

DOI:10.19853/j.zgjsps.1000-4602.2021.07.008

# 一种供水管网泄漏区域定位的机器学习方法

陈京钰, 肖诗云, 冯 新

(大连理工大学 建设工程学部, 辽宁 大连 116024)

**摘 要:** 当采用模式识别方法识别供水管网的泄漏时,如果将每一个节点作为分类器的一个标签,由于供水管网节点泄漏特征的相似性,模型训练的准确率会比较低,因此可以通过聚类泄漏特征相似的节点形成区域,以每个区域作为分类器的标签从而提高模型训练的准确率。提出了一种基于随机森林分类器的二分迭代法识别泄漏区域,根据上一级分类器识别的泄漏区域的节点泄漏变化矩阵,采用  $k$ -means 聚类将上一级迭代识别的泄漏区域聚类为两类(包含泄漏节点的区域和不包含泄漏节点的区域),从而识别包含泄漏节点的区域。随着候选泄漏区域的缩小,对识别泄漏区域有帮助的测点数量也逐渐减少,因此采用平均准确率减少(MDA)进行分类器特征(所需测点)的选择,在保证识别准确率不变的情况下减少分类器训练时所需的特征。与直接进行区域分块的识别方法相比,二分迭代法降低了选择区域分块数目时的盲目性,对于泄漏区域的识别更有目的性,提高了泄漏区域识别的准确率和效率。

**关键词:** 泄漏区域识别; 二分迭代法;  $k$ -means; 随机森林分类器; 特征选择; 平均准确率减少

中图分类号: TU990.3 文献标识码: A 文章编号: 1000-4602(2021)07-0058-08

## A Machine Learning Method for Leakage Localization of Water Distribution Network

CHEN Jing-yu, XIAO Shi-yun, FENG Xin

(Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024, China)

**Abstract:** When pattern recognition is used to identify the leakage of water distribution network, due to the similarity of leakage characteristics, the accuracy of model training will be relatively low if each node is used as the category label of the classifier. Therefore, nodes with similar leakage characteristics can be clustered to form regions, and each region can be used as a category label of the classifier to improve the accuracy of model training. A method called binary iteration based on random forest classifier was proposed to identify the leakage zone. According to the leakage matrix of the leakage zone identified by the last iteration,  $k$ -means clustering was used to cluster the nodes into two kinds (zones that contained leaks and zones that did not contain leaks), so as to identify the region containing the leaks. As the candidate leakage zone decreased, the number of sensors useful to identify the leakage also decreased, so mean decrease accuracy (MDA) was used to select the features required by the classifier, so as to reduce the features required by the classifier training under the condition that the

基金项目: 国家重点研发计划项目(2016YFC0802400); 中央高校基本科研业务费重点实验室专题(DUT20LAB133)  
通信作者: 冯新 E-mail: fengxin@dlut.edu.cn

identification accuracy was constant. Compared with the method that the leakage zone was identified directly, binary iteration method could reduce the blindness of selecting the division number, and it was more specific for the identification of leakage zone, which improved the accuracy and efficiency of the identification of the leakage zone.

**Key words:** leakage zone identification; binary iteration method;  $k$ -means; random forest classifier; feature selection; mean decrease accuracy (MDA)

供水系统是社会的重要基础设施,在经济发展和正常生活中占有重要地位。由于供水管网的老化及设计的不合理,供水系统中存在不同程度的水泄漏问题。发达国家每年约有15%的净化水从水分分配系统流失,而发展中国家平均有35%甚至高达60%的净化水流失,因此对供水管网泄漏进行精准定位是控制供水管网系统漏失率的关键之一。

随着智能化和数据时代的到来,数据挖掘技术在供水管网泄漏定位中的应用逐渐成为研究热点,其可以归纳为两种类型:①根据区域泄漏特征的泄漏区域定位方法;②根据邻近节点泄漏特征相似性的泄漏节点近似定位方法。

