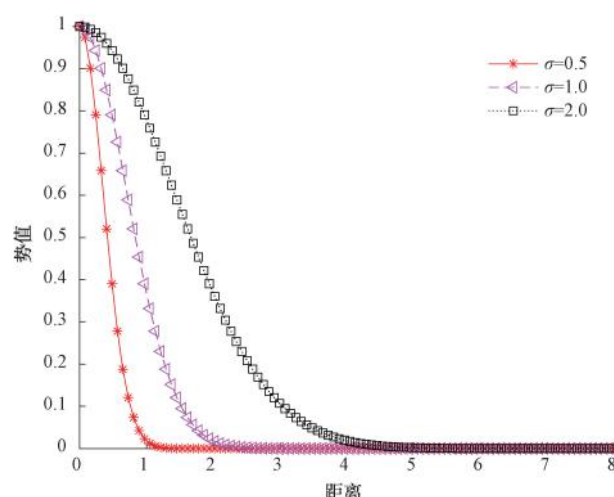


图1 不同辐射因子条件下势值随距离的变化

Fig. 1 Variation of Potential Values with Distances under Different Radiation Factor Conditions

值。熵的值越大,表明系统的不确定性越高。相反,如果数据对象所在位置的势值差异显著或分布不均,则不确定性较低,熵值较小。

势熵用以描述不同辐射因子对数据场的影响。给定一组数据点 x_1, x_2, \dots, x_n 的势值分别为 $\psi_1, \psi_2, \dots, \psi_n$, 则势熵定义为式(9)~式(11)。

$$H = - \sum_{i=1}^n \frac{\psi_i}{Z} \log \left(\frac{\psi_i}{Z} \right) \quad (9)$$

$$Z = \sum_{i=1}^n \psi_i \quad (10)$$

$$p_i = \frac{\psi_i}{\sum_{i=1}^n \psi_i} \quad (11)$$

其中: Z ——标准化因子;

H ——势熵;

ψ_i ——对应数据点的势值;

p_i ——为数据点的归一化势值,表示数据点 x_i 在总势能中的比重,概率 p_i 之和为1。

分析 H 的值总是介于0和 $\log(n)$ 之间,即 $0 \leq H \leq \log(n)$, 且 $H = \log(n) \Leftrightarrow \psi_1 = \psi_2 = \dots = \psi_n$, 此时每个 p_i 都等于 $\frac{1}{n}$, 势熵 H 达到最大值。

H 与 σ 之间的关系可以通过图2进行分析。随着 σ 的增加, H 先减小后增大。这一现象表明存在一个最优的 σ 值,使得 H 达到最小。由此,为了确定最优的 σ 值, σ 的选择问题就是 H 最小值的问

题。遗传算法通过模拟自然选择和遗传机制来搜索最优解,适用于解决此类优化问题,以选择效果最好的 σ 值。

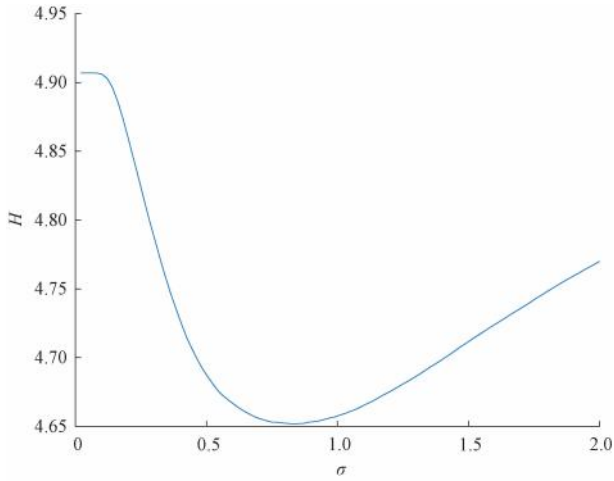


图2 势熵与辐射因子变化关系

Fig. 2 Relationship between Potential Entropy and Radiation Factors Variation

1.2.4 等势线和势心

势线和等势面是描述数据场中势能分布的几何工具。等势线是指连接数据场中具有相同势值的点的平滑曲线,而等势面则是在三维空间中,由具有相同势值的点构成的曲面,且这些点与辐射方向正交。在数据场中,辐射能量较高的区域接近数据源节点,即数据流的起点,导致等势面或等势线的分布更为密集;辐射能量较低的区域,等势线或面的分布则相对稀疏。假设有一个二维的数据场,其中包含一个正向辐射的源点,比如一个正电荷。在这个数据场中,可以画出一系列等势线,每条等势线都表示一个特定的势值。例如,如果将势值设为常数 ψ ,则所有与源点距离满足式(12)的点将连成一条等势线。

