

基于数据特征融合的管网信息物理异常诊断方法

马大中¹ 胡旭光¹ 孙秋野¹ 郑君¹ 王睿¹

摘要 随着管网物理空间和信息网络的深度融合,系统面临着物理和信息空间异常带来的运行风险.本文根据管网系统数据量大、耦合性强的特点,提出一种基于数据特征融合的信息物理异常诊断方法.首先通过站场信息数据构建信息增维矩阵并且通过矩阵预分析实现信息传输中断异常的判断.然后基于不同站场信息构建的信息增维协方差矩阵,通过矩阵特征值分布的变化情况对物理异常以及信息传输错误异常进行区分.在此基础上,为了对管网物理异常分类实现系统运行状态的有效分析,将管网信息增维协方差矩阵最大特征向量映射的二维图像作为输入,采用卷积神经网络进行研究,进而实现对物理异常的准确判断.最后通过某实际管网数据进行仿真分析,验证所提方法的有效性.

关键词 信息物理系统, 输油管网, 数据驱动, 异常诊断

引用格式 马大中, 胡旭光, 孙秋野, 郑君, 王睿. 基于数据特征融合的管网信息物理异常诊断方法. 自动化学报, 2019, 45(1): 163–173

DOI 10.16383/j.aas.2018.c180383

Cyber-physical Abnormity Diagnosis Method Using Data Feature Fusion for Pipeline Network

MA Da-Zhong¹ HU Xu-Guang¹ SUN Qiu-Ye¹ ZHENG Jun¹ WANG Rui¹

Abstract With deep fusion between pipeline physical network and cyber network, the system is facing operational risks caused by cyber-physical anomalies. According to pipe network features of big data and strong coupling, this paper proposes a cyber-physical abnormity diagnosis method using data feature fusion for pipeline network. Firstly, a cyber augmented matrix can be built by station cyber data, and cyber interrupt is diagnosed by matrix pre-analysis. Furthermore, the cyber augmented covariance matrix is established in light of the different station cyber, and physical anomaly and cyber transmission error anomaly are distinguished from each other by the variation condition of the matrix eigenvalue distribution. To effectively analyze the operating state of the physical network anomaly classification, the two-dimensional images which are mapped by the maximum eigenvectors of the cyber augmented covariance matrix of the pipeline network are regarded as the input signals, meanwhile, a convolutional neural network is utilized to carry out the analysis, thus the accurate judgment of the physical anomaly is realized. Eventually, the effectiveness of the proposed method is demonstrated through the time-domain simulation result obtained on a practical pipeline network.

Key words Cyber-physical systems (CPS), oil pipeline network, data driven, abnormity diagnosis

Citation Ma Da-Zhong, Hu Xu-Guang, Sun Qiu-Ye, Zheng Jun, Wang Rui. Cyber-physical abnormity diagnosis method using data feature fusion for pipeline network. *Acta Automatica Sinica*, 2019, 45(1): 163–173

随着信息和通信技术的不断进步与发展,工业系统依靠智能化的检测执行设备、可靠的网络通信

以及高度集成的决策分析信息促使工业系统物理设备及信息空间的联系更加紧密.异构空间融合发展的过程衍生出灵活高效的信息物理系统(Cyber-physical systems, CPS)^[1–3].由于系统具有实时、高效、高性能的特点,同时能够有效地实现工业过程稳定有效的监测与控制,信息物理系统成为未来工业 4.0 计划的核心基础.

作为工业信息物理系统的典型应用,输油管网系统已经从管道自动化网络发展成融合管道设备、通信体系以及信息网络的复杂智能化网络体系.由于输油管网在保障能源供应、维护能源安全方面具有重要价值,因此在日常生产调度中需要实时监测管网物理空间异常情况^[4].文献[5]采用泵状态、阀门开度以及管道设计参数等变量建立管道模糊分类

收稿日期 2018-05-31 录用日期 2018-09-12
Manuscript received May 31, 2018; accepted September 12, 2018

国家重点研发计划(2017YFF0108800),国家自然科学基金项目(61773109, 61573094),国家自然科学基金重大项目(61627809),中央高校基本科研业务费专项基金(N160404005)资助

Supported by National Key Research and Development Program of China (2017YFF0108800), National Natural Science Foundation of China (61773109, 61573094), Major Program of National Natural Foundation of China (61627809), and Fundamental Research Funds for the Central Universities (N160404005)

本文责任编辑 曹向辉
Recommended by Associate Editor CAO Xiang-Hui

1. 东北大学信息科学与工程学院 沈阳 110819
1. College of Information Science and Engineering, Northeastern University, Shenyang 110819

模型,从而完成对管道运行状态的判断分类.文献[6-7]从单条管道压力时间序列出发对泄漏异常情况进行诊断和定位,文献[6]采用结构相似度准则判断泄漏情况并且确定泄漏位置,同时文献[7]通过马尔科夫链提取管道压力数据特征进行异常检测并且采用相似性和连续小波定位方法找出泄漏源位置.文献[8]将小波提取的管道压力拐点时间作为特征对双支持向量机进行训练完成对管道泄漏点的定位.随着管网信息物理系统进一步发展,系统内部产生与累积了海量过程数据,机器学习方法由于其强大的数据挖掘能力,迅速的分析速度以及良好的泛化性能^[9-11],尤其是卷积神经网络(Convolutional neural network, CNN)自主挖掘图像深度特征的能力,因而广泛用于管道系统中,为管网系统的异常检测、分析和运行等提供了直接或辅助的决策.文献[12]提出一种融合一维卷积神经网络和支持向量机的方法检测管网泄漏情况.文献[13]通过卷积神经网络对管道漏磁图像进行训练,进而识别出管道的缺陷区域和非缺陷区域.通过对卷积神经网络增加视觉转化层,文献[14]可以准确地辨识不同形状和大小的缺陷特征.除了对缺陷进行识别以外,文献[15]使用卷积神经网络对管道焊缝法兰组件进行识别,并且结果表明卷积神经网络对漏磁图像具有良好的位移和畸变鲁棒性.

此外随着管网系统物理空间和信息网络深度融合和实时交互,信息网络的功能不断升级改造使得管网系统的可观测性不断增强,管网系统的功能和作用范围得到极大扩展.但是与此同时,信息网络也增加了管网系统的复杂性和不确定性,给管网系统带来新的安全问题.信息网络的异常会导致管网系统设备失去控制造成甩泵、阀门关断等紧急运行情况,如果处置不当甚至会诱发连锁故障使得管网全线停输^[16-17].文献[18]对管道SCADA系统信息传输中断的原因及保护措施进行了全面的分析.由上述文献可知,对信息传输中断或者数据传输错误异常进行快速检测具有重要意义,及早发现信息异常能够避免重大安全事故的发生.

综上所述,为了能够区分管网不同空间异常变化情况,本文以管网系统信息数据为基础,提出一种数据特征融合的管网信息物理异常诊断方法.通过对站场信息数据构建信息增维协方差矩阵实现数据特征的融合.在利用矩阵特征值分布情况实现信息异常判断的同时,将矩阵最大特征向量转换成的图像作为输入,通过卷积神经网络完成不同物理异常的分析.本文创新点在于:

1) 提出一种基于数据特征融合的管网信息物理异常诊断方法.首先通过信息增维协方差矩阵最大特征值对管网不同异常类型情况进行检测和分析,

进而在此基础上通过矩阵最大特征向量映射的图像完成对物理异常的分析.