节点发生泄漏时监测点的变化有明显的区域特征,因此可以通过泄漏前后监测点区域特征的变化定位泄漏区域。Costanzo等<sup>[1]</sup>根据摩阻系数的不同将供水管网分为不同的区域,对泄漏后的供水管网摩阻系数进行校正,采用贝叶斯方法分析泄漏前后摩阻系数的变化从而定位泄漏区域。基于候选泄漏节点的方法,Tao等<sup>[2]</sup>采用人工免疫算法识别出包含真实泄漏节点的泄漏节点群;Moser等<sup>[3]</sup>采用伪造错误域模型的方法识别出泄漏节点群。根据压力测点特征的相似性,刘畅等<sup>[4]</sup>利用多元统计学上的聚类分析方法,对供水管网故障模拟的数据进行聚类分析,从而对泄漏情况进行故障定位,为以后基于虚拟分区定位泄漏节点奠定了基础,随后,Zhang等<sup>[5]</sup>根据 $k$ -means聚类方法预先对大型供水管网进行虚拟分区,采用多类支持向量机进行泄漏区域的识别,并考虑了监测点布置、核函数的选择和参数优化对识别准确率的影响。

地理位置相邻较近的节点发生泄漏时,测点的变化特征会很相似,因此可以通过定位周围相似节点进而在其周围搜索真实泄漏节点。梁建文等<sup>[6]</sup>通过监测泄漏前后供水管网的水压变化,利用人工神经网络技术来推测其他所有未监测节点的水压变化,从而诊断泄漏位置和泄漏水量;王力等<sup>[7]</sup>以供

水管网压力、流量参数形成的时间序列数据为基础,应用支持向量机方法,通过压力梯度法实现漏点的准确定位。通过与延时模式相结合的方法,Soldevila等<sup>[8]</sup>将贝叶斯分类器应用于供水管网泄漏节点的定位,并且定位精度随时间的增加而增加;基于多级识别方法,Caputo等<sup>[9]</sup>提出了一种由区域到泄漏节点的泄漏节点定位方法,多层感知神经网络分为两级,第一级神经网络用于识别发生泄漏的区域,第二级神经网络用于估计漏损的大小和位置。

通过以上分析发现,以往均是采用聚类方法直接将供水管网分为不同的区域数目从而进行基于模式识别的泄漏区域的识别,具有分区(聚类)数目的盲目性,并且每次分区数目试算时并没有充分利用上一次试算的结果,使本次的试算模型训练中存在较多的冗余非泄漏区域,对于泄漏区域的识别是非必要的。因此,本研究提出了一种基于随机森林分类器的二分迭代法进行泄漏区域的识别,二分类降低了分区数目选择时的盲目性,通过随机森林分类器逐级剔除不包含泄漏节点的区域,从而进行泄漏区域的逐级识别;随着泄漏区域的逐级缩小,所需的对识别泄漏区域有帮助的测点数量将同步减少,因此将特征选择加入了模型的训练,从而在不影响识别准确率的情况下进行了测点数量的缩减,进一步减少了模型训练所需的样本。

## 1 泄漏区域定位方法

本研究从供水管网的节点合并、泄漏区域的二分类、特征选择及泄漏区域识别四个方面进行方法的介绍。①建立供水管网水力模型(采用EPANET软件),通过在节点添加泄漏系数从而采集两类节点泄漏数据集,节点的泄漏变化矩阵、用于训练随机森林分类器<sup>[10]</sup>的数据样本;②对于泄漏变化矩阵相似的节点,可以进行节点的预先合并,从而在不影响后续聚类的基础上,缩小供水管网所需产生样本的节点数量;③在缩小候选泄漏区域的同时,根据平均准确率减少(MDA)选择对候选泄漏区域有识别作

用的测点,从而在保证识别准确率不变的情况下减少所需测点;④根据上一级迭代识别的泄漏区域的泄漏变化矩阵,通过  $k$ -means<sup>[11]</sup>将上一级迭代识别的泄漏区域聚类为两类,结合随机森林分类器进行泄漏区域的识别,从而逐级缩小候选泄漏区域。本研究以模型训练级联识别准确率(每一次迭代的模型训练的准确率乘积)  $< 95\%$  为二分迭代识别结束的标准,泄漏区域识别的模型训练流程如图1所示。