$$\psi = \frac{1}{4\pi \times \epsilon_0 r} \quad (12)$$

其中: r ——距离源点的距离;

ϵ_0 ——真空电容率。

在三维空间中,这些等势线将形成一个等势面,比如一个球面。通过改变 ψ 的值,可以得到不同的等势线和等势面,它们越接近源点,分布就越密集,表明势能的变化率更高。

势场是由所有等势面或等势线覆盖数据域空间构成的,是基于势函数叠加作用的外在表现形式。

根据场强函数定义,势场中等势线或等势面的稀疏与稠密形象化地表示了数据的抽象度强弱关系,反映了数据对象整体辐射能量的特征。例如构造一个以每个用户作为一个数据源点的社交网络的数据场,一个等势面可能包含所有活跃度在特定范围内的用户。在这个势场中,等势面的稠密区域可能表示用户活跃度较高的社区中心,而稀疏区域则可能表示用户活跃度较低的边缘地带。

在势场中,势心是等势面或等势线分布的中心点,单一数据对象产生的数据场,势心位于数据对象本身的位置。例如,一个高压系统可以被视为一个数据源点,其辐射能量与气压大小成正比。在这种情况下,势心将位于高压系统的中心,即气压最高的点。对于由2个或2个以上的数据对象构成的数据场,这些对象可以被归类为一个类簇,其势心靠近辐射能量较大的数据对象的位置,且势心与重心位置相吻合。

1.3 数据场聚类算法

数据场聚类算法的基本思路是通过调节参数,将数据对象生成不同的数据场,获得不同的聚类效果,从而满足特定的聚类目标。该算法的基本步骤如下。

- (1) 设定关键参数,如辐射因子、网格大小等。
- (2) 为每个数据对象构建一个数据场,该数据场反映了数据对象在空间中的势能分布。
- (3) 计算数据场中每个点的势值 $f_y(x)$ 。
- (4) 在数据场中用平滑曲线连接具有相同势值的点生成等势线或等势面。
- (5) 在势场中选择势值最大的点作为聚类中心,即为势心。
- (6) 根据等势线的分布,将周围的数据对象聚集到一起,形成类簇。剩余的数据对象可以作为离群点处理。
- (7) 根据初始聚类结果,调整参数以优化聚类效果。
- (8) 评估聚类结果的质量。
- (9) 迭代改进参数调整和聚类过程,直至获得满意的结果。
- (10) 输出最终的聚类结果。

