

DOI:10.19853/j.zgjsps.1000-4602.2022.23.005

使用迭代分区识别算法的供水管网异常隔离

简彩, 高金良, 徐勇鹏

(哈尔滨工业大学 环境学院, 黑龙江 哈尔滨 150090)

摘要: 为了降低城市供水管网中异常事件(如爆管、漏失、非法用水等)风险,提出了一种能够及时、有效隔离供水管网异常事件的方法,该方法包括“分区”和“识别”两个步骤。首先,结合图论技术、聚类方法和人工微调将管网分为两个子区域,然后采用卷积神经网络算法、水力模拟判定异常事件的具体发生区域。分区和识别过程迭代进行,直到满足预设停止条件,使异常事件隔离在尽可能小的区域。为了提高计算效率,在水力模拟过程中使用了子组技术,同时在训练卷积神经网络算法中评估并选用最重要监测数据向量特征。案例分析结果表明,所提出的方法能够可靠、高效地隔离管网异常,在准确性及时效性方面均具有较好的表现。

关键词: 供水管网; 异常隔离; 迭代分区; 异常识别; 卷积神经网络

中图分类号: TU991 **文献标识码:** A **文章编号:** 1000-4602(2022)23-0031-07

Iterative Partition Identification Algorithm for Anomaly Isolation of Water Distribution Network

JIAN Cai, GAO Jin-liang, XU Yong-peng

(School of Environment, Harbin Institute of Technology, Harbin 150090, China)

Abstract: To minimize abnormal event risks in urban water distribution network (such as pipe burst, leakage and illegal water use), a method for timely and effective isolation of abnormal events was proposed, which included two steps: partition and identification. First, the pipe network was divided into two sub-regions by combination graph theory technology, clustering method and manual fine tuning. Then, the specific occurrence area of abnormal events was determined by convolutional neural network algorithm and hydraulic simulation. The partitioning and identification process iterated until preset stop conditions were met, so as to isolate the abnormal event in the smallest possible area. To improve the computational efficiency, the subgroup technique was used in the hydraulic simulation process, and the most important monitoring data vector features were evaluated and selected in the training convolutional neural network algorithm. The results of case analysis showed that the proposed method was capable of isolating network anomalies reliably and efficiently, and had a good performance in terms of accuracy and timeliness.

Key words: water distribution network; anomaly isolation; iterative partition; anomaly identification; convolutional neural network

基金项目: 国家重点研发计划项目(2018YFC0406200); 国家自然科学基金资助项目(51778178、51978203); 黑龙江省自然科学基金资助联合引导项目(LH2019E044)

通信作者: 简彩 E-mail: jiancheer@163.com

管网老化、施工质量不佳或缺乏维护导致城市供水管网各类异常事件频发,包括爆管、漏失、非法用水等。管网中一旦发生异常,相关人员需要及时检测、隔离异常区域,并查明现场位置,采取适当的措施来消除异常。针对供水管网异常事件隔离,目前开发的相关技术可以分为基于“硬件”与“软件”两类^[1-3],前者包括声学检测、热成像红外检测、探地雷达和遥感技术等^[4]。这类方法大多依赖于昂贵的设备且劳动强度高,不适用于大型供水管网。基于软件的方法分为基于瞬态流、基于模型与基于数据三类^[5]。基于瞬态流方法需要安装高频传感器设备,设施成本投入较高。基于模型方法能检测和定位异常,但需要提前校准水力模型。基于数据驱动方法需要收集大量数据,该过程可能会面临数据丢失、异常以及噪声等问题。因此,如何高效实现供水管网异常事件隔离值得进一步研究。

笔者提出了异常隔离定位方案,包括“分区”和“识别”两个步骤。首先,结合图论技术、聚类方法和人工微调将管网分为两个子区域;然后,采用卷积神经网络算法、水力模拟判定异常事件的发生区域。分区和识别过程迭代进行,直到满足预设停止条件,使异常事件隔离在尽可能小的区域。

