

doi: 10.11835/j.issn.1000-582X.2021.267

基于长时间序列预测的计量区给水管网爆管识别

文思齐, 龙天渝

(重庆大学 环境与生态学院, 重庆 400044)

摘要:为了迅速识别计量分区给水管网中的爆管,提出了一种新的预测-分类-校核的三阶段 Infomer-Z-score 算法。Infomer-Z-score 算法解决了传统方法数据处理效率低、不正常低用水量不处理的问题。在预测阶段中使用深度学习 Informer 算法预测管网长时间用水压力数据,提高用水压力预测的准确性和数据处理的效率。在分类阶段使用多阈值的分类方法提高了对用水压力数据随时间变化的鲁棒性。Infomer-Z-score 算法在爆管模拟检验中的真阳性率(TPR)为 90.9%、假阳性率(FPR)为 1.7%、检测准确率(DA)为 99.5%。长时间序列的压力预测不仅能用于爆管识别,而且还能有效的进行管网中的压力控制使爆管风险降低。

关键词:给水系统;爆管识别;深度学习;统计过程控制;压力控制

中图分类号:TU991

文献标志码:A

文章编号:1000-582X(2023)05-062-10

Burst detection in district metering areas based on long sequence time-series forecasting

WEN Siqi, LONG Tianyu

(College of Environment and Ecology, Chongqing University, Chongqing 400044, P. R. China)

Abstract: For the rapid detection of burst pipes in water supply networks of district metering area, a new prediction-classification-correction three-stage Infomer-Z-score algorithm was proposed. The Infomer-Z-score algorithm solves the problem of low data processing efficiency and abnormally low water consumption in traditional methods. In the prediction stage, the deep learning Informer algorithm was used to predict long sequence time-series of pressure data for the pipe network so as to improve the accuracy of water pressure prediction and the efficiency of data processing. In the classification stage, the robustness of water pressure data over time was improved by using a multi-threshold classification method. In the pipe burst simulation test, the Infomer-Z-score algorithm achieved a 99.5% (DA) detection accuracy with a 90.9% true positive rate (TPR), and a 1.7% false positive rate (FPR). Long sequence time-series pressure forecasting can be used not only for burst detection, but also for effective pressure control in the network to reduce the risk of bursts.

Keywords: water distribution system; pipe-burst detection; deep learning; statistical process control; pressure control

收稿日期:2021-09-06 网络出版日期:2021-12-20

基金项目:国家自然科学基金(41877472)。

Supported by National Natural Science Foundation of China(41877472).

作者简介:文思齐(1996—),男,硕士研究生,主要研究给水管网优化方向,(E-mail)cquwenxiaoshuai@163.com。

通讯作者:龙天渝(1960—),女,博士生导师,主要研究方向为城镇水资源与水环境,(E-mail)longty128@cqu.edu.cn。

城市供水管网为城市居民提供必需的生活用水,一旦发生爆管,将会影响居民的正常生活,浪费宝贵水资源的同时,还会造成微生物入侵管网、路面塌陷等问题^[1]。近年来,随着分区计量管理(district metering areas, DMAs)以及在独立区域计量区中数字化监控系统(supervisory control and data acquisition, SCADA)的应用^[2],供水公司可以利用实时的流量、压力传感器数据,通过实时数据驱动(data-driven)的方法对流量压力等数据进行分析,快速地找出其中的异常值,识别管道爆裂导致的漏失。

爆管识别的数据驱动方法通常包括预测-分类两阶段^[3],预测是指对未来正常运行的管网数据进行预测,而分类则是根据预测的数据和管网传感器实际观测的数据进行对比,分析两者之间的差异以判断管网是否发生爆管。Wang等^[4]在预测分类两阶段的基础上提出三阶段的数据驱动方法,在传统的两阶段算法后,增加一个校准阶段,将第二阶段中找出的异常值进行校准,用校准值后的时间序列进行下一步的预测,保持预测的精度。

早期用于预测阶段的方法主要是人工神经网络(artificial neural network, ANN)算法, Mounce等^[5]在2002年第一次采用人工神经网络的方式进行用水数据的预测。之后在该领域中人工神经网络模型被大量用于预测阶段的分析^[6,7],但当网络十分复杂时,人工神经网络的参数量太大,出现梯度弥散和梯度爆炸的现象导致学习困难。后来随着深度学习的不断发展^[8],预测算法多选用循环神经网络RNN或长短时记忆网络LSTM方法^[9]。但是RNN、LSTM在计算过程是顺序的,下一时刻的权重完全依赖于上一时刻的输出,无法充分利用计算机的并行计算提高计算效率;而且顺序计算过程中的信息丢失现象使得RNN类算法对长时间序列数据的预测误差过大。