数据场聚类算法的流程图如图3所示,该算法首先通过遗传算法优化辐射因子,然后构建势场并

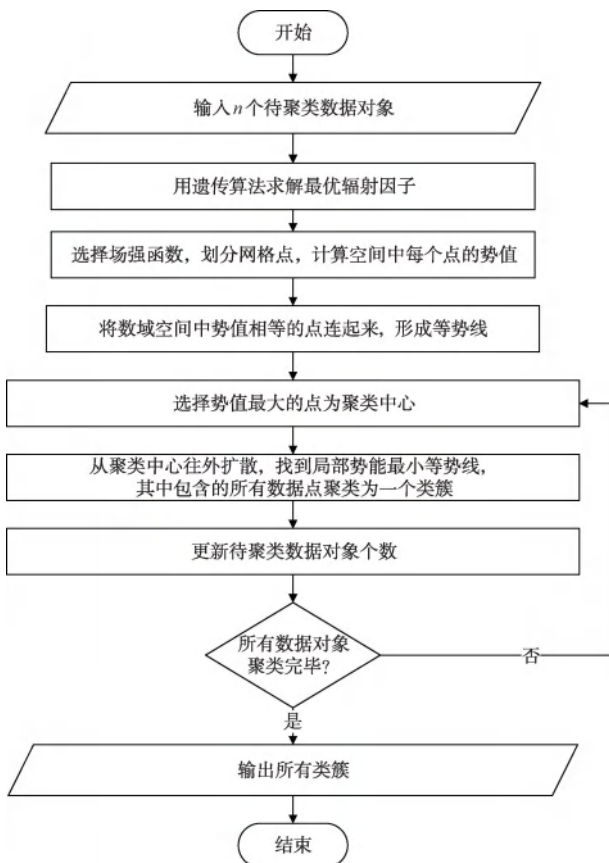


图 3 数据场聚类算法流程

Fig. 3 Flow of Data Field Clustering Algorithm

生成等势线,选择势值最大的点作为聚类中心,从该中心向外扩散形成类簇,更新待聚类数据对象数量,并循环进行直至所有数据对象被聚类,最终输出所有形成的类簇。此方法利用势场中势能的分布特性来指导聚类过程,适用于处理复杂数据结构,有效揭示数据的内在模式。

2 结果和讨论

2.1 数据场聚类算法试验数据

本文采用 L 市漏损率检测异常的 30 个 DMA 小区的 MNF、MNF/日均流量、日均漏损水量以及日均漏损率数据,对数据进行标准化处理后,再用 PCA 法得到最佳的聚类分量,最后用数据场聚类法对 DMA 小区漏损处理优先级进行评价。原始数据如表 1 所示。

2.2 数据场聚类算法试验结果分析

对原始数据进行 PCA 降维得到的数据加上 DMA 小区编号,最终形成散点图,如图 4 所示。

利用 PCA 降维得到的数据根据势熵函数对辐

表 1 L 市漏损率异常的 DMA 小区聚类原始数据
Tab. 1 DMA Clustering Raw Data of Abnormal Leakage Rates in L City

小区编号	MNF/ ($\text{m}^3 \cdot \text{h}^{-1}$)	MNF/ 日流量	日均漏 损量/ m^3	日均漏 损率
1	3.12	0.30	33.86	13.60%
2	6.78	0.36	70.17	15.72%
3	2.15	0.01	253.56	6.54%
4	7.63	0.37	56.47	11.35%
5	2.03	0.21	19.35	8.46%
6	1.79	0.38	31.26	27.47%
7	10.58	0.35	158.67	21.72%
8	5.69	0.46	48.35	16.30%
9	3.50	0.25	35.68	10.79%
10	4.37	0.27	56.72	14.76%
11	2.11	0.18	36.70	13.12%
12	6.09	0.21	69.36	9.79%
13	12.33	0.43	210.37	30.57%
14	7.91	0.16	162.46	13.55%
15	4.26	0.22	45.73	9.68%
16	5.68	0.32	52.66	12.54%
17	7.05	0.42	70.62	17.60%
18	6.88	0.41	65.40	16.33%
19	8.46	0.49	77.90	18.87%
20	4.25	0.78	29.32	22.33%
21	8.37	0.10	142.16	7.42%
22	4.65	0.26	53.92	12.33%
23	2.20	0.25	38.52	18.32%
24	7.66	0.14	121.88	9.56%
25	0.88	0.24	9.12	10.38%
26	4.33	0.26	53.57	13.47%
27	2.22	0.16	40.88	12.58%
28	8.65	0.42	88.14	17.69%
29	3.86	0.31	46.83	15.66%
30	2.57	0.19	58.50	17.95%

射因子进行优化。优化算法采用遗传算法,其中,种群大小设为 500,二进制字符串长度设为 50,交叉概率设为 0.8,变异概率设为 0.05,最大迭代轮次为 50,辐射因子区间为 $[0, 1]$ 。经过遗传算法优化得到最优辐射因子为 0.847 5,遗传算法适应度进化曲线如图 5 所示,迭代到第 21 轮目标函数值达到最优。