2) 针对管网系统的海量信息数据,采用随机矩阵谱理论实现数据特征的构建.通过管网各个站场信息数据构建信息增维协方差矩阵数据特征,从而能够在有效地降低数据输入量的基础上对管网进行实时分析.

3) 采用集中-分布协同检测形式完成对管网异常情况的检测.通过信息增维协方差矩阵最大特征值完成管网异常情况的判断,然后通过选取的管网全部站场的矩阵最大特征向量完成物理异常诊断,从而实现站场异常诊断与信息物理异常分类的有机结合,提高了异常诊断响应能力.

本文其余内容安排如下:第1节根据管网特性对站场数据进行梳理,建立起管网的信息物理系统模型;第2节介绍基于信息增维协方差矩阵的管网信息物理异常诊断方法;第3节通过仿真实例验证本文所提出方法的有效性;最后,给出了本文的结论.

1 管网信息物理模型

1.1 管网图论

输油管网模型是信息物理异常诊断的基础,所建立的模型结构既要从全局上描述整个系统,又要兼顾各个站场在系统中的功能和地位.输油管网具有网络状拓扑结构并且管道内油品具有方向性,因此本文采用有向拓扑图理论进行研究分析.假设管网系统表示为有向图 $G=(V, E)$,其中 $V=\{v_1, v_2, \dots, v_n\}$ (n 为节点数目), $E=\{e_1, e_2, \dots, e_m\}$ (m 为边连接数目).在管网信息物理系统中,节点表示站场物理及信息节点,并且不同节点通过连接边完成能量及信息流传递,最终实现管网系统的平稳运行和完成额定输油计划.

1.2 管网结构

管网图论模型是设备硬件等物理组元和管道数据等信息组元的综合反映,同时也是系统集成的基础.因为管网信息物理系统是一个融合物理实体设备与信息数据网络的海量异构系统,所以从物理信息融合交互角度对管网的数据结构进行划分,具体结构如图1所示.

在物理网络中,节点表示实际存在的数据站场并且通过长距离密闭输油管道实现油品的输送.在实际输油管网系统中,通常采用管道两端压力和流量评价和分析管网运行情况,因此本文定义用于管网系统异常分析的信息数据为压力、流量.定义 t 时

刻站场 k 的信息数据 $\mathbf{z}_k(t)$ 为

$$\mathbf{z}_k(t) = h(\mathbf{x}) + \alpha \quad (1)$$

其中, $h(\mathbf{x})$ 为非线性测量函数, $\mathbf{x} \in \mathbf{R}^{n_k}$ 为站场 k 的测量向量, α 为测量噪声, 站场编号 $k = 1, 2, \dots, n$.

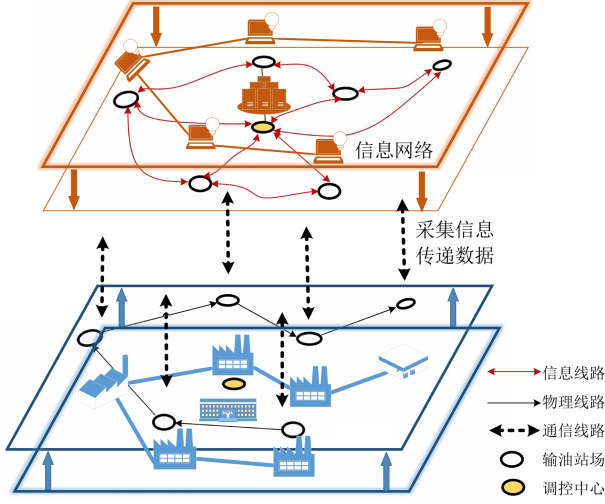


图1 管网信息物理系统结构

Fig.1 CPS structure of pipeline network

为提高采集的数据可靠性, 管网会安装多个设备采集同一信息数据, 所以同一采样时刻信息数据可以描述为向量 $\mathbf{z}_k(t)$

$$\mathbf{z}_k(t) = [z_k^1(t), z_k^2(t), \dots, z_k^j(t)]^T = [h(\mathbf{x}) + \alpha_1, h(\mathbf{x}) + \alpha_2, \dots, h(\mathbf{x}) + \alpha_j]^T \quad (2)$$

其中, j 为采集同一信息数据量的设备个数.

多元信息网络是由实际物理设备采集到的实时信息数据构成的, 用于判断和分析管道的运行状态. 因为管网系统为有向图结构, 为了能够实时反映站场数据流情况, 定义 n_k 维的站场信息节点向量 $\mathbf{d}_k(t)$ 为

$$\mathbf{d}_k(t) = [\mathbf{z}(t)^{\text{in}}, \mathbf{z}(t)^{\text{out}}]^T = [p^{\text{in}}, f^{\text{in}}, p^{\text{out}}, f^{\text{out}}]^T_{n_k \times 1} \quad (3)$$

其中, $\mathbf{z}(t)^{\text{in}}$ 为站场 k 的进站信息数据向量, $\mathbf{z}(t)^{\text{out}}$ 为站场 k 的出站信息数据向量. p 和 f 分别表示为站场压力和流量, 上标 in 和 out 分别表明数据的采集点为进站位置和出站位置.

假设管网系统信息节点向量 $\mathbf{d}_k(t)$ 通过 t_s 次采样后构成行为信息数据, 列为时间长度的矩阵 D_k ($D_k \in \mathbf{R}^{n_k \times t_s}$). 为进一步通过管网信息数据情况判断出管网异常情况, 本文将站场信息节点向量 $\mathbf{d}_k(t)$ 进行增维实现突出信息数据变化变化的目的. 由于管网中各个站场的压力变送器、流量计设备类型及数量各不相同, 所以设置增维变量 \hat{f}_k , 根据实际管网站场情况对信息节点矩阵实现不同程度地增维.

因为信息节点矩阵 D_k ($D_k \in \mathbf{R}^{n_k \times t_s}$) 每一行表示为信息数据的时间序列, 因此为了能够突显信息数据的变化以及保持数据变化的连续性, 对矩阵 D_k ($D_k \in \mathbf{R}^{n_k \times t_s}$) 行向量依次增维 \hat{f}_k 次, 构成如式 (4) 所示的增维矩阵 \hat{D}_k .

$$\hat{D}_k = [D_k, D_k, \dots, D_k]_{\hat{n}_k \times t_s} = [\hat{\mathbf{d}}_k^1, \hat{\mathbf{d}}_k^2, \dots, \hat{\mathbf{d}}_k^{\hat{n}_k}]^T = [\mathbf{z}^{\text{in}}, \mathbf{z}^{\text{in}}, \dots, \mathbf{z}^{\text{in}}, \mathbf{z}^{\text{out}}, \mathbf{z}^{\text{out}}, \dots, \mathbf{z}^{\text{out}}]^T \quad (4)$$

其中, $\hat{n}_k = \hat{f}_k \times n_k$, $\hat{f}_k < \lfloor t_s/n_s \rfloor$, $\lfloor \cdot \rfloor$ 表示不超过“.”的最大整数.

2 基于数据特征的信息物理异常诊断

为了阐述本文提出的异常诊断方法, 本节以站场 k 的信息物理异常诊断为例进行说明, 并且定义物理异常为工况调整和泄漏, 信息异常为信息传输中断和信息传输错误. 在第 1 节建立管网信息物理系统增维矩阵的基础上, 依据随机矩阵理论^[19] 构建信息增维协方差矩阵. 接着通过站场信息增维协方差矩阵最大特征值的变化区别出物理异常和信息传输错误异常, 最后通过矩阵最大特征向量融合映射成的图像完成不同物理异常的区分.