Google研究团队在2017年提出注意力机制(Attention)机制并运用于自身的Transformer模型来进行机器翻译任务,并且取得了很好的效果^[10]。注意力机制对输入信息同时并行处理,通过嵌入(Embedding)来考虑信息的排序和位置信息。Zhou等^[11]提出了一个基于Transformer的模型Informer用于预测长时间的时间序列,通过改进注意力机制(使用ProbSparse self-attention)解决了Transformer应用于长时间序列数据中计算复杂度高、内存占用大、解码器预测速度慢等问题。

在应用数据驱动方法识别爆管时,通常采用流量数据,该方法虽然能够检测出超过正常用水量的异常事件,但却把全部低于正常用水量的流量都视为正常。如果使用管网压力数据来判断爆管,爆管会瞬间使管网出现低压状态而容易快速识别。此外,使用压力数据还可以在管网流量远远小于正常用水量(即管网出现压力过高时)进行警报,调节管网的压力使其恢复到正以减少爆管发生。

为此,笔者提出了一种压力数据驱动的爆管识别Informer-Zscore算法,该方法包括预测-分类-校准三阶段。将所提出的方法用于某管网的爆管识别,并其效果进行评价,以检验该方法的合理性。

1 研究方法

1.1 预测阶段

RNN、LSTM和Transformer、Informer处理时间序列数据中本质上都采用编码器-解码器(Encoder-Decoder)的深度学习框架,即输入一段时间序列数据,输出一段时间序列数据,解决Sequence到Sequence的映射问题。其中编码器Encoder的作用是对输入时间序列数据进行编码,提取数据的特征并转换为一个向量(称为context)。解码器Decoder的作用是逐项解码context,生成输出序列中的元素。RNN与LSTM的问题在于编码和解码context的过程中随着时间推移顺序处理,遍历context的过程中遗忘早期的信息。Transformer同时计算所有特征间的注意力(Attention)并在编码和解码context的过程中只关注相似度高的部分,但Transformer的运算量太大,内存占用过多,无法预测长时间序列的数据。例如输入一个句子进行翻译,RNN与LSTM会从头到尾逐字逐句进行翻译,当句子太长时会忘记句子最开始的语境信息;Transformer会计算句子中所有词汇的两两关系,翻译过程中会重点考虑意思相近或语境相同的词,但当句子太长包含词汇太多时计算所有词间的关系太复杂、内存占用太多。

Informer算法在编码与解码的过程中使用ProbSparse Self-attention机制,仅仅只考虑注意力机制中贡献最大的部分,对比Transformer中的attention计算量更小,内存占用低,使得输出时间序列对输入的时间序列

重要信息加以选择性地学习。综上所述,本文的预测阶段采用 Zhou 等^[11]的长序列时间序列预测算法 Informer,该算法的结构单元如图 1 所示,原论文的公式以及思路将在下文详细阐述。

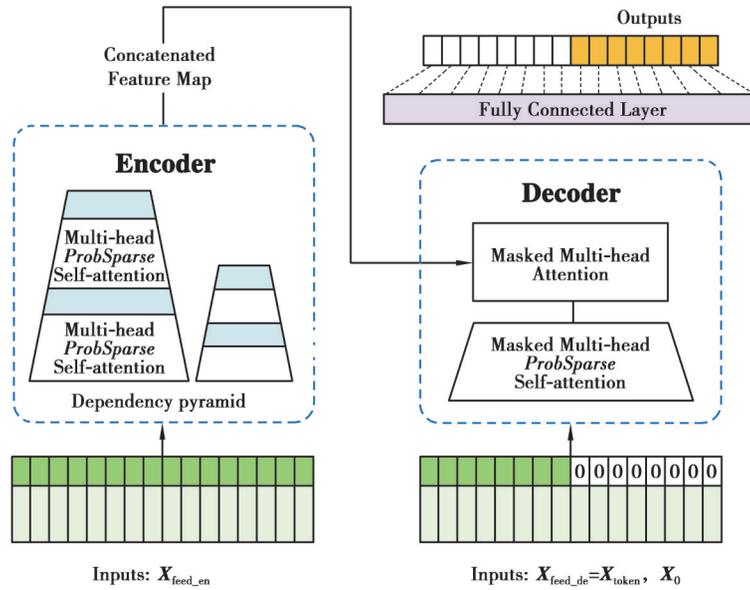


图 1 Informer 模型的结构单元^[11]

Fig. 1 Illustration of an Informer model unit^[11]