1 供水管网的分区方法

本研究的总体实施步骤如图1所示。

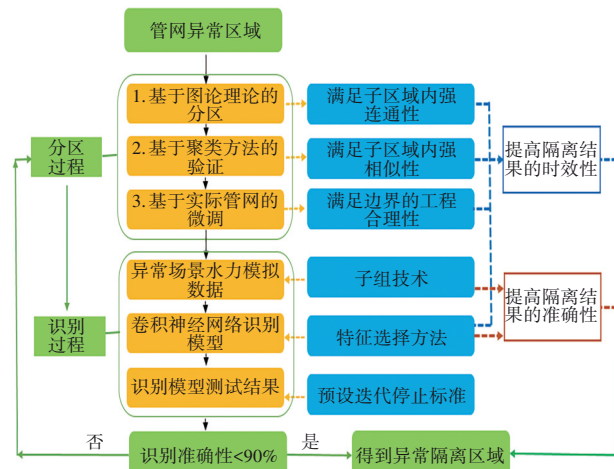


图1 方案实施步骤

Fig.1 Implementation steps of the proposed scheme

采用图论—聚类—人工微调的分区方法:第一步,基于图论方法二分区,以保证子区域内的强连通性;第二步,基于聚类方法验证第一步得到的两个子集,保证子区内强相似性;第三步,根据实际

供水管网的拓扑结构和水力特征进行手动微调,提高分区结果的合理性。

1.1 基于图论技术的划分

对供水管网的划分采用了基于经典模块化方法的图论技术,包括:①将供水管网区域映射成一个无向图。其中,无向图的顶点代表网络节点、水库和水塔,边代表管道、泵和阀门。②基于模块化指数将无向图划分为具备强连通性特征的顶点集群。模块化指数是一种拓扑指数,代表了识别网络结构中集群的可能性。一般来说模块化指数越高,集群识别的效果就越好,其公式为:

$$Q = \frac{1}{2m} \sum_{v\omega} \left[A_{v\omega} - \frac{k_v k_\omega}{2m} \right] \delta(c_v, c_\omega) \quad (1)$$

式中: v, ω 为无向图的两个顶点; $A_{v\omega}$ 为网络图中邻接矩阵的元素; $k_v = \sum_{\omega} A_{v\omega}$ 、 $k_\omega = \sum_v A_{v\omega}$ 分别为连接到顶点 v 、顶点 ω 边数的总和; c_v 为顶点 v 所属的集群,如果 $c_v = c_\omega$,那么 $\delta(c_v, c_\omega) = 1$,否则 $\delta(c_v, c_\omega) = 0$; $m = \frac{1}{2} \sum_{v\omega} A_{v\omega}$,为网络图中边的数目。

以模块化指数最大为目标来二分的管网区域,分别记为子区域A与子区域B。

1.2 基于聚类方法的验证

使用聚类方法验证前一步划分子区域内节点间的强相似性。假设在子区域A中存在 N 个节点、 N_p 个压力传感器和 N_q 个流量传感器。传感器的正常监测值 $S^0 = [P_1^0, P_2^0, \dots, P_{N_p}^0, Q_1^0, Q_2^0, \dots, Q_{N_q}^0]$ 。若某个节点 $i (1 \leq i \leq N)$ 在某一时刻发生异常事件(通过添加一个异常流量来模拟),传感器的异常监测值 $S_i^u = [P_{i1}^u, P_{i2}^u, \dots, P_{iN_p}^u, Q_{i1}^u, Q_{i2}^u, \dots, Q_{iN_q}^u]$, $\Delta S_i = S_i^u - S^0 = [P_{i1}^u - P_1^0, P_{i2}^u - P_2^0, \dots, P_{iN_p}^u - P_{N_p}^0, Q_{i1}^u - Q_1^0, Q_{i2}^u - Q_2^0, \dots, Q_{iN_q}^u - Q_{N_q}^0]$ 则称为传感器的异常向量。将相同异常流量分别添加到子区域A中的每个节点,可得到异常矩阵 $\Delta S_A = \{\Delta S_1, \Delta S_2, \dots, \Delta S_N\}^T$,同理可以得到子区域B的异常矩阵 ΔS_B 。由于流量和压力的单位不同,因此在聚类前对异常矩阵进行零均值归一化处理,以消除因维度差异导致的聚类误差。然后基于k-means聚类方法进行如下操作:①分别计算子区域A和B数据点的样本平均值,作为两个初始中心 E_A 和 E_B ;②对于异常矩阵 ΔS_A 中每个特征向量 ΔS_i ,计算其到聚类中心 E_A 和 E_B 的距离,记为 D_A^i