为保证信息增维协方差矩阵 $S_{\hat{D}_k}$ 数据 $\hat{d}_k^i(t)$ 的可靠性, 首先对增维矩阵 \hat{D}_k 进行预分析从而保证站场 k 未出现信息传输中断异常情况. 当站场 k 的增维矩阵 \hat{D}_k 数据满足数据预分析式 (5) 时, 说明 t_s 时刻管网系统发生信息传输中断.

$$\hat{d}_k^i(t_s - t_b) = \dots = \hat{d}_k^i(t_s - 1) = \hat{d}_k^i(t_s) \quad \text{s. t.} \quad \text{num}(i) = \vec{n} \cdot \hat{f}_k \quad 1 \leq \vec{n} \leq \hat{n}_k \quad (5)$$

其中, t_b 为系统允许的最大数据中断时间.

在确保站场 k 数据可靠性的基础上, 将随机噪声引入矩阵 \hat{D}_k 保证信息数据间的独立性, 减少由于增维过程带来的数据高度一致性, 从而完成构建信息增维协方差矩阵 $S_{\hat{D}_k}$ 的第一步. 构建的信息增维矩阵 \tilde{D}_k 如下所示.

$$\tilde{D}_k = \hat{D}_k + A_k \times G \quad (6)$$

其中, A_k 为增加的随机噪声幅值, G 为标准高斯噪声矩阵.

接着根据式 (7) 逐行对信息增维矩阵元素 $\tilde{d}_k^i(t)$ 进行标准化处理, 得到标准信息增维矩阵 \bar{D}_k .

$$\bar{d}_k^i(t) = \frac{\tilde{d}_k^i(t) - \bar{d}_k^i}{\sigma_{\tilde{d}_k^i}} \quad (7)$$

其中, $i = 1, 2, \dots, \hat{n}_k, t = 1, 2, \dots, t_s, \bar{d}_k^i$ 和 $\sigma_{d_k^i}$ 分别为矩阵 \bar{D}_k 第 i 行的均值和标准差.

最终得到信息增维协方差矩阵 $S_{\bar{D}_k}$ 为

$$S_{\bar{D}_k} = \frac{1}{\hat{n}_k} \bar{D}_k \bar{D}_k^T = U_{\bar{D}_k} L_{\bar{D}_k} U_{\bar{D}_k}^{-1} \quad (8)$$

其中, 矩阵 $U_{\bar{D}_k}$ 为特征向量矩阵, 对角矩阵 $L_{\bar{D}_k}$ 包含矩阵 $S_{\bar{D}_k}$ 的全部特征值 $\lambda_{\bar{D}_k}^i (i = 1, 2, \dots, \hat{n}_k)$.

根据随机矩阵 MP 律可知, 当管网系统处于平稳运行状态时, 信息数据不会出现明显地变化, 矩阵 $S_{\bar{D}_k}$ 特征值 $\lambda_{\bar{D}_k}^i$ 的分布满足^[19]

$$1 + r - 2\sqrt{r} \leq \lambda_{\bar{D}_k}^i \leq 1 + r + 2\sqrt{r} \quad (9)$$

其中, 矩阵 \bar{D}_k 维度比 $r = \hat{n}_k/t_s$.

如果管网系统发生异常进而导致信息数据改变情况时, 特征值 $\lambda_{\bar{D}_k}^i$ 会偏离式 (9) 的特征值分布范围, 尤其是最大特征值 $\lambda_{\bar{D}_k}^{\max}$ 会远超过理论值.

因此为了检测站场 k 的信息传输异常和物理异常情况, 本文采用矩阵最大特征值 $\lambda_{\bar{D}_k}^{\max}$ 分析判断站场 k 的信息数据变化情况, 当站场最大特征值满足下式时, 表明 t_s 时刻站场 k 发生异常情况.

$$\lambda_{\bar{D}_k}^{\max} \geq \gamma \cdot \lambda^{\text{set}} \quad (10)$$

其中, γ 为设定的异常阈值, λ^{set} 为预先定义的增维协方差矩阵 $S_{\bar{D}_k}$ 的最大理论特征值.

根据输油管网特性可知, 当站场 k 的信息增维矩阵最大特征值 $\lambda_{\bar{D}_k}^{\max}$ 满足式 (10) 时, 需要进一步综合管网其余站场信息数据进行物理信息异常的分析. 为了能够区分信息异常和物理异常, 首先定义最大站场数据传输时间 t_g 为

$$t_g = t_g^{\text{set}} + t^{\text{thr}} \quad (11)$$

其中, t^{thr} 为数据传输延迟时间, 站场数据传输时间 t_g^{set} 为

$$t_g^{\text{set}} = \frac{L_{kg}}{v_{kg}} \quad (12)$$

其中, L_{kg} 为站场 k 和 g 间的管道长度, v_{kg} 为站场间负压波速.

当相邻站场 g 的信息增维协方差矩阵 $S_{\bar{D}_g}$ 经过长度为 t_g 的采样时间后未满足式 (10), 则表明 t_s 时刻站场 k 的信息数据改变是由信息传输错误异常引起的.

在信息增维协方差矩阵特征值实现信息异常判断的基础上, 为了对物理异常进行分析, 本文采用卷积神经网络进行物理异常类型的识别. 首先从管网站场的信息增维协方差矩阵最大特征向量中选取 \tilde{f}_k 维的最大特征向量, 然后将选取的最大特征向量按照管网站场输油顺序进行排列, 将 t_n 秒数据构成维度为 $\tilde{f}_k \times t_n$ 的矩阵; 接着根据矩阵内的数值情况将矩阵内元素映射为 $0 \sim 255$ 范围内的数字, 从而得到 $\tilde{f}_k \times t_n$ 维的灰度图; 最后通过伪彩色变换实现灰度图向 RGB 图像的转换. 在得到输入图像后, 通过卷积神经网络的卷积核和池化窗进行特征自主提取, 使最大特征向量包含的信息经过非线性模型转化为更抽象的特征, 最终完成物理异常的识别^[20-21].

针对管网系统物理异常分类问题构建的卷积神经网络如图 2 所示, 由 4 个卷积层、4 个批正则化层、4 个池化层、1 个全连接层以及 1 个 Softmax 回归层组成.

图 2 中的卷积层包含卷积运算和非线性运算两种计算结构, 卷积层图像的特征向量是由卷积核的尺寸决定的, 将上一层的输出与当前层 32 个大小为 5×5 的卷积核进行卷积运算, 然后通过激活函数对卷积结果进行非线性运算, 也就是说通过非线性函数和加入偏置项得到最终当前层的输出特征映射 \mathbf{X}_{c+1} .

$$\mathbf{X}_{c+1} = f_c(\mathbf{W} \cdot \mathbf{X}_c + \mathbf{b}) \quad (13)$$

其中, \mathbf{W} , \mathbf{X}_c , \mathbf{b} 分别为第 c 卷积层的权重系数向量、图像特征向量以及偏置系数向量. f_c 为第 c 层的非线性激活函数, 本文采用 Relu 函数作为激活函数.