1.1.1 编码器 Encoder 的输入部分

式(1)~(3)为图 1 左侧编码器 Encoder 的信息嵌入(Embedding)计算公式,即编码器的输入部分。

$$P_{(p, 2i)} = \sin(p/10\,000^{2i/d_{\text{model}}}), \quad (1)$$

$$P_{(p, 2i+1)} = \cos(p/10\,000^{2i/d_{\text{model}}}), \quad (2)$$

$$X'_{\text{feed_en}} = au'_i + P_{(L_x(t-1)+i)} + \sum_p [S_{(L_x(t-1)+i)}]_p, \quad (3)$$

式中: P 为位置编码(positional encoding)信息(即数据在长时间序列中的位置和排序信息), p 为该数据在整个时间序列的绝对坐标, d_{model} 为映射特征的维度, L_x 为输入的长度, S 为输入序列的时间信息。式(1)为整个时间序列数据中奇数位置的 P 计算公式,式(2)为偶数位置的 P 计算公式,式(3)中 $X'_{\text{feed_en}}$ 为 Encoder 的输入,包含了时序、位置信息, au'_i 为输入的时间序列数据通过一维卷积 Conv1d 映射的高维特征向量, $P_{(L_x(t-1)+i)}$ 为位置编码的综合信息, $\sum_p [S_{(L_x(t-1)+i)}]_p$ 为全局的时间戳。

1.1.2 编码器 Encoder 的编码部分

$$Z = \text{Softmax}\left(\frac{\bar{Q}K^T}{\sqrt{d}}\right)V, \quad (4)$$

$$X'_{j+1} = \text{MaxPool}(\text{ELU}(\text{Conv1d}([X'_j]_{\text{AB}}))). \quad (5)$$

式(4)中 Z 为 ProbSparse Self-attention 机制公式计算的注意力机制权值,其中 \bar{Q} 、 K 、 V 为输入 $X'_{\text{feed_en}}$ 不同线性变换后的活跃查询向量(active query)、键向量(key)、值向量(value),即输入时间序列中重要的学习信息以及长时间序列数据之间相关关系。Multi-ProbSparse Self-attention 就是将 Embedding 之后的输入按映射的维度切割成 n 个,分别做 ProbSparse attention 之后再合并在一起。

式(5)中 AB 代表 attention block(注意力机制计算模块),表示输入 X'_j 经过 AB 层、Conv1d 映射、激活函数(ELU)与 MaxPooling 层(MaxPool)得到下一时刻的编码输入 X'_{j+1} 。同时 ProbSparse Self-attention 机制的 Distilling 使得自身级联降低。式(4)(5)的结果经过线形层得到编码器 Encoder 的输出,并传入解码器 Decoder。

1.1.3 解码器 Decoder 的解码部分

式(6)为解码器 Decoder 的输入公式

$$X'_{\text{fecd_dc}} = \text{Concat}(X'_{\text{token}}, X'_0) \in \mathbb{R}^{(L_{\text{token}} + L_y) d_{\text{model}}}, \quad (6)$$

式中: $X'_{\text{fecd_dc}}$ 为编码器 Decoder 的输入部分, X'_{token} 为编码器 Decoder 输入中的已知部分, L_{token} 为该部分长度, X'_0 为编码器 Decoder 输入中的掩盖部分, L_y 为掩盖部分长度。即 Decoder 输入一段时间序列 $X'_{\text{fecd_dc}}$, 把后半部分使用 0 掩盖, 整个模型的目的即为预测被遮盖的 Decoder 的部分。解码过程中同样通过 ProbSparse Self-attention 学习数据间的相关信息, 计算过程与公式同公式(4), 首先计算 X'_{token} 间的 ProbSparse Self-attention 注意力权值, 最后结合编码器 Encoder 中传入的信息经过 Self-attention 层进行预测, 通过一个全连接层获得输出, 预测出输入部分被 0 掩盖的信息(即 X'_0)。

1.2 分类阶段

分类阶段中常用的方法是统计过程控制理论(statistical process control, SPC), SPC 将统计理论运用在分类阶段, 判断差异到底是由爆管引起非随机误差的还是系统自带的随机误差。McBean 等^[12]使用累积和算法(cumulative sum CUSUM)来确定管网中的漏损事故, 通过计算每个数据与目标值之间偏差的累积和, 累积和超过预先设定的阈值则识别为异常数据; Palau 等^[13]使用 Hotelling T^2 方法来检测管网中的突发事件, 使用多变量数据间的马氏距离来识别超出控制限的异常数据。由于管网压力数据分类是一个单变量的异常值识别问题, SPC 方法中西部电器公司规则(Western Electronic Company rules, WEC)和 z 标准分数(Z-score)单变量方法简单有效, 都是将超出数倍数据标准差的数据视为异常数据, 但 WEC 方法识别异常值的规则过于繁琐, 故选用 Z-score 多阈值的 SPC 方法来检测预测阶段中出现的异常值。