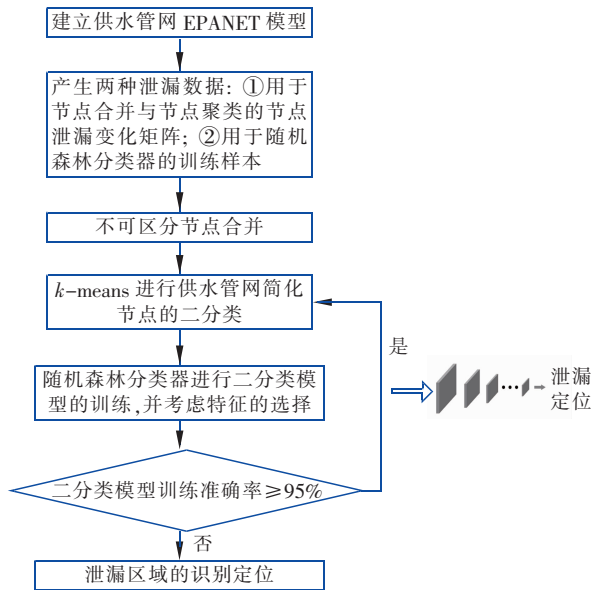


图1 泄漏区域定位的模型训练流程

Fig. 1 Flow chart of model training for leakage zone localization

$$\Delta s = \begin{bmatrix} \Delta s_1 \\ \Delta s_2 \\ \vdots \\ \Delta s_n \end{bmatrix} = \begin{bmatrix} \Delta P_{11} & \Delta P_{12} & \cdots & \Delta P_{1x} & \Delta Q_{11} & \Delta Q_{12} & \cdots & \Delta Q_{1y} \\ \Delta P_{21} & \Delta P_{22} & \cdots & \Delta P_{2x} & \Delta Q_{21} & \Delta Q_{22} & \cdots & \Delta Q_{2y} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \Delta P_{n1} & \Delta P_{n2} & \cdots & \Delta P_{nx} & \Delta Q_{n1} & \Delta Q_{n2} & \cdots & \Delta Q_{ny} \end{bmatrix} \quad (2)$$

### 1.3 节点合并

当供水管网发生泄漏时,由于供水管网拓扑结构的复杂性及测点数量的有限性,会存在一些泄漏特征非常相似的节点,因此可以预先进行节点的合并,从而减少供水管网模型训练时所需的产生训练样本的节点数量。本研究设定供水管网压力测点的精度为  $0.1 \text{ kPa}$ ,流量测点的精度为  $0.01 \text{ L/s}$ ,根据节点泄漏变化矩阵,若节点  $i$  和节点  $j$  ( $1 \leq i \leq n, 1 \leq j \leq n, i \neq j$ ) 满足:

$$\begin{cases} |\Delta P_{im} - \Delta P_{jm}| < 0.1 \text{ kPa} & (1 \leq m \leq x) \\ |\Delta Q_{ir} - \Delta Q_{jr}| < 0.01 \text{ L/s} & (1 \leq r \leq y) \end{cases} \quad (3)$$

### 1.1 泄漏模型

供水管网的节点泄漏量与压力相关,压力越大则漏失量也越大。因此可以在 EPANET 中使用喷射系数来生成泄漏事件。泄漏量可以表示成节点压力的幂函数:

$$Q = YP^\alpha \quad (1)$$

式中:  $Q$  为节点泄漏量;  $Y$  为节点泄漏系数;  $P$  为节点压力;  $\alpha$  为压力灵敏度指数,指数的选择与管材有关,本研究假设管道是金属材质,因此  $\alpha$  取  $0.5$ <sup>[12]</sup>。