为了能够提升卷积神经网络的模型拟合能力, 在图 2 卷积神经网络结构中加入批正则化层, 对

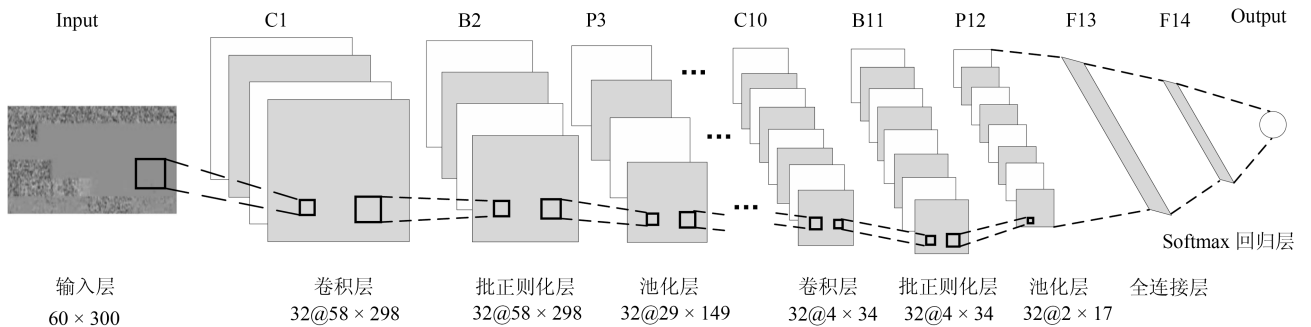


图 2 管网卷积神经网络模型

Fig. 2 CNN structure of pipeline network

卷积层输出特征进行正则化处理, 使得图像数据特征的分布更加均匀。

接着池化层的每个神经元对局部接收区域进行池化操作, 起到二次特征提取及实现非线性降采样运算的作用, 进而得到池化层输出图像的特征向量。本文中池化层的池化窗大小为 2×2 , 步长为 2, 并且神经元表达式为

$$\mathbf{X}_{c+1} = f_c(\Phi \times \text{down}(\mathbf{X}_c) + \mathbf{b}) \quad (14)$$

其中, $\text{down}(\cdot)$ 为最大值池化函数。 Φ 和 \mathbf{b} 分别为权重系数向量和偏置系数向量。

在经过 12 层网络的图像特征提取之后, 将特征图输入到具有 1024 个神经元的全连接层, 通过全连接将提取的特征图连接成一个一维向量。进而将其输入到 Softmax 回归层进行物理异常的分类。

基于上述分析可知: 当站场 k 信息数据发生改变时, 根据信息增维协方差矩阵最大特征值和最大特征向量的变化可以诊断出管网信息物理异常, 并且基于数据特征的管网异常诊断方法的流程图如图 3 所示。

3 仿真结果与分析

为了验证所提方法的有效性, 本文采用 Matlab 完成对不同算例数据的仿真及分析。并且每个算例中均包含不同类型的异常样本及诊断过程的描述。所有测试结果均在一台配备 Intel Core i7-6700 CPU、8 GB 内存的计算机上实现。并且仿真数据选取某输油管网近 5 年的工况调整数据、放油测试以及调试时产生的信息异常数据。为了能够满足卷积神经网络对于样本的要求, 本文通过对现有物理异常数据的泛化实现样本数量的增加。所以最后训练和测试卷积神经网络的数据集共包含 26 500 个数据样本, 其中物理异常工况调整样本为 14 000 个, 不同程度地泄漏样本为 12 500 个。在管网输送油品到各个站场时, 管网调控中心及相应站场会对站内阀门、主输泵等设备进行操作, 由于此类操作有时会使管网信息数据产生类似泄漏情况的变化, 并且可能会传播到管网沿线其他站场, 所以在本文中定义由正常操作引起的数据改变为工况调整样本; 泄漏样本定义为管网放油测试以及程序调试时通过阀室及站内调节阀模拟产生的管网泄漏数据。训练样本和测试样本分别为总样本数量的 80% 和 20%。为了模拟实际现场情况, 检验模型的泛化能力, 在测试过程中, 将部分工况调整及泄漏样本作为未知样本进行测试, 其比例占全部测试样本的 20%。仿真参数为: 管网站场增维矩阵的维度为 30×60 , 站场信息增维协方差矩阵的最大特征向量的维度为 30×30 , $f_k = 10$, $t_n = 300$, 因此卷积神经网络的输

入图像维度是 60×300 , 最大允许数据中断时间 t_b 为 10 s 以及最大传输延迟长度 t^{thr} 为 3 s。异常阈值 γ 为 1.2, 管网信息增维协方差矩阵最大理论特征值 λ^{set} 为 0.1。卷积神经网络训练参数为初始化学率为 10^{-4} , 损失函数为交叉熵函数, 最大迭代次数为 20 次, 卷积层权重系数向量参数由卷积神经网络训练获得, 其余参数如图 2 所示。

3.1 算例一

为了说明本文提出的检测方法对信息传输中断异常的有效性, 以图 4 所示的管网某管道调试时的压力曲线为例进行说明。管道上、下游站场分别为管网中的第 2 个和第 3 个分输站场。管道下游站场通过式 (5) 进行分析计算, 当采样时刻达到 326 s 时, 计算第 317 s 至第 326 s 的增维矩阵 \hat{D}_2 和 \hat{D}_3 的数据可靠性, 发现下游站场数据满足式 (5), 因此判断下游站场发生信息传输中断异常。

此外, 选取如图 5 所示的管网站场 2 和站场 3 的异常压力曲线对信息传输错误异常进行说明。从图 5 中得知: 采样时间为 324 s 时, 下游站场压力由 1.748 MPa 下降至 1.746 MPa, 超过了平时压力波动范围。

基于数据特征融合的异常诊断方法首先通过式 (5) 验证增维矩阵 \hat{D}_2 和 \hat{D}_3 满足数据可靠性, 进而采用式 (6)~(8) 构建信息增维协方差矩阵 $S_{\hat{D}_2}$ 和 $S_{\hat{D}_3}$ 。通过图 6(a) 可知, 在采样时间 $t_s = 325$ s 时, 站场 3 的信息增维协方差矩阵最大值 $\lambda_{\hat{D}_3}^{\text{max}}$ 为 0.1294, 超过了式 (10) 所设定的阈值 0.12。因此, 判断下游站场 (站场 3) 发生异常情况。为进一步确定异常类型, 对上游站场 (站场 2) 的压力进行分析。当上游站场采样时间超过最大站场数据传输时间 $t_g = 61$ s 时, 通过图 6(a) 可知上游站场信息增维协方差矩阵的最大值 $\lambda_{\hat{D}_2}^{\text{max}}$ 在此时间段内并未超过阈值。因此根据图 3 流程图可知, 下游站场 (站场 3) 发生信息传输错误异常。

3.2 算例二

选取如图 7 所示的管网工况调整曲线对工况调整的物理异常进行检测。

按照第 2 节所示诊断流程, 首先分别对管网 6 个站场进行数据预分析, 然后再构建站场信息增维协方差矩阵 $S_{\hat{D}_1} - S_{\hat{D}_6}$ 。接着通过式 (10) 得到管网 6 个站场的信息增维协方差矩阵 $S_{\hat{D}_1} - S_{\hat{D}_6}$ 的最大特征值。管网站场最大特征值计算结果如图 8 所示。通过图 8 可知, 站场 4 的最大特征值 $\lambda_{\hat{D}_4}^{\text{max}}$ 在 800 s 时超过设置的阈值 0.12 ($\lambda_{\hat{D}_4}^{\text{max}}(800) = 0.1411$)。然后根据图 3 所示流程图判断其他站场的最大特征值曲线, 通过图 8 其余站场最大特征值曲线发现相邻站场均在不同时间发生最大特征值改变, 因此判断