当管网中发生爆管事故时, 观测到的用水压力数据 X_i 会和预测用水压力数据 \hat{X}_i 相差很大, 此时可以通过两者间的压力残差 R_i 对异常数据进行辨别。

$$R_i = X_i - \hat{X}_i. \quad (7)$$

由于残差始终随着时间不断进行波动, 而且用水行为的周期性导致了残差也随着周期不断波动。基于这个特性, 把残差按照时间段进行划分, 因为传感器的数据采集间隔为 5 min, 每天的压力残差转换为 288 长度的列向量(00:00, 00:05, ..., 23:55), 一共有 n 天的数据, 所以残差矩阵如下所示, 尺寸为 $288 \times n$:

$$\begin{bmatrix} R_{1,1} & \cdots & R_{n,1} \\ \cdots & \cdots & \cdots \\ R_{1,288} & \cdots & R_{n,288} \end{bmatrix}。$$

通过分析每一行的残差, 即获得特定时刻的残差阈值, 即一天中有 288 个残差阈值的多阈值分类方法。根据管网的该压力特性, 采用 z 标准分数(Z-score)方法, 即使用 z 标准分数表示一个压力数据点与样本压力平均值的标准差:

$$z = \frac{r_i - \bar{R}_i}{\sigma_i}. \quad (8)$$

式中, z 值表示 t 时刻预测值 r_i 与该特定时段残差样本均值的差异, 根据规则如果标准分数 z 超过下面的正常阈值时, 则识别非随机模式:

$$\bar{R}_n - Z \sqrt{\text{Var}[R_n]} < r_i < \bar{R}_n + Z \sqrt{\text{Var}[R_n]}, \quad (9)$$

式中, Z 为 z 标准分数的经验阈值, 一般可取值 2.00、2.65、3.00、3.50 以及更大的经验阈值。经验阈值过小将误报管网中的正常事件, 取得太大将漏报管网中的小型爆管事件, 综合上述条件和参考有关文献[14], 取经验阈值为 3.00。具体来说, 对于管网中的压力传感器, 当观测值高于预测值, 即超过正常条件下的水压时。供水管网的水压过高, 有爆管的风险, 而且背景渗漏会加剧, 需要及时对管网进行主动地压力管理。当观测值突然低于预测值时, 表明管网中可能发生了爆管事故, 需要马上进行设备的检修排查。

为了避免管网的波动性造成用水数据偶尔超过阈值导致事故的误报, 分类阶段将残差连续两次超过阈值时才会触发报警, 即本文中的爆管检测时间为 10 min。

1.3 校正阶段

对于大多数的预测分类两阶段算法中, 识别到异常值后的校正环节很少被讨论。Wang 等^[14]认为传统预

测分类两阶段算法中分类阶段已经识别出的异常值必须经过校正才能作为正常值进行下一步的预测。实际工程证明在预测分类两阶段后增加一个校正阶段很有必要,避免因为异常数据进入预测阶段导致未来连续的错误预测。除了按照上述策略对分类阶段的异常值校正外,这里还增加了对正常用水压力预测值的校正,避免正常的用水压力预测值因为残差的逐渐累积直到发生较大的偏差被误报为事故。预测值校正公式采用:

$$X_t^c = \hat{X}_t + \frac{1}{2} R_t, \quad (10)$$

$$X_t^* = \hat{X}_t + \bar{R}. \quad (11)$$

式中: X_t^c 表示没有发生事故的校正值=预测值+该次预测残差的二分之一。 X_t^* 表示发生事故的校正值,等于预测值+该时刻残差的均值。校正值将视为本时刻的正常用水压力值输入到下一时刻的预测阶段模型中,通过滑动窗口输入的方式实现对管网异常用水压力数据的连续识别。

1.4 评价指标

采用平均百分比误差(mean absolute percentage error, MAPE)来评估长时间序列预测阶段的准确性,使用真阳性率(true positive rate, TPR)、假阳性率(false positive rate, FPR)、检测准确率(detection accuracy, DA)来评价分类阶段对于事故的检测效果。

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \left| \frac{X_i - \hat{X}_i}{X_i} \right|, \quad (12)$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%, \quad (13)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \times 100\%, \quad (14)$$

$$\text{DA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \times 100\%. \quad (15)$$

式中:TP(true positive),表示事故发生并且被正确识别;FN(false negative),事故发生但未被检测到;TN(true negative),表示事故没发生系统也未检测到事故;FP(false positive),表示事故未发生但系统检测到事故。MAPE表示预测值与真实值之间的差异大小,TPR表示所有事故中被系统检测出来的比例,FPR表示误报事故的比例,DA表示所有条件下系统正确判断的比例。即MAPE、FPR越小,DA、TPR越大,该检测方法越有效。