和 D_B^i ,如果 $D_A^i \leq D_B^i$,则节点 i 保留在子区域A中,否则将其划分到子区域B中,对异常矩阵 ΔS_B 的每个特征向量同样进行上述操作;③对调整后的子区A和B,重新计算其聚类中心,重复步骤①和②,直至聚类中心基本不变,得到最终分区结果。

1.3 基于实际管网特点的微调

为了避免因满足第二步的相似性而破坏了第一步的连通性,导致独立环状网络中的某些节点被单独划分到另一个子区中,或者是同一水源的供水区域被明显切割,需要对管网分区边界进行再次确认,并采取人工微调手段对前两步的分区结果进行改进。人工微调的核心原则是优化分区结果,确保分区满足水力要求和工程的合理性。

2 异常事件的识别方法

假设在持续时段内供水管网中仅有一个异常事件发生,采用水力模型进行模拟并得到相应异常场景下的流量和压力数据,然后将其作为卷积神经网络模型的输入数据。输出数据是类别标签(1、2分别代表子区A和子区B)。最后,训练好的卷积神经网络模型使用实际的监测值来确定异常子区。

2.1 水力模拟与子组技术

利用水力模拟软件EPANET模拟多种异常场景,通过更改异常事件的规模和位置产生相应的训练数据集。真实的供水管网规模庞大,如果在子区域的所有节点上均进行水力模拟,则需要大量的运算和训练时间。如果只在随机选择的部分节点上进行模拟又难以训练出泛化性能良好的模型。为了解决这个难题,本研究采用子组技术:在每个子组中确定一个父节点,在父节点上生成训练数据,并在子组中其他节点上随机复制。例如,如果一个子区域有1 000个节点,将10个相近节点组成一个子组,那么搜索空间的大小将从1 000减少到100。采用子组技术后,模拟数据量会大大减少,训练网络的速度也会加快。子组技术中的子组数量由子区域中的节点总数和训练的需求来确定,邻近节点的确定可以在建立水力模型时同步完成。

2.2 卷积神经网络模型与特征重要性

人工神经网络(ANN)作为近年来应用最为广泛的机器学习算法之一,衍生出许多深度学习算法。卷积神经网络(CNN)是其中的重要代表,具有局部连接、权重共享、池化操作和多层结构等优势。

这些优势使得CNN能够比传统的深度学习算法更高效地学习到数据间的非线性关系,拥有更强的鲁棒性和适应性。本研究同时采用了传统的人工神经网络算法——反向传播神经网络(BP神经网络)建立识别模型,通过对比以验证CNN模型在异常识别应用中的优越性。80%的水力模拟数据用于训练,10%用于验证,10%用于测试。

训练识别模型输入值是模拟异常场景下监测设备的记录值。不过,随着分区迭代的进行,子区域面积及其中的监测设备数量不断减少,所需要用于模型识别的监测数据逐渐降低,相应地,识别模型的输入特征向量选取也应调整。此时如果继续采用整个管网中所有传感器的异常数据进行训练,不仅无法提升模型的性能,还会导致模型过于复杂,出现过拟合等风险。为了避免该问题的发生,需要识别模型的重要输入特征,消除冗余特征。而筛选出的重要输入特征向量也反映了当前模型中对应监测仪表的重要性。本研究采用递归特征消除算法进行特征筛选^[6]。