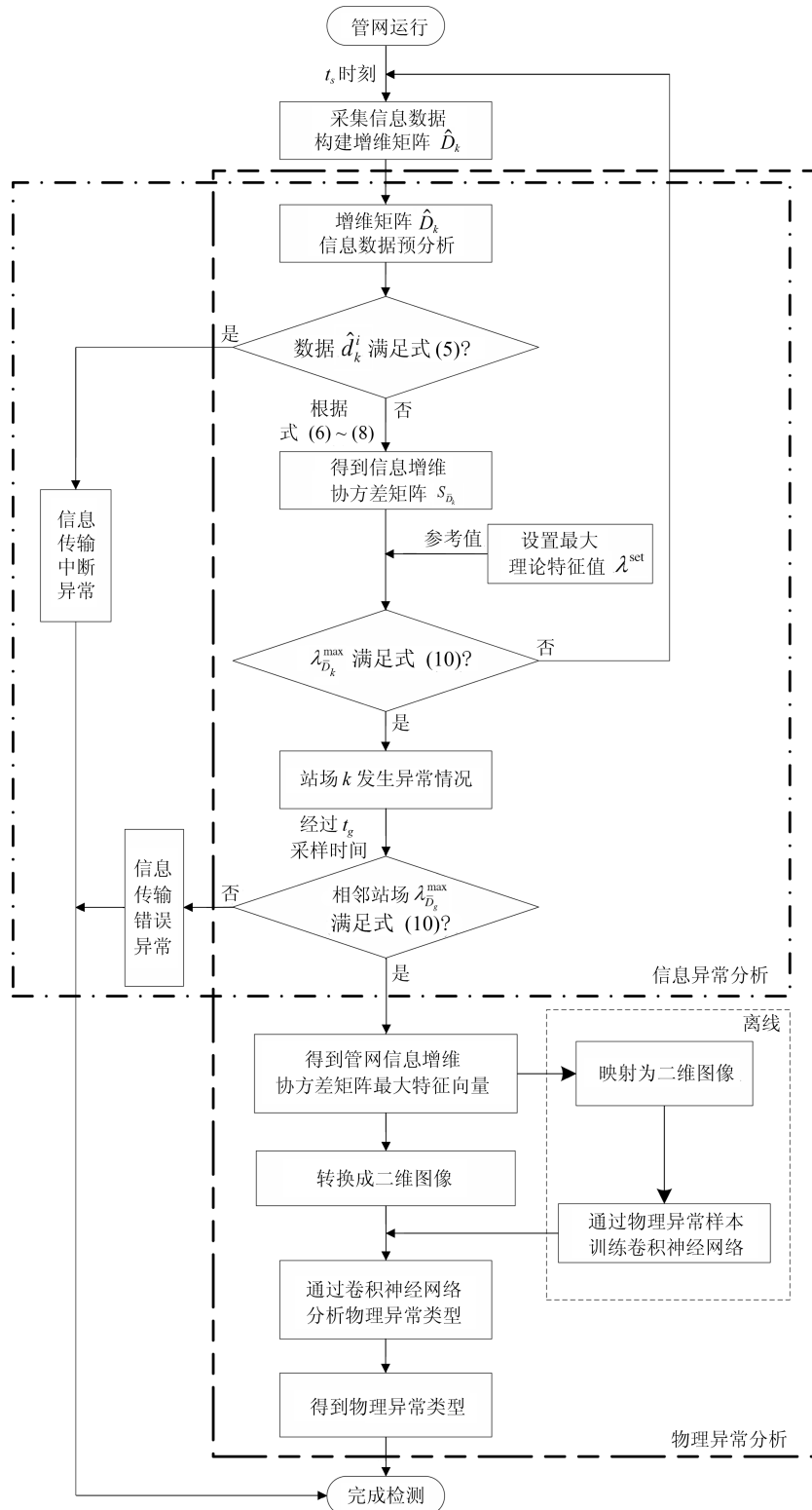


图 3 管网信息物理异常诊断流程图

Fig. 3 Flowchart of pipeline network for cyber-physical abnormality diagnosis

站场 4 发生物理异常.

然后选取管网站场信息增维协方差矩阵最大特征向量组成 60×300 维的矩阵. 如图 9 所示, 为了

判断物理异常类型, 首先将最大特征向量构成的矩阵通过灰度图映射为彩色图, 接着将其作为卷积神经网络的输入, 通过之前训练好的网络得到站场 4

发生的物理异常为工况调整. 卷积神经网络对于图 7 所示的异常压力曲线的分析结果与实际工况调整相符 (站场 4 进行输油调压).

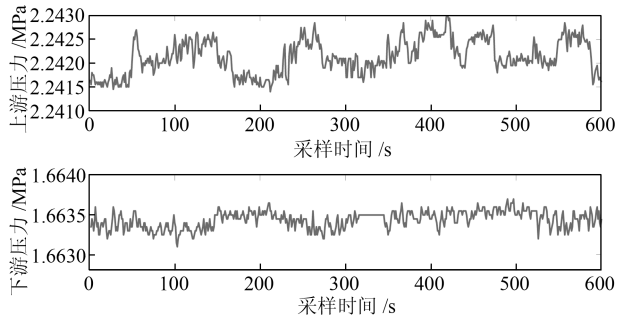


图 4 管道信息中断压力曲线

Fig. 4 Pipeline cyber interrupt pressure

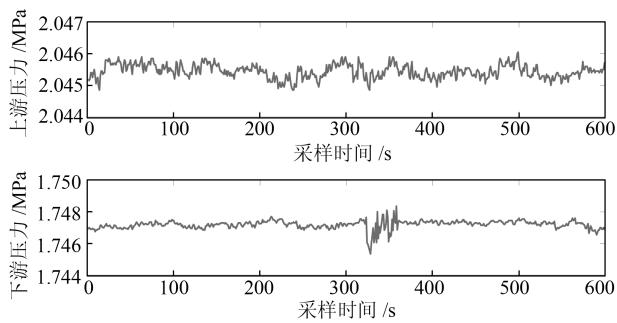


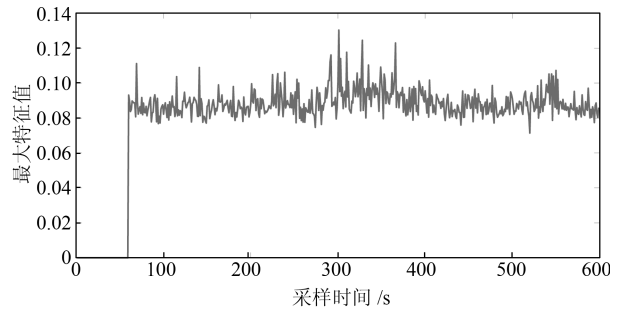
图 5 管道信息错误压力曲线

Fig. 5 Pipeline cyber error pressure

信息物理异常诊断方法除了需要对工况调整作出准确地识别以外, 还需要对管网站场发生的泄漏情况进行分析. 因此本算例通过站场 2 放油测试产生的压力曲线 (图 10) 进行泄漏异常诊断.