2 结果与讨论

2.1 实验管网与数据

基于ENPNET的水力计算包WNTR(water network tool for resilience)^[15]来模拟供水管网的水力过程。WNTR是基于python语言编写的开源包工具,在EPANET的基础上提供更多功能,而且由于是python语言编写,与深度学习算法间数据交流更加方便。

这里所采用的管网是Jung等^[16]优化的Apulian管网(如图2所示),Apulian管网被广泛用于各种爆管检测算法的测试。该管网有1个水源、11个环路、23个节点,以及34条管段。主干管直径为300 mm,支管直径为100,150,200,250 mm,水库高程为51.4 m,节点的平均高程为9.6 m,系统的平均总需水量为282 L/s。在管网内部设置5个检测点,在管网入口处(节点1)放置压力传感器,其他传感器放置位于管网的内部(节点9、15、21、22)传感器每5 min记录一次压力数据,管网结构以及压力传感器位置如图2所示,压力数据随用水量动态地变化。

使用WNTR水力计算包生成2021年1月1日至2021年4月12日共102 d的用水压力数据,其中2021年1月1日到4月10日(100 d)的用水压力数据无爆管发生,在4月11日和4月12日各合成11次爆管事故,总计22次爆管事故。采用设置压力传感器的节点1、9、15、21和22的模拟压力数据进行算法训练和事故分类。

除了原始的顺序时间序列数据外,还通过提取原始时间序列中每个时间点的数据,形成每日不同时刻时间序列。管网中压力传感器每 5 min 记录一次压力数据,每次记录的压力数据作为一个样本 (sample)。

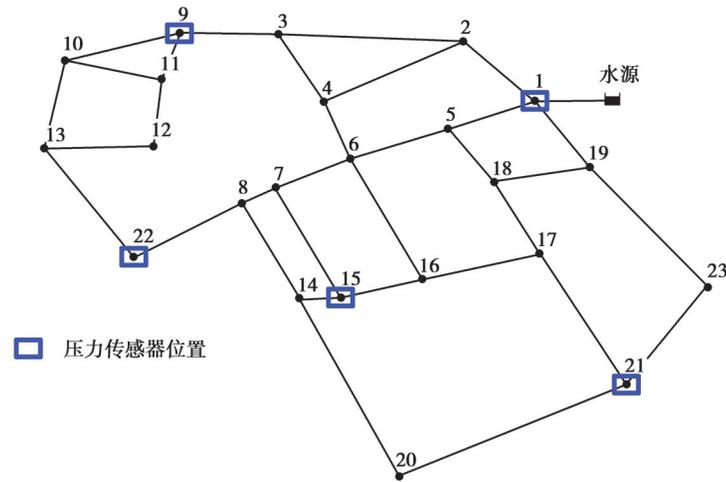


图 2 Apulian 给水管网以及仪表布局

Fig.2 Apulian network layout with meter locations

例如每天有 288 个压力数据读数,把原始顺序时间序列 $\{p_1, p_2, \dots, p_{102}\}$ 转换为日间隔时间序列 $\{p_1^1, p_2^1, \dots, p_{102}^1\}, \{p_1^2, p_2^2, \dots, p_{102}^2\}, \dots, \{p_1^{288}, p_2^{288}, \dots, p_{102}^{288}\}$, 其中 p 代表用水压力的测量值,下标为第几天的用水压力,上标为每天的不同时刻。顺序时间序列代表了用水压力数据的趋势性,日间隔时间序列表示用水压力数据的周期性。

2.2 预测过程

预测阶段采用 Informer 分析以往压力长时间序列间的相关关系(输入多个样本),预测出未来长时间序列(输出多个样本)。为了对每个样本提取更多的压力预测信息,组合日间隔时间序列数据形成样本的 3 组特征,如表 1 所示:1)样本时间 t 相邻的 $t-1$ 时刻前 7 d 的用水压力数据;2)样本时间 t 的前 7 d 同时刻用水压力数据;3)样本时间 t 相邻的 $t+1$ 时刻前 7 d 用水压力数据。

表 1 压力数据样本特征

Table 1 Pressure data characteristics and input and output of the informer model

样本特征	输入的压力值时刻
1) $t-1$ 时刻的前 7 d 压力数据	$t-1-7(24\text{ h}), t-1-6(24\text{ h}), t-1-5(24\text{ h}), t-1-4(24\text{ h}),$ $t-1-3(24\text{ h}), t-1-2(24\text{ h}), t-1-1(24\text{ h})$
2) t 时刻的前 7 d 压力数据	$t-7(24\text{ h}), t-6(24\text{ h}), t-5(24\text{ h}), t-4(24\text{ h}), t-3(24\text{ h}),$ $t-2(24\text{ h}), t-1(24\text{ h})$
3) $t+1$ 时刻的前 7 d 压力数据	$t+1-7(24\text{ h}), t+1-6(24\text{ h}), t+1-5(24\text{ h}), t+1-4(24\text{ h}),$ $t+1-3(24\text{ h}), t+1-2(24\text{ h}), t+1-1(24\text{ h})$