场强函数采用核力场函数,每个数据点的质量

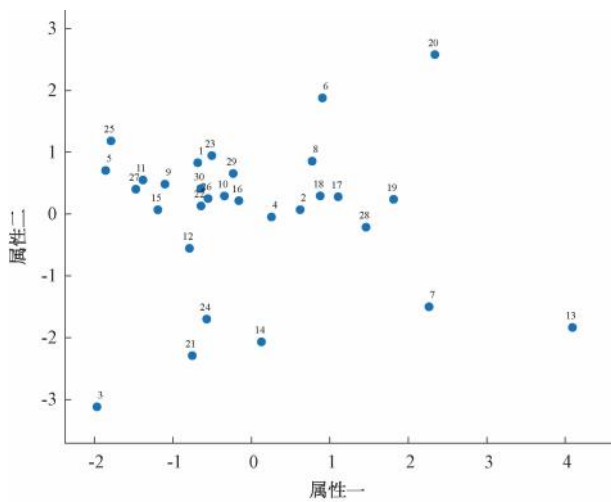


图 4 DMA 小区 PCA 降维属性散点图

Fig. 4 Attribute Scatter Plot of DMA Community PCA Dimensionality Reduction

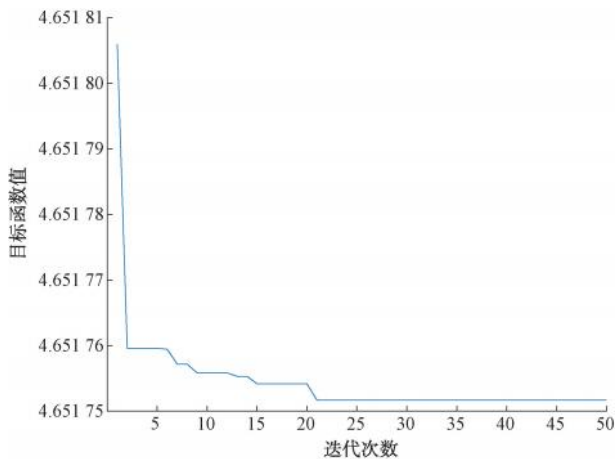


图 5 遗传算法适应度进化曲线

Fig. 5 Evolutionary Curve of Fitness of Genetic Algorithm

默认为 1,根据遗传算法优化得到的辐射因子可以计算空间内每个网格点的势值,并可以根据势值画出三维空间势值图和势分布等高线图(即等势线图),分别如图 6 和图 7 所示。

由图 6 和图 7 可知,空间中势值有一个特别大的峰值,以及几个较小的峰值,这说明大多数的 DMA 小区属于同一类,极少数的 DMA 小区属于其他类。多数漏损值异常的 DMA 小区都是轻微漏损,严重漏损属于极少数情况,所以聚类结果符合实际情况,具体的聚类结果如图 8 所示。

图 8 中不同颜色代表不同的类簇,也就是不同的 DMA 小区漏损分级预警结果。根据不同类簇中心点的势值大小,划分预警级别,类簇中心势值越小,预警级别越高,成功将漏损事件划分为 5 个聚类