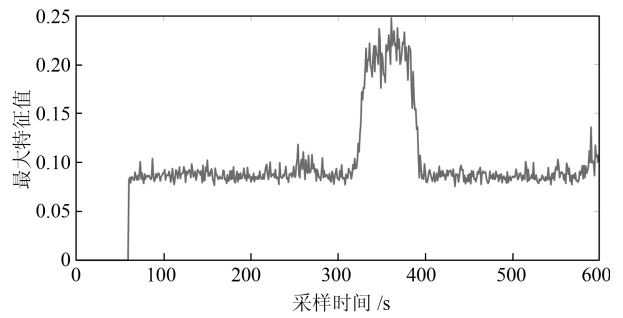
通过图 10 可知, 站场 2 压力曲线从 380s 附近开始下降, 并且从图 11 可以看出站场 2 的最大特征

值从 382s 开始超过了设定的阈值. 并且在 430s 时站场 3 的最大特征值呈现出上升趋势. 两个站场在最大站场数据传输时间内相继出现最大特征值上升的情况, 表明管网发物理异常.



(a) 上游压力最大特征值曲线

(a) Max eigenvalue curve of upstream pressure



(b) 下游压力最大特征值曲线

(b) Max eigenvalue curve of downstream pressure

图 6 管道压力最大特征值曲线

Fig. 6 Max eigenvalue curves of pipeline pressure

为了进一步判别该异常的类型, 在经过 251s 的采样后, 将管网站场信息增维协方差矩阵的最大特征向量转化为如图 12 所示的 RGB 图像. 利用训练好的卷积神经网络对该图像进行图像特征提取,