每个压力数据样本一共有 21 个特征,样本的标签为该时刻 t 的用水压力值。Informer 模型输入 24 h 的顺序时间序列(288 个样本),预测接下来 30 min 的正常用水压力数据(6 个样本)。长时间序列(288 个样本)的输入考虑了用水行为的趋势性,每个样本中日间隔时间序列组合的特征表示用水行为的周期性。预测阶段 Informer 的参数如下所示:seq_len=288、label_len=12、pred_len=6、d_model=512、d_ff=2 048、activation='gelu'、enc_in=22、dec_in=22、c_out=1、e_layers=2、d_layers=1;batch_size=32、learning rate=0.000 1。

数据集分为训练集(前 70 天中 20 160 个样本)、验证集(70 ~ 90 d 中 5 760 个样本)、测试集(90~100 d 中

2 880个样本),爆管应用阶段(101、102 d中576个样本)。模型在训练集上训练之后,在验证集上计算误差,通过调整参数直至模型在验证集上的误差取得最小值,训练好的模型在测试集上评估最终的预测能力。爆管应用阶段数据将在分类阶段用于与预测值进行对比验证爆管事件。

2.3 爆管检测结果分析

实验在2021年4月11日与4月12日间进行22次爆管事故的模拟,爆管事故兼顾爆管的大小和持续时间,当管网中任意传感器观测到连续的异常用水压力波动时发出警报,最早发出警报的传感器位置被认为最靠近事故的发生地点。22次爆管事故最后一次事故是在不正常用水的高压之后发生的爆管。爆管规模大小与规模如表2所示。

表2 爆管事故及检测情况信息
Table 2 Data of burst pipe node and detection

事故发生位置 (节点位置)	发生日期	开始时间	结束时间	爆管泄漏流量与正 常流量之比/%	爆管时压力与正常 用水压力之比/%	检测结果	最早报警传 感器节点位 置
14	2020-04-11	00:05:00	00:30:00	12.4	91.2	TP	15
15	2020-04-11	02:00:00	02:20:00	15.6	88.6	TP	15
16	2020-04-11	04:00:00	04:30:00	10.1	93.2	TP	15
8	2020-04-11	06:05:00	06:25:00	5.6	95.6	TP	22
7	2020-04-11	08:03:00	08:19:00	5.2	95.9	TP	15
6	2020-04-11	10:06:00	10:30:00	4.1	96.8	TP	15
20	2020-04-11	12:05:00	12:30:00	6.2	94.8	TP	15
17	2020-04-11	14:00:00	14:15:00	4.5	96.0	TP	21
5	2020-04-11	16:05:00	16:30:00	3.8	98.1	TP	1
22	2020-04-11	18:03:00	18:25:00	4.2	96.7	TP	22
21	2020-04-11	20:00:00	20:25:00	5.1	95.9	TP	21
10	2020-04-12	00:05:00	00:30:00	4.4	96.6	TP	9
19	2020-04-12	02:00:00	02:20:00	4.6	96.4	TP	1
3	2020-04-12	04:00:00	04:30:00	4.1	96.8	TP	9
2	2020-04-12	06:05:00	06:25:00	3.6	98.3	TP	1
4	2020-04-12	08:03:00	08:19:00	4.2	96.8	TP	9
11	2020-04-12	10:06:00	10:30:00	4.1	97.0	TP	9
12	2020-04-12	12:05:00	12:30:00	4.3	96.8	TP	9
13	2020-04-12	14:00:00	14:15:00	4.2	96.7	TP	22
1	2020-04-12	16:05:00	16:30:00	3.3	98.9	TP	1
5	2020-04-12	18:03:00	18:25:00	2.8	99.3	FN	-
6	2020-04-12	20:00:00	20:25:00	3.1	99.1	FN	-

在 4 月 11 日 11 次爆管事故中,爆管流量大于 3.8% 的事故系统全部能够识别,即便事故发生地点离传感器稍远,爆管发生后的压力信号仍然能被系统识别出来。在 4 月 12 日合成的 11 次小型爆管事故中,系统在识别用水高峰期的爆管事故时(18:03:00、20:00:00)出现问题,其中 18:03:00 发生的小型爆管事故被掩盖在用水高峰的压力波动中,20:00:00 的爆管是发生在不正常用水导致的高压情况,该次高压发出警报,但未能检测出之后的爆管事件,实际中如果能迅速调整不正常低用水导致的高压,或许能避免该次事故的发生。4 月 12 日位于节点 1 的压力传感器用水压力以及残差情况如图 3 ~ 4 所示,大部分用水异常压力数据被识别。