### 1.2 泄漏变化矩阵

假设供水管网有  $n$  个节点、 $x$  个压力计、 $y$  个流量计。当供水管网节点未发生泄漏时,正常的压力监测点的值与流量监测点的值为  $s = [P_1^0, P_2^0, \cdots, P_x^0, Q_1^0, Q_2^0, \cdots, Q_y^0]$ , 节点  $i$  发生泄漏时的压力值与流量值为  $s_i = [P_{i1}, P_{i2}, \cdots, P_{ix}, Q_{i1}, Q_{i2}, \cdots, Q_{iy}]$ , 假设  $P_u^0$  和  $Q_v^0$  分别为供水管网正常运行状态下的压力监测点  $u$  及流量监测点  $v$  的值,  $P_{iu}$  和  $Q_{iv}$  为节点  $i$  发生泄漏时的压力监测点  $u$  及流量监测点  $v$  的值,则节点  $i$  泄漏前后监测点的值的变化为  $\Delta s_i = [P_{i1} - P_1^0, P_{i2} - P_2^0, \cdots, P_{ix} - P_x^0, Q_{i1} - Q_1^0, Q_{i2} - Q_2^0, \cdots, Q_{iy} - Q_y^0]$ ; 对每个节点依次添加相同的泄漏系数  $Y$ , 从而得到供水管网的节点泄漏变化矩阵为  $\Delta s = \{\Delta s_1, \Delta s_2, \cdots, \Delta s_n\}^T$ , 如果对于泄漏节点  $i$ ,  $\Delta P_{iu} = P_{iu} - P_u^0$ 、 $\Delta Q_{iv} = Q_{iv} - Q_v^0$ , 那么供水管网全部节点的泄漏变化矩阵可以表示为:

式中:  $\Delta P_m$  是指节点泄漏系数为  $Y$  时,压力测点  $m$  处压力变化值;  $\Delta Q_r$  是指节点泄漏系数为  $Y$  时,流量测点  $r$  处的流量变化值; 则将节点  $i$  和节点  $j$  合并为一组,模型训练时只取本组内的一个节点产生训练样本。

### 1.4 $k$ -means

供水管网节点发生泄漏时会有明显的区域特征,因此可以根据节点泄漏时测点特征的相似性,采用  $k$ -means 进行供水管网的区域分块,将每一个区域作为分类器的一个标签进行模型训练,从而提升分类器的模型训练准确率。 $k$ -means 根据距离相



似度将供水管网的  $n$  个节点划分为  $k$  ( $k \leq n$ ) 个类别,应用简单、聚类速度快且效果好。聚类结束的目标是最小化平方误差  $E$ ,如式(4)所示,使  $k$  个类别满足类内距离最小且类间距离最大,即类内相似度最大且类间相似度最小。此外,由于流量与压力的量纲不同,变化相差较大,因此聚类之前需要进行泄漏变化矩阵的归一化处理,消除由于量纲不同而引起的聚类误差。本研究采用 0 均值标准化进行泄漏变化矩阵的数据归一化。假设将供水管网的  $n$  个节点划分为  $C_1, C_2, \dots, C_k$  类,随机选择  $k$  个节点的泄漏特征作为初始的  $k$  个质心向量,则初始平方误差  $E_0$  为:

$$E_0 = \sum_{i=1}^k \sum_{s_r \in C_i} \|s_r - u_j\|_2^2 \quad (4)$$

式中:  $u_j$  为类别  $C_i$  的质心,  $u_j = \frac{1}{|C_i|} \sum_{s_r \in C_i} s_r$ ,  $|C_i|$  为类别  $C_i$  中节点的个数。

随后通过迭代的方法更新质心并计算平方误差  $E$ ,假设迭代  $N$  次,满足  $E_N - E_{N-1} < e'$  ( $e'$  为迭代收敛值)或者质心不再更新变化(达到最大的迭代次数)时迭代结束,输出  $C = \{C_1, C_2, \dots, C_k\}$ ,即为供水管网节点聚为  $k$  类的结果,本研究取  $k=2$ 。

### 1.5 随机森林

随机森林是基于大量决策树的集成学习技术<sup>[13-14]</sup>,决策树分为分类树和回归树,由于本研究是对泄漏区域的识别定位,因此选择分类树。随机森林分类器的算法流程如图2所示。