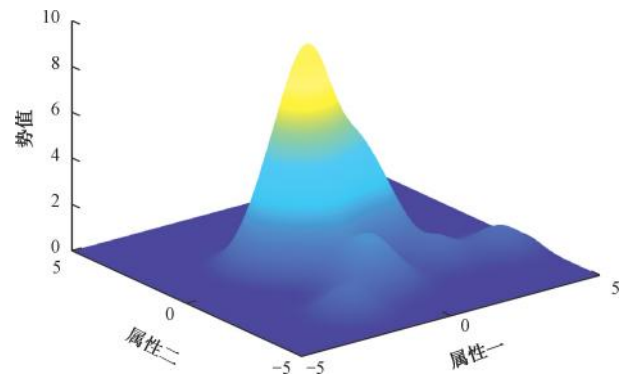


图 6 DMA 小区三维空间势值

Fig. 6 3D Spatial Potential Values of DMA Community

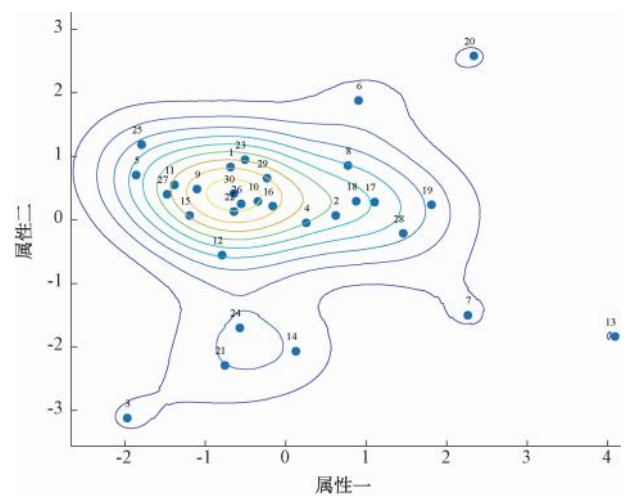


图 7 DMA 小区数据场等势线

Fig. 7 Equipotential Lines of DMA Community Data Field

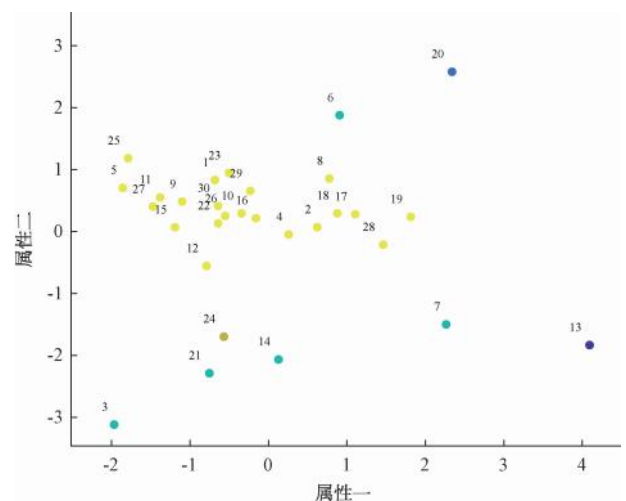


图 8 DMA 小区漏损预警分级聚类结果

Fig. 8 Clustering Results of DMA Community Leakage Warning Classification

预警级别,其中1级预警占比为73.3%(22个小区),5级预警仅为3.3%(1个小区),符合现实情况。具体划分如表2所示。

表2 DMA小区分级预警
Tab. 2 DMA Community Warning Classification

预警级别	DMA 小区编号
1	1,2,4,5,8,9,10,11,12,15,16,17,18,19,22,23,25,26,27,28,29,30
2	24
3	3,6,7,14,21
4	20
5	13

2.3 漏损预警机制设定

为了增加分级预警系统的稳定性和现实中工作人员的可操作性,本文在基于数据场聚类的基础上,采用持续最高等级时长作为实际预警级别。

为了让系统预警方案更贴近现场运维工作的实际需求,且降低用户表预测值带来的误差,系统持续监测每个DMA小区在8 h内处于最高聚类等级的时长。预警机制流程如图9所示,DMA小区位于聚类最高等级的时长大于2 h且小于等于4 h,即为轻度预警;时长大于4 h且小于等于6 h,即为中度预警;时长大于6 h为重度预警。值班人员收到现场