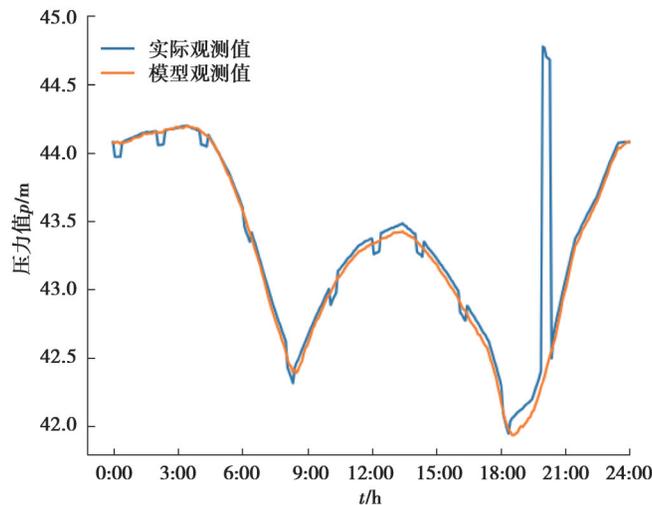


图 3 4 月 12 日节点 1 压力图

Fig.3 Node 1 pressure chart on 12 April

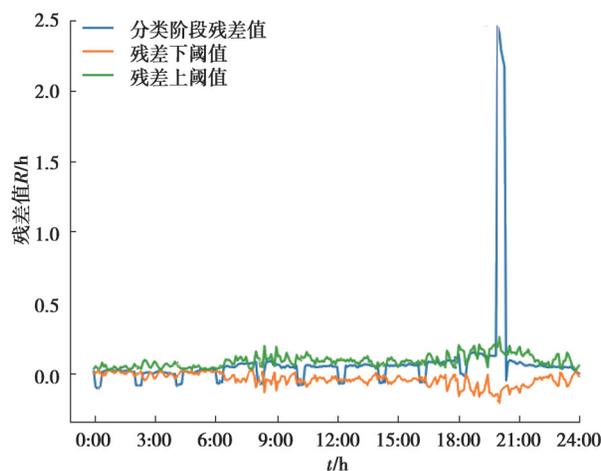


图 4 4 月 12 日节点 1 残差图以及检测情况

Fig.4 Node 1 residual plots and testing on 12 April

总的来说,管网在 2 d 内模拟了 22 次爆管事故,一共检测出 20 起真实事故,另误报 1 次事故,该方法的 TPR=90.9%、FPR=1.7%、DA=99.5%,对爆管事故的检测具有可靠性。

2.3.1 Informer 预测性能评估

Informer 预测是爆管检测算法的第一步,预测的效果直接影响爆管识别的准确度。为了检验 Informer 的预测效果,ANN 以及 LSTM 算法也被用于预测阶段进行对比。LSTM 与 ANN 的参数均使用 Informer 的调参策略并选择其最佳的预测结果进行对比。结果表明,Informer 在长时间序列的预测优于 LSTM,ANN。相同

输入条件下,Informer的MAPE均值为0.8%,远远超过LSTM的4.8%以及ANN的10.2%。Informer独特的ProbSparse Self-attention机制使它对长时间序列数据的预测精度更佳,这为接下来的分类奠定了好的基础。

2.3.2 基于Z-score的多阈值机制评估

Z-score多阈值的分类能尽可能地使得异常的残差数据被识别,同时也避免因为个别时间点压力数据的随机波动导致事故误报。4月11日中节点15的多阈值残差图如图5所示,在低用水量的时间段(00:00:00—06:00:00),残差阈值小;高用水量水时间段(08:30:00、18:00:00、23:00:00),用水波动性较大,所以残差的上下阈值较大。Z-score多阈值的方法避免用水稳定时的阈值导致波动性较大出现时间点的情况。因此在检测到系统的连续两次超过残差阈值时才会认定为爆管,避免用水峰高峰出现误判事故。