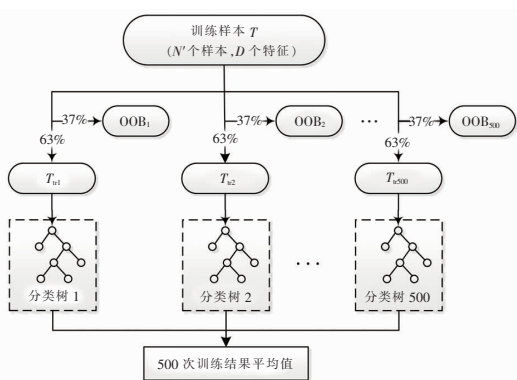


图2 随机森林分类器算法流程

Fig.2 Flow chart of random forest classifier

① 假设供水管网产生的泄漏样本为  $T$ , 有  $N'$  个样本,  $x + y = D$  个监测点特征;

② 设定随机森林中有  $n'$  个子训练集  $T_{tr1}, T_{tr2}, \dots, T_{trn'}$ , 也即有  $n'$  棵决策树  $tree_1, tree_2, \dots, tree_{n'}$ ; 对

于每个子训练集  $T_{tri}$ , 采用自举采样法 (bagging) 从训练样本  $T$  中有放回地随机选出新的子训练集, 每个子训练集  $T_{tri}$  和训练样本  $T$  的大小一致, 即对于每个子训练集随机抽样  $n'$  次生成  $n'$  个随机样本, 因此每个子训练集  $T_{tri}$  中会有重复样本。对于训练样本  $T$  中的某个样本,  $n'$  次均没有被抽到的概率为  $(1 - 1/n')^{n'}$ , 若  $n' \sim \infty$ , 那么  $(1 - 1/n')^{n'} \approx 1 - 1/e \approx 37\%$ , 也即训练样本  $T$  中约有 37% 的样本不是每次都能被抽到, 这一部分未被抽到的样本称为袋外数据 (OOB), 用于评价每棵树的准确率, 如图2所示。由于随机森林分类器的精度随着决策树量的增大而逐渐增大并趋于极限, 因此本研究选取默认值  $n' = 500$ <sup>[14]</sup>。

③ 对于每个子训练集  $T_{tri}$  的节点, 随机地从  $D$  个特征中抽取  $m$  ( $m \leq D$ ) 个子特征用于创建树, 本研究的  $m$  取默认值  $\sqrt{D}$ , 计算每个特征的基尼指数 (Gini Index), 如式(5)所示。给定训练子集  $T_{tri}$  和连续特征  $d$ , 其中,  $d = 1, 2, \dots, m$ , 训练子集  $T_{tri}$  有  $c$  类样本, 第  $k$  类有  $|c_k|$  个样本, 特征  $d$  在  $T_{tri}$  上出现了  $a$  个不同的取值, 把这些值由小到大进行排序, 记为  $\{d^1, d^2, \dots, d^a\}$ , 基于划分点  $t [t = (d^i + d^{i+1})/2, 1 \leq i \leq a - 1]$  可将  $d$  分为两个子集  $d_i^-$  和  $d_i^+$ , 其中  $d_i^-$  是包含那些在特征  $d$  上取值不大于  $t$  的样本,  $d_i^+$  是包含那些在特征  $d$  上取值大于  $t$  的样本, 对相邻的特征取值  $d^i$  与  $d^{i+1}$ , 划分点  $t$  在区间  $[d^i, d^{i+1})$  中任意值所产生的划分结果相同, 因此有  $a - 1$  个划分点作为候选划分点, 子训练集  $T_{tri}$  及特征  $d$  在划分节点  $t$  处的基尼系数为:

$$\begin{cases} \text{Gini}(T_{tri}) = \sum_{k=1}^c p_k (1 - p_k) = 1 - \sum_{k=1}^c p_k^2 \\ \text{Gini}(T_{tri,d}) = \frac{|T_{tri,d}^-|}{|T_{tri,d}|} \text{Gini}(T_{tri,d_i^-}) + \frac{|T_{tri,d}^+|}{|T_{tri,d}|} \text{Gini}(T_{tri,d_i^+}) \end{cases} \quad (5)$$

式中:  $p_k$  代表样本属于第  $k$  类的概率。基尼指数  $\text{Gini}(T_{tri,d})$  表示在样本集合  $T_{tri,d}$  中一个随机选中的样本被错分的概率, 基尼指数值越大, 样本被错分的概率越大。选择基尼指数最小的  $t$  值及其对应的特征进行分裂, 之后再构造各个分支重复上述流程。