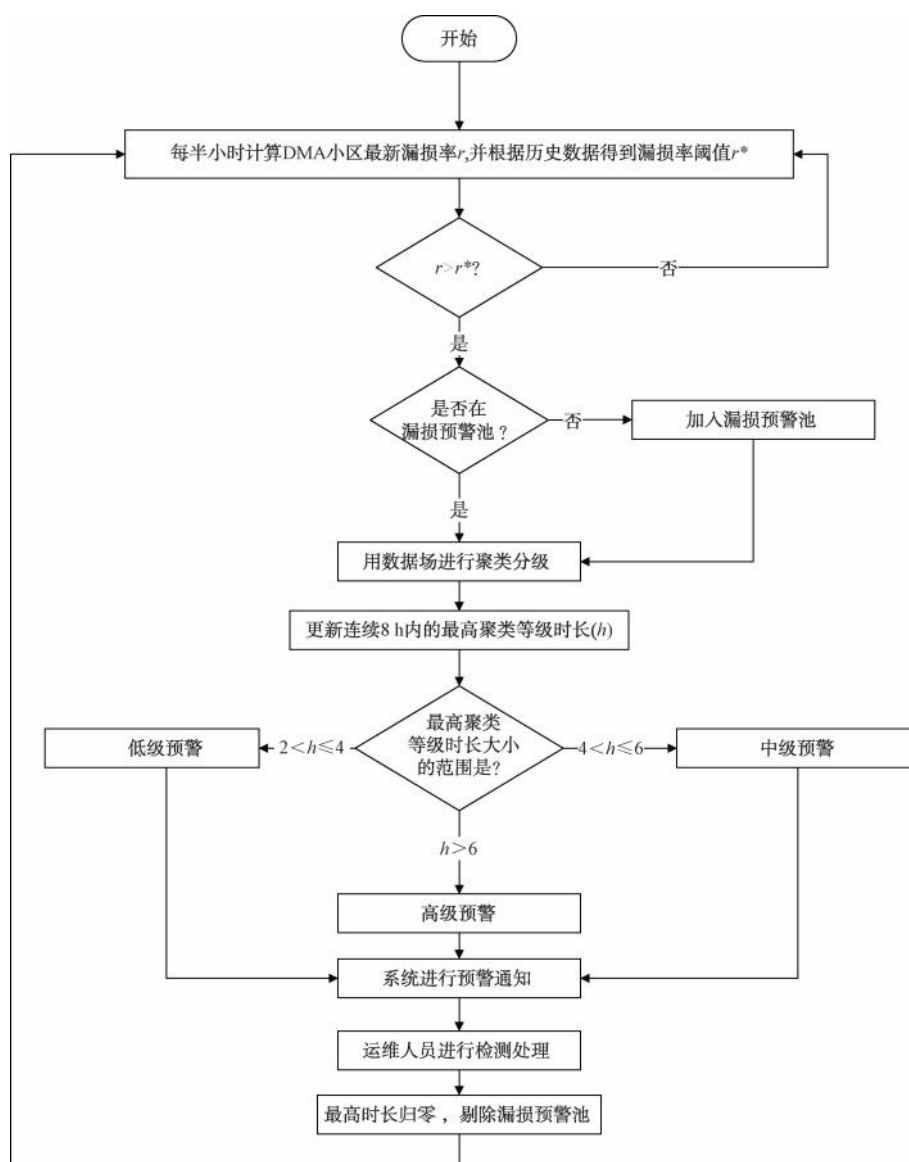


图9 预警机制流程

Fig. 9 Flow of Warning Mechanism

运维人员对漏损异常 DMA 小区完成检测的反馈后,可手动将该 DMA 小区的预警持续时长归零。

本文算法被应用于漏损预警监控系统,系统界面如图 10、图 11 所示。

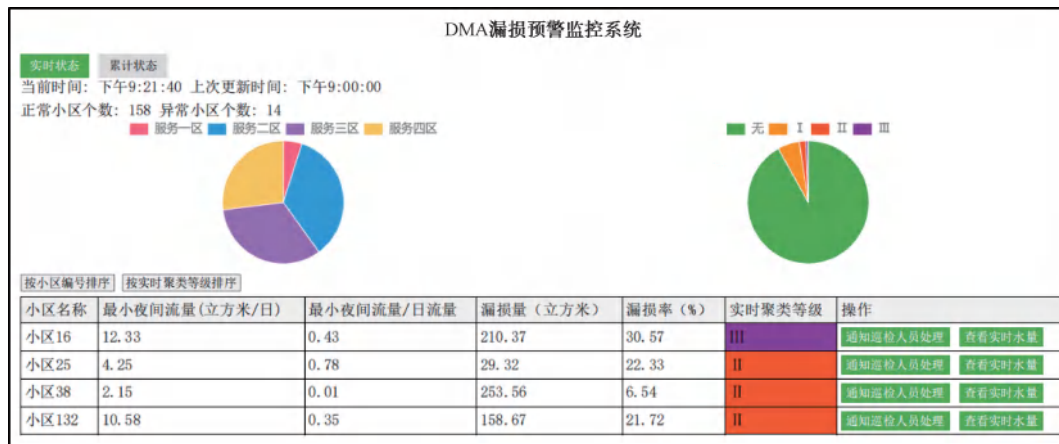


图 10 DMA 漏损预警监控系统实时状态

Fig. 10 Real Time Status of DMA Leakage Warning and Monitoring System

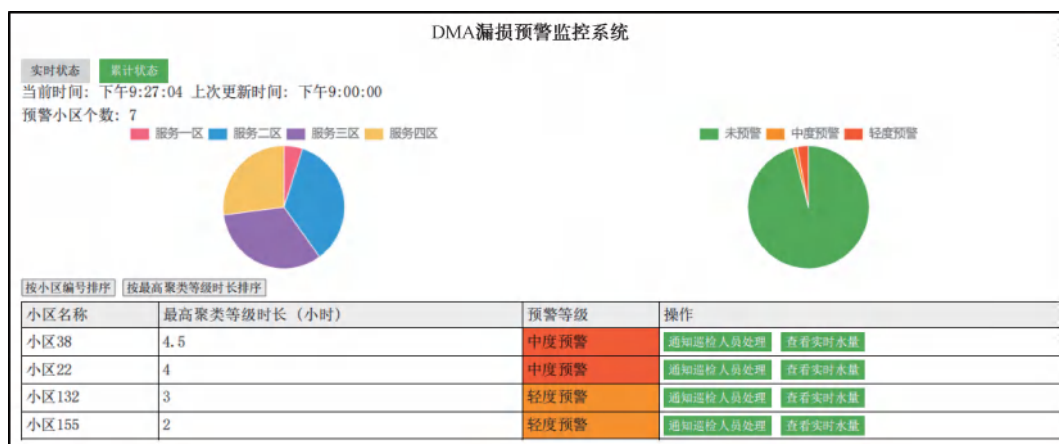


图 11 DMA 漏损预警监控系统累计状态

Fig. 11 Cumulative Status of DMA Leakage Warning and Monitoring System