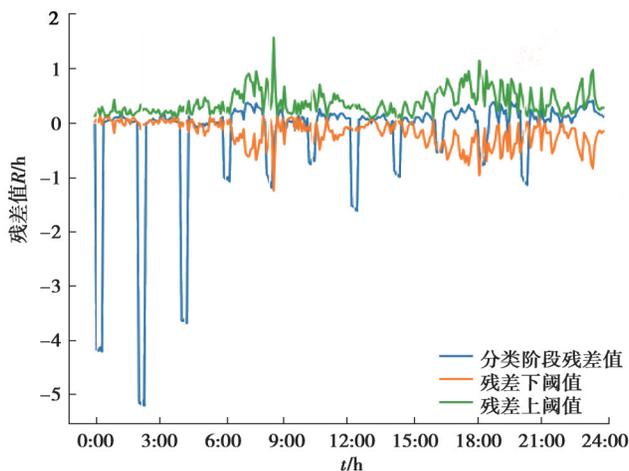


图5 4月11日节点15残差图以及检测情况

Fig.5 Node 15 residual map and detection on 11 April

使用该Z-score多阈值方法识别22次爆管事故,一共检测出20起爆管,另误报1次,该方法的TPR=90.9%、FPR=1.7%、DA=99.5%。当识别变为发现超过一次阈值即认为发生爆管时,误报事故变为5次,TPR=20%,表明该方法能够快速有效地识别管网中的不同程度的漏损并避免误报。

3 结论

提出了一种长时间序列的数据驱动的爆管检测算法,并使用Apulian管网模型检验。结论如下:

1)模拟结果表明Informer-Z-score方法的预测阶段MAPE为0.8%,远远超过其他RNN、ANN算法。在分类阶段,该方法的TPR=90.9%、FPR=1.7%、DA=99.5%,多种结果表明该方法具有良好地爆管检测功能。

2)目前该方法能够及时识别管网模型中的爆管事故,并且长序列的用水数据预测能够为供水公司提供压力管理以及管网运行参考。如预测长时间序列的压力数据时,当观测实际管网压力过高时,可以及时进行管网的压力管理,预防爆管发生。

3)下一步将该算法应用于实际中的给水管网爆管识别,对该方法的实用性能做进一步的探究。

参考文献

- [1] Shortridge J E, Guikema S D. Public health and pipe breaks in water distribution systems: Analysis with Internet search volume as a proxy[J]. Water Research, 2014, 53: 26-34.
- [2] Chen J D, Boccelli D L. Forecasting hourly water demands with seasonal autoregressive models for real-time application[J]. Water Resources Research, 2018, 54(2): 879-894.
- [3] Wu Y P, Liu S M. A review of data-driven approaches for burst detection in water distribution systems[J]. Urban Water Journal, 2017, 14(9): 972-983.
- [4] Wang X T, Guo G C, Liu S M, et al. Burst detection in district metering areas using deep learning method[J]. Journal of Water Resources Planning and Management, 2020, 146(6): 04020031.

- [5] Mounce S R, Day A J, Wood A S, et al. A neural network approach to burst detection[J]. *Water Science and Technology*, 2002, 45(4/5): 237-246.
- [6] Romano M, Kapelan Z, Savić 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.
- [7] Aksela K, Aksela M, Vahala R. Leakage detection in a real distribution network using a SOM[J]. *Urban Water Journal*, 2009, 6(4): 279-289.
- [8] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [9] Guo G C, Liu S M, Wu Y P, et al. Short-term water demand forecast based on deep learning method[J]. *Journal of Water Resources Planning and Management*, 2018, 144(12): 04018076.
- [10] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[EB/OL]. (2017-12-06)[2021-08-10]. <https://arxiv.org/abs/1706.03762>.
- [11] Zhou H Y, Zhang S H, Peng J Q, et al. Informer: beyond efficient transformer for long sequence time-series forecasting[EB/OL]. (2021-03-28)[2021-08-10]. <https://arxiv.org/abs/2012.07436>.
- [12] McBean E A. Failure mechanisms and monitoring methodologies pertinent for detection of exposure risks in water distribution networks[C]//*Security of Water Supply Systems: from Source to Tap*, 2006, 8:9-17. DOI:10.1007/1-4020-4564-6_2.
- [13] Palau C V, Arregui F J, Carlos M. Burst detection in water networks using principal component analysis[J]. *Journal of Water Resources Planning and Management*, 2012, 138(1): 47-54.
- [14] Vaghefi M, Mahmoodi K, Akbari M. A comparison among data mining algorithms for outlier detection using flow pattern experiments[J]. *Scientia Iranica*, 2018, 25(2): 590-605.
- [15] Klise K A, Bynum M, Moriarty D, et al. A software framework for assessing the resilience of drinking water systems to disasters with an example earthquake case study[J]. *Environmental Modelling & Software*, 2017, 95: 420-431.
- [16] Jung D, Kang D, Liu J, et al. Improving the rapidity of responses to pipe burst in water distribution systems: a comparison of statistical process control methods[J]. *Journal of Hydroinformatics*, 2015, 17(2): 307-328.

(编辑 郑洁)