④ 每一棵分类树均贡献了一个子训练集识别准确率。求得 500 棵树的平均识别准确率即为训练模型的最终准确率。由于本研究是采用多级二分类进行识别分类, 因此泄漏区域最终定位的识别准确

率为级联分类器训练准确率的乘积。

随机森林分类器决策树的并行运算加快了训练速度,且不需要选择核函数<sup>[6]</sup>及假设先验知识<sup>[7]</sup>,随机产生  $n'$  个训练子集和随机选择  $m$  个子特征的随机性保证了其能防止过拟合,该方法已被成功应用到计算机视觉识别<sup>[15]</sup>及故障诊断<sup>[16]</sup>等实际领域,并取得了不错的效果。

### 1.6 监测维度的选择

当采用二分迭代法进行泄漏区域识别时,随着迭代次数的增加,泄漏区域也逐渐缩减,对泄漏区域有识别效果的测点也在逐步减少,因此可以在不影响识别准确率的情况下通过特征选择的方式进行测点的选择。

本研究采用随机森林分类器的平均准确率减少(MDA)<sup>[17]</sup>进行特征重要性排序,进而进行特征的选择。计算方法如下:

① 对随机森林中的任一分类树  $i$  ( $1 \leq i \leq 500$ ),使用相应的 OOB(袋外)数据计算其袋外数据误差,记为  $\text{errOOB}_{i1}$ ;

② 依次随机打乱袋外数据样本在特征  $d$  处的值,再次计算它的袋外数据误差,记为  $\text{errOOB}_{i2}$ ;

③ 假设随机森林中有 500 颗树,那么特征  $d$  的重要性为  $\text{MDA} = \sum_{i=1}^{500} \frac{\text{errOOB}_{i1} - \text{errOOB}_{i2}}{500}$ 。

MDA 越大则特征越重要,通过计算 MDA 的值从而进行特征的重要性排序,根据特征重要性由大到小进行特征个数的增加,当增加至训练模型的识别准确率不变时即为本次迭代所需的传感器测点,也即本次的监测维度。

## 2 供水管网泄漏区域定位仿真算例

本研究通过示例模型,从节点合并、泄漏区域的二分类、特征选择及泄漏区域识别四个方面进行了方法的分析说明。

### 2.1 供水管网模型简介

采用 EPANET 建立供水管网水力模型,如图 3 所示。该供水管网由 1 个水库、1 个蓄水池、375 个节点和 469 个管段组成,布设 9 个压力计和 3 个流量计。基本需水量为 148 L/s,最大需水量为 162.8 L/s,最小需水量为 118.4 L/s。本研究假设泄漏时只发生一个泄漏节点。由于泄漏量太小不会引起传感器的变化、泄漏量太大会让居民首先发现,因此本示例选取的泄漏系数在 0.5 ~ 1.0 之间。

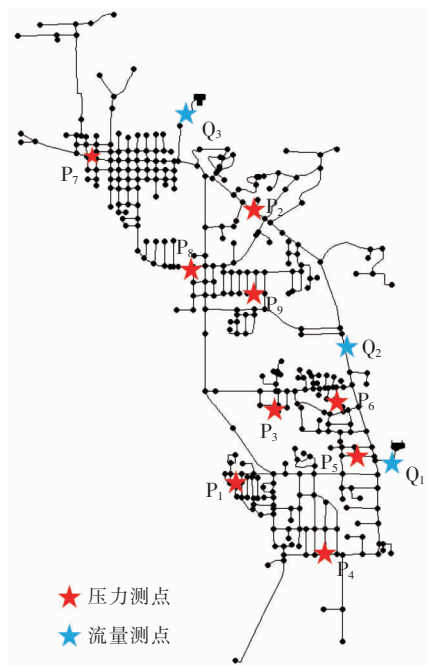


图3 示例供水管网

Fig. 3 Layout of example water distribution system

### 2.2 节点合并

取泄漏系数  $Y=1.0$  产生供水管网的泄漏变化矩阵,根据泄漏变化矩阵对示例供水管网进行不可区分节点的合并。示例供水管网节点的合并结果如图 4 所示,不可区分节点用红色点标出。可知,节点合并之后,示例供水管网由 375 个节点转换为 280 个节点组,合并的节点均位于供水管网的支端及节点较密集的区域。