图 10 为 DMA 漏损预警监控系统实时状态图,是当前半点时刻状态下漏损预警显示页面。工作人员可以根据运行经验,自主选择是否通知巡检人员处理,如不处理,当前半点时刻的预警信息自动推送至图 11 预警累计状态信息中。本文算法在漏损监控系统中的应用降低了漏损事件响应时长,同时提高了漏损监控的数字化水平和漏损管理的标准化水平。

3 结论

本文利用了基于数据场的聚类算法,首先将同城大量的 DMA 小区异常漏损率、异常漏损量、MNF、MNF/日流量应用 PCA 法得到最佳聚类分量,再用数据场聚类算法对 DMA 小区漏损率所处的预

警级别进行分级,同时根据实际工作场景建立动态预警机制,为具有同城多 DMA 小区管理需求的水务企业提供了不同优先级处理漏损预警的科学方案。本文对于工作人员排查漏损小区的优先级次序具有指导性作用,但是未将漏损预警机制和水务公司管理制度进行结合,未来工作将对 DMA 小区漏损检测进行制度和流程上的优化,从而进一步提升水务公司漏损管控的效率。

参考文献

- [1] 李旭,林峰,李允琛,等. 基于组合流量法的城市供水系统小区漏损诊断和预警[J]. 净水技术, 2024, 43(2): 135-142, 183.
LI X, LIN F, LI Y C, et al. Diagnosis and early warning for