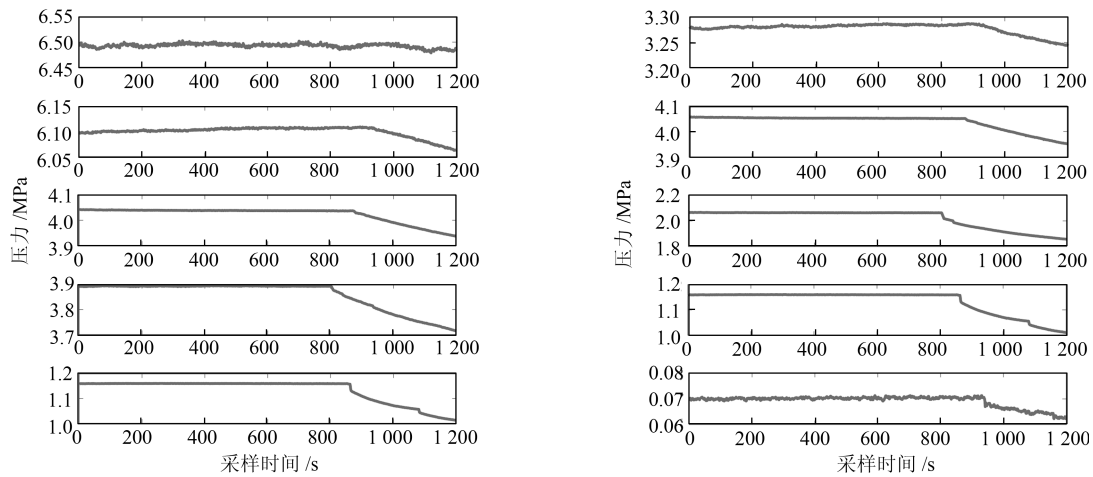


图 7 管网工况调整压力曲线

Fig. 7 Pipeline network operation adjustment pressure

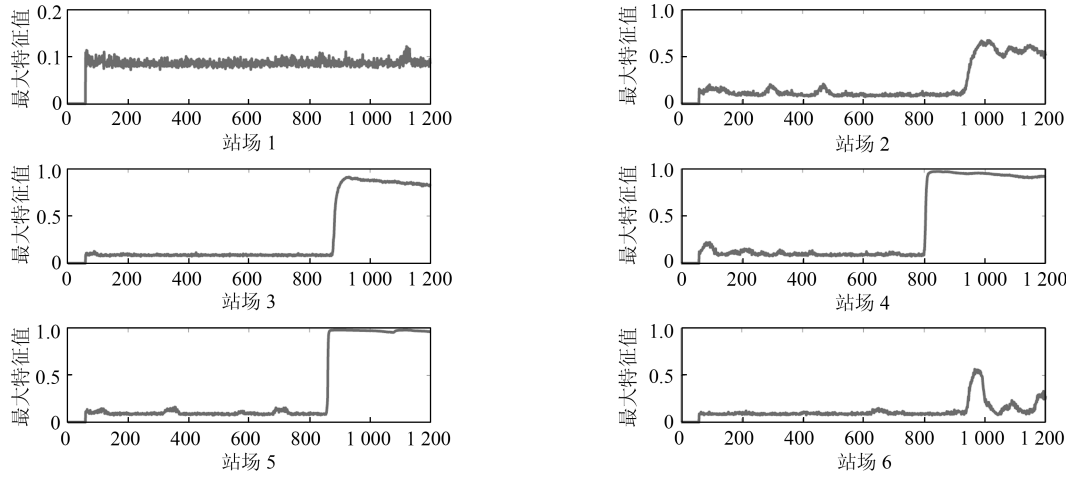


图 8 管网压力最大特征值曲线

Fig. 8 Max eigenvalue curves of pipeline network pressure

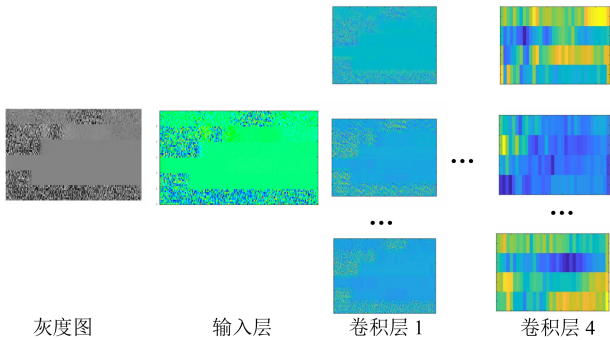


图 9 卷积神经网络输出特征图

Fig. 9 Output feature of CNN

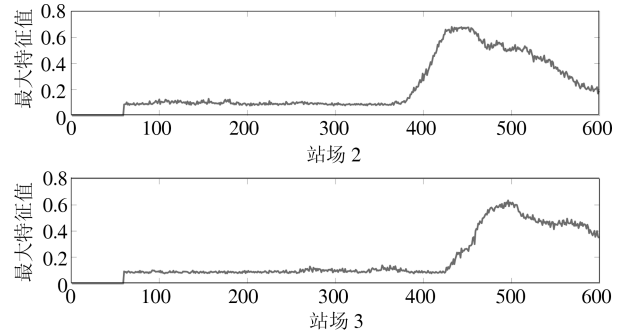


图 11 管道压力最大特征值曲线

Fig. 11 Max eigenvalue curves of pipeline network pressure

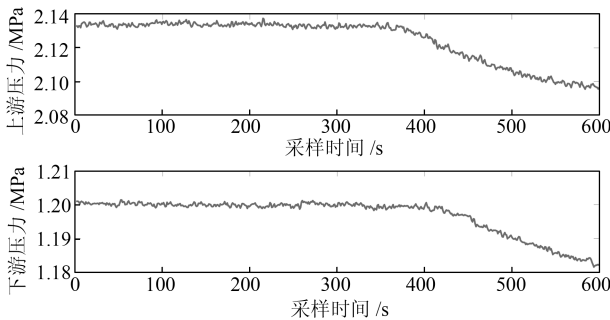


图 10 管道泄漏压力曲线

Fig. 10 Pipeline leakage pressure

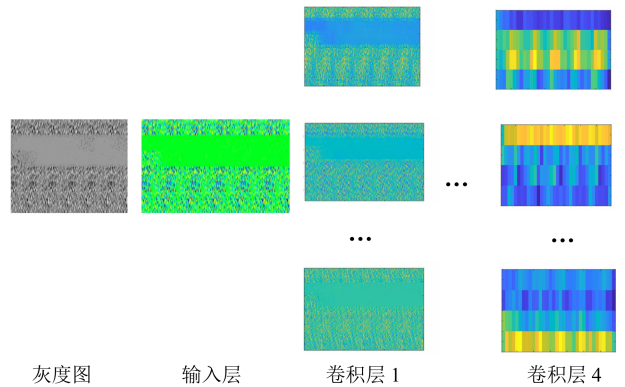


图 12 卷积神经网络输出特征图

Fig. 12 Output feature of CNN

卷积层 1 和卷积层 4 得到的图像特征向量如图 12 所示. 在完成特征提取后, 将一维特征向量输入到 Softmax 回归层, 最终得到站场 2 的物理异常类型为泄漏.

在管网物理泄漏异常诊断当中, 异常诊断方法需要识别出站场出现的缓慢泄漏情况. 因此, 选取如图 13 所示的站场 3 和站场 4 的压力变化曲线进行阐述. 该变化曲线是由阀室开阀放油测试产生的.

如图 13 所示, 在采样时刻 $t_s = 450s$ 附近, 上、下游站场压力开始缓慢下降并且压力下降值小于 $0.01 MPa$. 对管网 6 个站场的信息节点矩阵通过增维变量 \hat{f}_k 进行增维, 经过式 (5)~(10) 分析得到如图 14 所示的站场 3 和站场 4 的最大特征值曲线: 在 $480s$ 附近, 上、下游站场最大特征值均超过最大特

征值异常阈值 1.2, 表明站场 4 发生物理异常情况. 接着将管网 6 个站场的信息增维协方差矩阵最大特征向量转化为图 15 所示的输入层图像. 然后通过图 2 所示的卷积层提取图像特征, 接着将提取到的一维特征向量输入到图 2 所示的 Softmax 回归层中进行判别, 最后得到站场 4 的物理异常是由泄漏引起的.

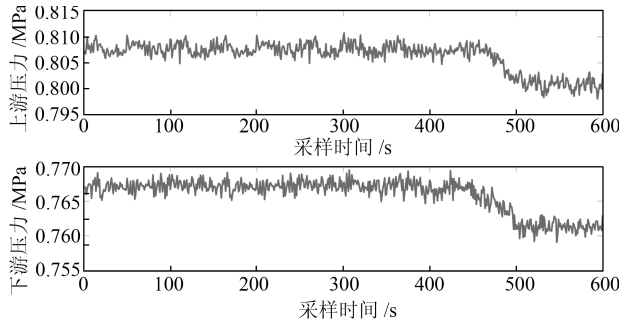


图 13 管道泄漏压力曲线
Fig. 13 Pipeline leakage pressure

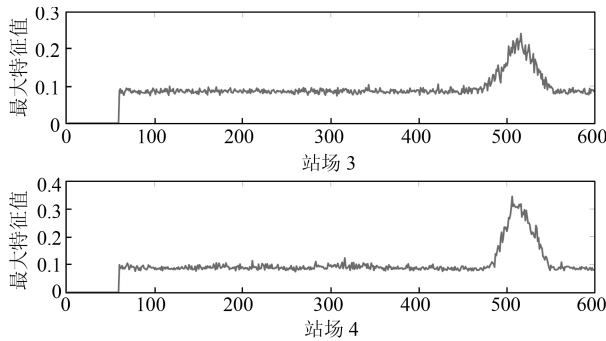


图 14 管道压力最大特征值曲线
Fig. 14 Max eigenvalue curve of pipeline network pressure

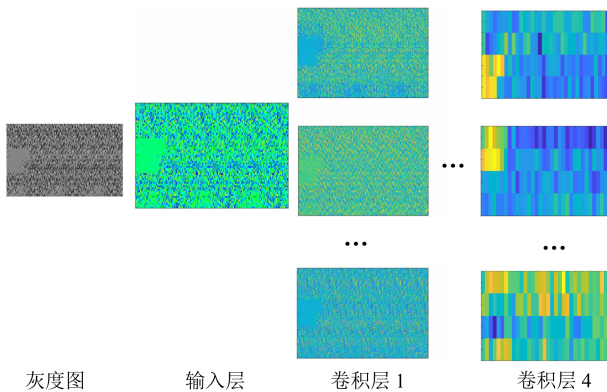


图 15 卷积神经网络输出特征图
Fig. 15 Output feature of CNN

综合算例二的结果可知, 通过数据特征融合的信息物理异常诊断方法能够有效地判断出不同的物理异常类型. 为了检验卷积神经网络对工况调整及泄漏异常的识别能力, 对测试样本进行 50 次实验,

在 50 次统计测试实验中, 全部测试样本中工况调整样本数量为 2800 个, 泄漏样本的个数为 2500. 并且统计结果如表 1 所示.

表 1 物理异常统计结果

Table 1 The statistical result of physical abnormality

待识别类型	正确分类数	错误分类数	精度 (%)
工况调整	2696	104	96.3
泄漏	2355	145	94.2

从表 1 可以看出工况调整样本的错误分类数为 104 个, 并且泄漏测试样本的错误分类数为 145 个. 两种物理异常的判断结果均在 95% 附近, 表明卷积神经网络对于不同物理异常样本的识别有着较好的分类效果. 如果在训练过程中包含全部物理异常类型的样本, 那么测试精度会达到 99% 以上.

3.3 算例三

为了进一步表明文中提出方法在实际应用中的优势, 将本文提出的方法和参考文献 [7]、反向传播神经网络 (Back propagation neural network, BPNN)、支持向量机 (Support vector machine, SVM) 的方法进行了对比.

在仿真对比中, 参考文献 [7] 采用发生物理异常的管道两端压力进行分析, 并且采用相同的样本对 BPNN、SVM 进行训练, 其中 BPNN 是由三层神经网络组成, 隐含层神经元的个数为 6, 并且隐含层的传递函数为 tan-sigmoid; SVM 的核函数为径向基函数, 并且通过网格搜索算法来优选 SVM 参数. 不同方法间的精度对比结果如表 2 所示.

表 2 不同方法物理异常精度对比

Table 2 Comparison of accuracy among different methods

诊断方法	灵敏度	特异度	准确率
本文	94.2%	96.3%	95.3%
文献 [7]	91.0%	88.6%	89.7%
BPNN	92.4%	91.8%	92.1%
SVM	93.0%	92.6%	92.8%

从表 2 中可以看出, 本文所提的方法在两种物理异常的识别中均优于其他三种方法. 由于成品油是在密闭管道内输送, 压力会随着距离不断衰减, 参考文献 [7] 主要针对单条管道进行异常检测, 所以该方法无法对管网沿线的操作进行有效地识别, 从而导致误报警出现. 因此, 该方法对于工况调整和泄漏的识别精度较低. BPNN 和 SVM 方法采用管网数据进行识别, 所以方法灵敏度、特异度以及准确率都比参考文献 [7] 要高. 虽然三种不同的方法均能够得到较好的效果, 但是灵敏度、特异度以及准确率均低

于本文所提的方法. 出现该现象的原因在于卷积神经网络的输入为管网信息增维协方差均值的最大特征向量图像, 而非手工提取的特征, 信息更加全面, 能够挖掘出隐藏在图像背后的深层特征.