图4 示例供水管网简化结果

Fig. 4 Layout of simplified example water distribution system

### 2.3 泄漏区域的二分迭代识别

以两个单独发生的泄漏节点为例进行所述识别方法的详细说明。假设示例供水管网的泄漏节点发



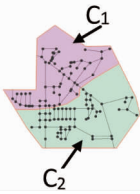
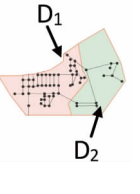
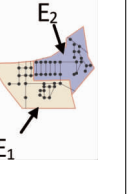

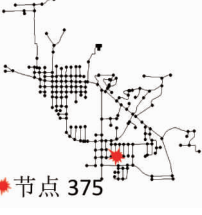




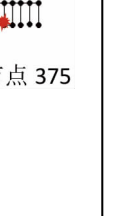




下减少了所需的特征,从而大大减少训练时所需的样本容量。

表2 二分迭代法识别泄漏节点375所在区域的过程

Tab.2 Identification of zone containing leakage node 375 by binary iteration method

迭代级数	1	2	3	4	5	6
聚类结果						
聚类过程	$A_1, A_2$	$A_1 \rightarrow (B_1, B_2)$	$B_2 \rightarrow (C_1, C_2)$	$C_2 \rightarrow (D_1, D_2)$	$D_1 \rightarrow (E_1, E_2)$	$E_2 \rightarrow (F_1, F_2)$
识别的泄漏区域						
特征选择	$Q_2$	$P_7, Q_3$	$P_2, P_7, P_9, Q_2$	$P_9, Q_1, Q_2$	$P_9, Q_1, Q_2$	$P_9, Q_1, Q_2$

两种情况的泄漏识别结果如图5所示。

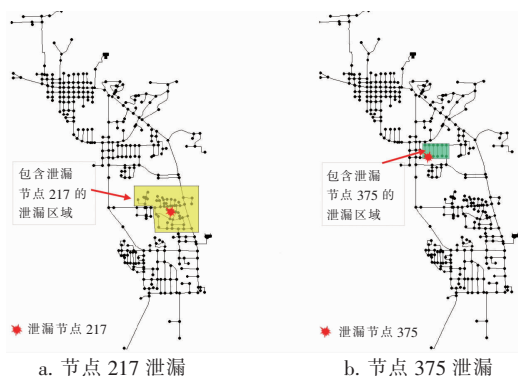


图5 泄漏区域识别结果

Fig.5 Leakage zone identification by binary iteration method

### 3 结论

本研究基于模式识别的二分迭代法对泄漏区域进行识别:通过合并泄漏特征相似的节点,减少了供水管网进行泄漏区域识别时所需节点样本量;随后通过二分迭代法进行了泄漏区域的逐步缩减定位,并且在二分类识别时进行了测点的选择,从而在保证识别准确率不变的情况下减少了所需测点量。与直接进行区域分块识别方法相比,二分迭代法降低了选择区域分块数目(聚类数目)的盲目性,减少了模型训练的试算次数,提高了泄漏区域识别的准确率和效率。

本研究假设供水管网只发生单节点泄漏,但实际供水管网往往存在多节点泄漏,因此迫切需要提出一种基于模式识别的多节点泄漏的识别方法。传感器的数目及泄漏系数的取值范围对模式识别的结果均会有影响,因此在未来的研究中需要讨论传感器数目及泄漏量的大小对泄漏识别结果的影响。

### 参考文献:

- [1] COSTANZO F, FIORINI MOROSINI A, VELTRI P, et al. Model calibration as a tool for leakage identification in WDS: a real case study[J]. Procedia Engineering, 2014, 89: 672-678.
  - [2] TAO T, HUANG H D, LI F, et al. Burst detection using an artificial immune network in water-distribution system[J]. Journal of Water Resources Planning and Management, 2014, 140(10): 04014027.
  - [3] MOSER G, PAAL S G, SMITH I F. Leak detection of water supply networks using error-domain model falsification[J]. Journal of Computing in Civil Engineering, 2017, 32(2): 04017077.
  - [4] 刘畅, 刘志强, 郝红海. 聚类分析在管网故障点定位中的应用[J]. 安徽农业科学, 2007, 35(7): 2177-2178.
- LIU Chang, LIU Zhiqiang, HAO Honghai. Application of fault location of cluster analysis in pipeline network[J].