- water leakage loss of residential community in urban water supply system based on combined flow method[J]. *Water Purification Technology*, 2024, 43(2):135-142, 183.
- [2] 张坤林, 张强, 金盛, 等. 基于分级分区计量的供水管网漏损预警系统应用[J]. *给水排水*, 2020, 56(10): 115-120.
ZHANG K L, ZHANG Q, JIN S, et al. Application of early warning system for water supply pipe network leakage based on graded metering areas[J]. *Water & Wastewater Engineering*, 2020, 56(10): 115-120.
- [3] 刘晓飞, 刘肃, 何元春, 等. 广州市小区居民夜间合法用水量研究[J]. *中国给水排水*, 2017, 33(9): 45-48, 52.
LIU X F, LIU S, HE Y C, et al. Analysis of residential nighttime water consumption in Guangzhou[J]. *China Water & Wastewater*, 2017, 33(9): 45-48, 52.
- [4] 张文体, 甄非非, 梁伟文. DMA 最小夜间水量的预测与应用[J]. *给水排水*, 2024, 50(s1): 397-403.
ZHANG W T, ZHEN F F, LIANG W W. Prediction and application of minimum night consumption in DMA[J]. *Water & Wastewater Engineering*, 2024, 50(s1): 397-403.
- [5] 李岚, 吴珊, 寇晓霞, 等. 基于独立计量区的夜间最小流量的研究进展[J]. *给水排水*, 2018, 54(6): 135-141.
LI L, WU S, KOU X X, et al. Research progress of minimum night flow based on district metering area[J]. *Water & Wastewater Engineering*, 2018, 54(6): 135-141.
- [6] 高云, 王兴双, 周柯, 等. MNF 法在 DMA 小区漏损管理中的综合分析运用[J]. *供水技术*, 2022, 16(6): 36-38.
GAO Y, WANG X S, ZHOU K, et al. Comprehensive analysis and application of MNF method in DMA community leakage management[J]. *Water Technology*, 2022, 16(6): 36-38.
- [7] 倪世海. 基于大数据分析在黄山市供水管网漏损精细化管理的实践[J]. *中国建设信息化*, 2023(11): 75-79.
NI S H. Practice of fine leakage management of water supply network in Huangshan City based on big data analysis[J]. *Informatization of China Construction*, 2023(11): 75-79.
- [8] 王彤, 李钟毓, 康炳卿, 等. 基于管网分区流量数据的漏损检测方法研究[J]. *水电能源科学*, 2023, 41(7): 127-131.
WANG T, LI Z Y, KANG B Q, et al. Research on leakage detection based on flow data of pipeline network[J]. *Water Resources and Power*, 2023, 41(7): 127-131.
- [9] 冉雨晴, 吴玮, 狄鑫. 基于遗传算法优化 BP 神经网络的管网漏失定位模型研究[J]. *水电能源科学*, 2021, 39(5): 123-126, 122.
RAN Y Q, WU W, DI X. Study on leakage location model of water supply network based on BP neural network optimized by genetic algorithm[J]. *Water Resources and Power*, 2021, 39(5): 123-126, 122.
- [10] 王积建. 供水管网漏损和定位的实时三级预警模型[J]. *科技通报*, 2023, 39(11): 68-72, 94.
WANG J J. Real-time three-level warning model for leakage and location of water supply network[J]. *Bulletin of Science and Technology*, 2023, 39(11): 68-72, 94.
- [11] MCMILLAN L, FAYAZ J, VARGA L. Domain-informed variational neural networks and support vector machines based leakage detection framework to augment self-healing in water distribution networks[J]. *Water Research*, 2024, 249: 120983. DOI: 10.1016/j.watres.2023.120983.
- [12] LI D, QIU S, YANG S. An abnormal detection and early warning system for sugarcane diseases based on unsupervised learning[C]//IEEE. 2023 5th International Conference on Applied Machine Learning (ICAML), 2023: 231-235.
- [13] XU H, ZHANG L, MA L, et al. Analysis and research on electric heating risk early warning based on embedded feature selection and DBSCAN adaptive clustering[C]//IEEE. 2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT), 2022: 1-8.
- [14] 高金良, 朱小玉, 郭文娟, 等. 应用 KS 聚类算法处理 DMA 分区流量数据[J]. *给水排水*, 2021, 47(12): 135-140.
GAO J L, ZHU X Y, GUO W J, et al. Apply KS clustering algorithm to process traffic data of DMA partition[J]. *Water & Wastewater Engineering*, 2021, 47(12): 135-140.
- [15] PARK S, HA J H. Development of computational algorithms for daily water leak detection in district metered areas based on the principal component analysis[J]. *Desalination and Water Treatment*, 2021, 227: 93 - 103. DOI: 10.5004/dwt.2021.27354.
- [16] 淦文燕, 李德毅, 王建民. 一种基于数据场的层次聚类方法[J]. *电子学报*, 2006(2): 258-262.
GAN W Y, LI D Y, WANG J M. An hierarchical clustering method based on data fields[J]. *Acta Electronica Sinica*, 2006(2): 258-262.