3 案例管网应用与分析

以C-town供水管网^[7]作为基准网络,验证方案的有效性。管网拓扑结构如图2所示。

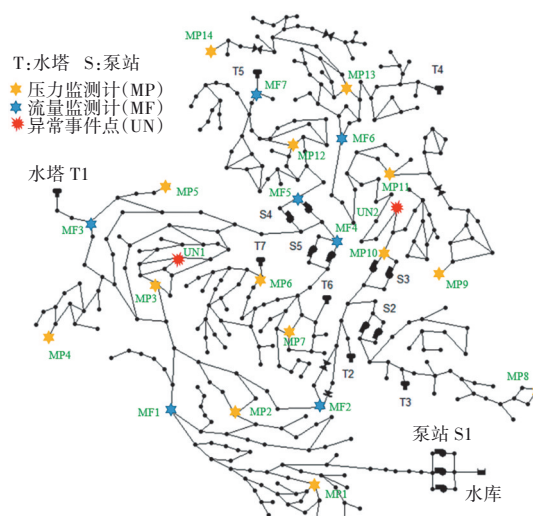


图2 显示监测仪表位置的案例供水管网拓扑示意

Fig.2 Case water distribution network structure showing sensor locations

图2表示了一个虚拟管网,包括1个水库、7个水塔、5个泵站、388个节点、429条管道。假设管网中存在14个压力监测计、7个流量监测计。异常漏失事件1(位置为UN1)的流量为5 L/s。

在分区和识别过程中,为了加快建模速度,基于夜间最小流量进行水力模拟。采用模型训练耗时和对测试样本的识别准确率来评价模型的表现。设定迭代分区的停止条件为模型识别准确率低于90%。第一次分区结果见图3。

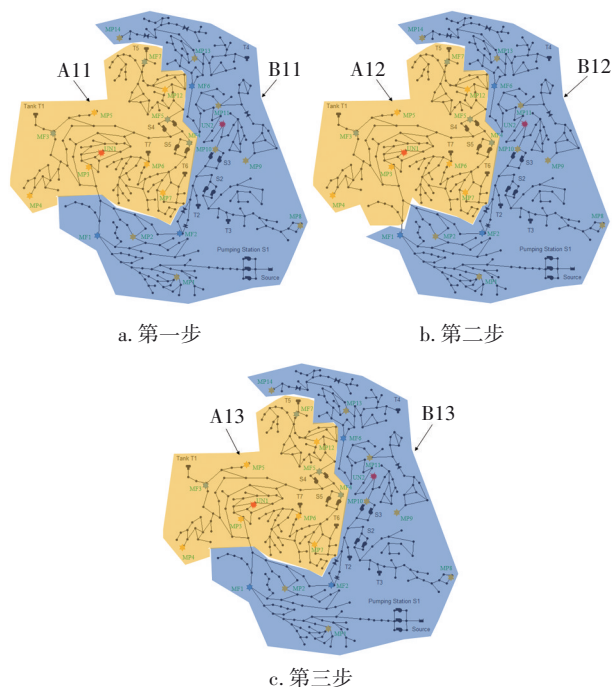


图3 异常事件1的第一次分区结果

Fig.3 The first partition results of abnormal event 1

第一步,对整个管网采用图论技术进行划分,得到两个子区域(A11,B11),如图3(a)所示。第二步,采用聚类方法对这两个子集进行验证,遍历每个子集中的元素,确保其距离本子集的聚类中心更近,否则将其归于另一个子集,最终得到调整后的两个子区域(A12,B12),如图3(b)所示。第三步,必要时采用人工微调手段,在确保分区满足水力要求和工程合理性的基础上进一步优化分区结果。采用三步分区方法得到两个子区域(A13,B13)后,需要识别异常事件发生在哪个子区域。在水力模拟中采用子组技术,子组中包含的节点数目不同,识别模型的表现也不同,结果见表1。