- Journal of Anhui Agricultural Sciences, 2007, 35 (7): 2177 – 2178 (in Chinese).
- [5] ZHANG Q Z, WU Z Y, ZHAO M J, *et al.* Leakage zone identification in large-scale water distribution systems using multiclass support vector machines[J]. Journal of Water Resources Planning and Management, 2016, 142 (11): 04016042.
- [6] 梁建文, 肖笛, 张宏伟, 等. 给水管网故障实时诊断方法(II)[J]. 自然科学进展, 2002, 12(9): 970 – 974. LIANG Jianwen, XIAO Di, ZHANG Hongwei, *et al.* On-line fault diagnosis of water distribution systems (II) [J]. Progress in Natural Science, 2002, 12(9): 970 – 974 (in Chinese).
- [7] 王力, 刘光尧, 曾佩佩. 基于 SVM 的机场供水管网泄漏辨识定位系统研究[J]. 水资源与水工程报, 2014, 25(1): 38 – 41. WANG Li, LIU Guangyao, ZENG Peipei. Study on position system of leakage identification for water supply network at airport based on SVM[J]. Journal of Water Resources and Water Engineering, 2014, 25(1): 38 – 41 (in Chinese).
- [8] SOLDEVILA A, FERNANDEZ-CANTI R M, BLES A J, *et al.* Leak localization in water distribution networks using Bayesian classifiers [J]. Journal of Process Control, 2017, 55: 1 – 9.
- [9] CAPUTO A C, PELAGAGGE P M. An inverse approach for piping networks monitoring [J]. Journal of Loss Prevention in the Process Industries, 2002, 15 (6): 497 – 505.
- [10] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5 – 32.
- [11] LAROSE D T. Discovering Knowledge in Data: An Introduction to Data Mining [M]. New Jersey: John Wiley & Sons Inc., 2005.
- [12] FUCHS-HANUSCH D, STEFFELBAUER D, GÜNTHER M, *et al.* Systematic material and crack type specific pipe burst outflow simulations by means of EPANET2 [J]. Urban Water Journal, 2016, 13(2): 108 – 118.
- [13] BREIMAN L, FRIEDMAN J, STONE C J, *et al.* Classification and Regression Trees [M]. New York: CRC Press, 1984.
- [14] KUHN M, JOHNSON K. Applied Predictive Modeling [M]. New York: Springer, 2013.
- [15] 徐岳峰, 周书仁, 王刚, 等. 基于深度图像梯度特征的人体姿态估计[J]. 计算机工程, 2015, 41(12): 200 – 205. XU Yuefeng, ZHOU Shuren, WANG Gang, *et al.* Human body attitude estimation based on gradient feature of depth images [J]. Computer Engineering, 2015, 41 (12): 200 – 205 (in Chinese).
- [16] LI C, SANCHEZ R V, ZURITA G V, *et al.* Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals [J]. Mechanical Systems and Signal Processing, 2016, 76: 283 – 293.
- [17] 王学渊, 陈志刚, 钟新荣, 等. 基于随机森林的管网漏失诊断方法[J]. 计算机应用, 2018, 38 (S1): 20 – 23. WANG Xueyuan, CHEN Zhigang, ZHONG Xinrong, *et al.* Pipeline network leakage diagnosis based on multi-source random forest fusion [J]. Journal of Computer Applications, 2018, 38 (S1): 20 – 23 (in Chinese).

作者简介: 陈京钰 (1989 – ), 女, 河南南阳人, 博士研究生, 研究方向为供水管网的泄漏监测。

E-mail: 875965143@qq.com

收稿日期: 2019 – 09 – 29

修回日期: 2020 – 04 – 16

(编辑: 任莹莹)