由表1可知,子组包含的节点数目越多,CNN和ANN识别模型的训练耗时越短,即计算速度越快,但是在水力模拟数据测试样本上的识别准确率会下降。不过,即使是在子组包含30个节点的前提下所建立的识别模型,也能够准确地识别出异常子区域为A13,说明所建立的识别模型非常可靠。在建

立CNN识别模型过程中,对监测仪表所提供的数据向量进行特征重要性评估。按照重要性由大到小的顺序,训练CNN模型时最重要的10个向量特征所对应的仪表是:MF5、MF4、MP3、MF1、MP10、MF2、MP2、MP7、MF6、MP12。因为潜在异常区域是整个管网,所以此时所有节点都可能是潜在的异常位置,所有监测仪表都很重要。但实际上,安装在敏感度较高的节点位置重要性更强,因为它更容易收集到受异常影响的数据信息。

表1 异常事件1的第一次识别结果

Tab.1 The first identification results of abnormal event 1

项目	CNN 训练耗 时/s	CNN测 试结 果/%	CNN 识别结 果	ANN 训练耗 时/s	ANN测 试结 果/%	ANN 识别结 果
子组包含10个节点	537	98.7	A13	442	95.3	A13
子组包含20个节点	366	97.9	A13	328	92.4	A13
子组包含30个节点	249	95.4	A13	261	90.8	A13
平均值	384	97.3	—	344	92.8	—

重复上述过程,第二次的分区和识别结果如图4和表2所示。

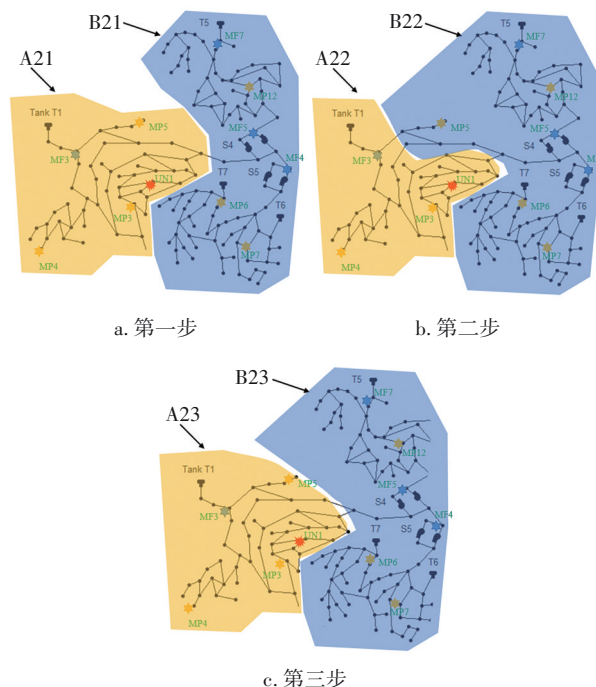


图4 异常事件1的第二次分区结果

Fig.4 The second partition results of abnormal event 1

表2 异常事件1的第二次识别结果

Tab.2 The second identification results of abnormal event 1

项 目	CNN 训练耗 时/s	CNN测 试结 果/%	CNN 识别结 果	ANN 训练耗 时/s	ANN测 试结 果/%	ANN 识别结 果
子组包含 10 个节点	481	96.4	A23	442	92.8	A23
子组包含 20 个节点	323	94.9	A23	337	91.6	A23
子组包含 30 个节点	226	92.2	A23	236	87.9	A23
平均值	343	94.5	—	338	90.8	—

注：(A23,B23)中,识别出 A23 包含异常。

因为潜在异常区域面积缩小,因此模拟的潜在异常节点分布范围和数目也随之减少,传感器的重要性也发生了变化。按照重要性由大到小的顺序,本过程训练 CNN 模型时最重要的 8 个特征是 MP3、MF3、MF1、MP2、MF5、MF4、MP4、MP5。继续进行分区和识别,第三次的结果如图 5 和表 3 所示。按照重要性由大到小的顺序,本过程训练 CNN 模型时最重要的 6 个特征是 MP3、MF3、MF5、MF4、MP4、MP5。此时卷积神经网络模型的识别准确率平均值下降至 89.9%,低于预设的迭代停止标准,分区过程停止。通过对三次分区和识别结果的综合分析可得出:第一,所提出的分区方法能够合理地二分管网区域,为后续识别过程奠定基础。基于图论和聚类的方法实现了自动划分,而人工微调则可以避免明显不合理的划分。例如图 4 中,通过人工微调,将一个独立环状的管网节点调整到同一个分区中,因为同一个独立环包含的节点敏感性更为相似。第二,所提出的识别算法能够准确有效地识别出异常所在的子区。子组的大小不仅影响训练速度,还影响识别的准确性。子组包含的节点越多,识别算法的训练速度越快,但是识别准确性会下降,因此在选择子组大小时应该权衡上述要素。另一方面,随着分区迭代的进行,子区范围缩小,区域内节点数目减少,因此模拟的计算量减少,卷积神经网络模型的训练速度提升。但是由于子区域内覆盖的监测仪表数目下降,仪表之间相互补充数据信息的作用减弱,所以识别准确率会下降。

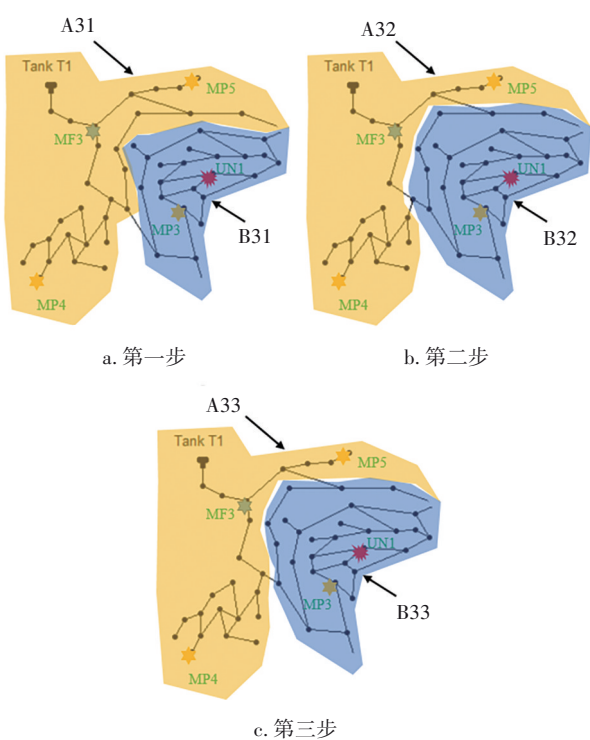


图5 异常事件1的第三次分区结果

Fig.5 The third partition results of abnormal event 1

表3 异常事件1的第三次识别结果

Tab.3 The third identification results of abnormal event 1

项 目	CNN 训练耗 时/s	CNN测 试结 果/%	CNN 识别结 果	ANN 训练耗 时/s	ANN测 试结 果/%	ANN 识别结 果
子组包含 10 个节点	417	92.4	B33	385	87.3	B33
子组包含 20 个节点	309	90.1	B33	313	84.6	B33
子组包含 30 个节点	234	87.3	B33	227	82.4	B33
平均值	320	89.9	—	308	84.8	—

注：(A33,B33)中,识别出 B33 包含异常。

从表 1~3 还可以看出,相比于人工神经网络模型,卷积神经网络模型的识别准确率更高,但是训练耗时略长,这是因为卷积神经网络模型的结构更复杂。尽管如此,人工神经网络的训练速度优势也并不明显。而且,随着分区迭代过程的进行,输入特征向量减少,卷积神经网络模型的结构复杂度明显下降,两者的训练速度差距逐渐缩小。

总体上看,经过三次分区,已经成功将异常事件隔离在一个较小的区域内,此时可以采用声学设备或者热成像设备等对隔离区域进行仔细探查,或

神经网络模型,CNN模型的识别准确性更高。

与文献[8]相比,本研究成果的优势在于:第一,对供水管网二分的过程更加合理。文献[8]根据节点的异常敏感性矩阵进行分区,仅能保证子区域内的强相似,缺陷是无法保证子区域内水流的强流通性。实际的城市供水管网结构复杂,可能存在多个环状和树状结构。如果不考虑水的连通性,那么分区的结果可能无法满足水力要求,比如树状管网的上下游被分割开,那么其末端的管网节点就成了孤立的节点,与区域内其他节点之间的水流不连通,这样的分区结果明显不合理。本研究提出的三步分区方法则弥补了该缺陷,同时满足子区域内的强相似性和强连通性的原则,还额外考虑分区边界的工程合理性,因此本方案适用于所有的实际管网,具备较强的工程实践性。第二,本研究采用了子组技术和特征选择技术,使得对异常区域识别的过程更加准确、高效。此外,卷积神经网络算法相比于传统的机器学习算法能够更准确地识别出异常子区。案例管网的应用结果也证明了上述框架方法能够可靠、高效地隔离管网异常,在准确性和时效性方面均有令人满意的表现。综上,本研究所提出的方案实现了对之前方法的完善和提升。

4 结论

本研究提出的迭代分区和识别算法的整体方案能够可靠、高效地隔离管网异常事件。联合图论技术、聚类方法和人工微调手段能够将供水管网区域合理二分,为后续异常识别准确性的提高奠定基础。采用子组技术,并对卷积神经网络识别模型进行输入特征筛选,能够提高方案的时效性。

参考文献:

- [1] MOUNCE S R, MOUNCE R B, BOXALL J B. Novelty detection for time series data analysis in water distribution systems using support vector machines[J]. *Journal of Hydroinformatics*, 2011, 13(4): 672-686.
- [2] ROMANO M, KAPELAN Z, SAVIC D A. Automated detection of pipe bursts and other events in water distribution systems[J]. *Journal of Water Resources Planning and Management*, 2014, 140(4): 457-467.
- [3] HU Z, CHEN B, CHEN W, *et al.* Review of model-based and data-driven approaches for leak detection and location in water distribution systems[J]. *Water Supply*, 2021, 21(7): 3282-3306.
- [4] 吴以文,杜坤,吴汉清,等. 基于LSSVM交互预测的供水管网爆管检测[J]. *中国给水排水*, 2022, 38(9): 58-63.
WU Yiwen, DU Kun, WU Hanqing, *et al.* Water supply network burst detection based on least squares support vector machine interactive prediction[J]. *China Water & Wastewater*, 2022, 38(9): 58-63 (in Chinese).
- [5] BRANISAVLJEVIC N, KAPELAN Z, PRODANOVIC D. Improved real-time data anomaly detection using context classification[J]. *Journal of Hydroinformatics*, 2011, 13(3): 307-323.
- [6] GUYON I, WESTON J, BARNHILL S, *et al.* Gene selection for cancer classification using support vector machines[J]. *Machine Learning*, 2002, 46(1/3): 389-422.
- [7] OSTFELD A, SALOMONS E, ORMSBEE L, *et al.* The battle of the water calibration networks (BWCN)[J]. *Journal of Water Resources Planning and Management*, 2012, 138(5): 523-532.
- [8] CHEN J, XIN F, XIAO S. An iterative method for leakage zone identification in water distribution networks based on machine learning[J]. *Structural Health Monitoring*, 2020, 20: 1938-1956.

作者简介:简彩(1993-),女,河南信阳人,博士研究生,主要研究方向为供水管网需水量预测和异常分析。

E-mail:jiancheer@163.com

收稿日期:2022-06-08

修回日期:2022-07-14

(编辑:任莹莹)