此外, 为了对比不同方法间的计算量, 测试时间定义为方法得到一次样本结果所需的平均计算时间. 不同方法的训练时间及测试时间结果如表 3 所示.

表 3 不同方法物理异常计算时间对比
Table 3 Comparison of computing time among different methods

诊断方法	训练时间	测试时间
本文	438.7582 s	1.6701×10^{-2} s
文献 [7]	20.1538 s	2.1252×10^{-6} s
BPNN	16.9072 s	8.6301×10^{-7} s
SVM	1.4099 s	1.9650×10^{-5} s

从表 3 中可以看到, 本文所提方法的训练时间比另外三种方法的时间要长. 其原因在于本文所提的方法需要将信息增维协方差矩阵的最大特征向量转化成图像, 进而使用卷积神经网络进行训练和测试, 并且卷积神经网络的参数及结构较为复杂, 所以训练时间较长. 因为网络训练是离线进行的, 所以不同的训练时间对于在线实时检测来说是一样的. 虽然本文提出的方法是采用管网数据进行测试, 测试时间较长, 但是仍然没有超过 1 s.

因此, 本文提出的方法能够兼顾管网异常诊断精度以及实时性, 能够在较短的时间内准确地判断出运行情况, 从而确保管道的安全.

4 结论

针对管网不同类型的信息物理异常, 本文提出一种基于数据特征融合的管网信息物理异常诊断方法. 从信息网络角度将站场压力、流量作为信息数据进行分析, 通过对信息数据增维的方式突出信息数据的变化情况. 接着通过构建的信息增维协方差矩阵预分析以及矩阵最大特征值的变化情况得到不同信息异常类型的分析结果. 当多个站场最大特征值发生改变时, 将协方差矩阵最大特征向量映射的二维图像作为卷积神经网络的输入, 最终实现对站场工况调整和泄漏两种不同物理异常的分析. 通过大量实际数据样本的仿真研究表明, 本文提出的方法能够有效地检测管网信息物理异常情况.

此外为了能够提高测试精度, 可以采用增加训练样本以及多次多折交叉验证进行卷积神经网络的训练. 虽然本文提出的方法能够诊断出管网信息物理异常, 但是需要进一步确定泄漏源位置. 因此下一步的研究方向为继续研究深度学习机理, 进一步探索泄漏源位置定位方法.

References

- Wang Zhong-Jie, Xie Lu-Lu. Cyber-physical systems: a survey. *Acta Automatica Sinica*, 2011, **37**(10): 1157–1166 (王中杰, 谢璐璐. 信息物理融合系统研究综述. *自动化学报*, 2011, **37**(10): 1157–1166)
- Wen Jing-Rong, Wu Mu-Qing, Su Jing-Fang. Cyber-physical system. *Acta Automatica Sinica*, 2012, **38**(4): 507–517 (温景容, 武穆清, 宿景芳. 信息物理融合系统. *自动化学报*, 2012, **38**(4): 507–517)
- Park K J, Zheng R, Liu X. Cyber-physical systems: milestones and research challenges. *Computer Communications*, 2012, **36**(1): 1–7
- Li Jian, Chen Shi-Li, Huang Xin-Jing, Zeng Zhou-Mo, Jin Shi-Jiu. Review of leakage monitoring and quasi real-time detection technologies for long gas & oil pipelines. *Chinese Journal of Scientific Instrument*, 2016, **37**(8): 1747–1760 (李健, 陈世利, 黄新敬, 曾周末, 靳世久. 长输油气管道泄漏监测与准实时检测技术综述. *仪器仪表学报*, 2016, **37**(8): 1747–1760)
- Liu Jin-Hai, Feng Jian. Research on leak fault intelligent detection method for fluid pipeline based on fuzzy classification. *Chinese Journal of Scientific Instrument*, 2011, **32**(1): 26–32 (刘金海, 冯健. 基于模糊分类的流体管道泄漏故障智能检测方法研究. *仪器仪表学报*, 2011, **32**(1): 26–32)
- Liu Wei, Liu Hong-Zhao. Pipeline leak detection and location method based on structural similarity criteria. *Journal of Central South University (Science and Technology)*, 2017, **48**(1): 134–140 (刘伟, 刘宏昭. 基于结构相似度的管道泄漏检测定位法. *中南大学学报 (自然科学版)*, 2017, **48**(1): 134–140)
- Liu Jin-Hai, Zang Dong, Wang Gang. Leakage detection and location method of oil and gas pipelines based on Markov features. *Chinese Journal of Scientific Instrument*, 2017, **38**(4): 944–951 (刘金海, 臧东, 汪刚. 基于 Markov 特征的油气管道泄漏检测与定位方法. *仪器仪表学报*, 2017, **38**(4): 944–951)
- Kan Zhe, Lang Xian-Ming, Wang Xiao-Lei. Leakage location of branch pipeline based on cyber-physical system architecture. *Information and Control*, 2018, **47**(1): 22–28 (阚哲, 郎宪明, 王晓蕾. 基于信息物理系统架构分支管道泄漏定位. *信息与控制*, 2018, **47**(1): 22–28)
- Wang Wei-Ning, Wang Li, Zhao Ming-Quan, Cai Cheng-Jia, Shi Ting-Ting, Xu Xiang-Min. Image aesthetic classification using parallel deep convolutional neural networks. *Acta Automatica Sinica*, 2016, **42**(6): 904–914 (王伟凝, 王励, 赵明权, 蔡成加, 师婷婷, 徐向民. 基于并行深度卷积神经网络的图像美感分类. *自动化学报*, 2016, **42**(6): 904–914)
- Chang Liang, Deng Xiao-Ming, Zhou Ming-Quan, Wu Zhong-Ke, Yuan Ye, Yang Shuo, et al. Convolutional neural networks in image understanding. *Acta Automatica Sinica*, 2016, **42**(9): 1300–1312 (常亮, 邓小明, 周明全, 武仲科, 袁野, 杨硕, 等. 图像理解中的卷积神经网络. *自动化学报*, 2016, **42**(9): 1300–1312)
- Li Yong, Lin Xiao-Zhu, Jiang Meng-Ying. Facial expression recognition with cross-connect LeNet-5 network. *Acta Automatica Sinica*, 2018, **44**(1): 176–182 (李勇, 林小竹, 蒋梦莹. 基于跨连接 LeNet-5 网络的面部表情识别. *自动化学报*, 2018, **44**(1): 176–182)

- 12 Kang J, Park Y J, Lee J, Wang S H, Eom D S. Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems. *IEEE Transactions on Industrial Electronics*, 2018, **65**(5): 4279–4289
- 13 Feng J, Li F M, Lu S X, Liu J H, Ma D Z. Injurious or non-injurious defect identification from MFL images in pipeline inspection using convolutional neural network. *IEEE Transactions on Instrumentation and Measurement*, 2017, **66**(7): 1883–1892
- 14 Lu S X, Feng J, Zhang H G, Liu J H, Wu Z N. An estimation method of defect size from MFL image using visual transformation convolutional neural network. *IEEE Transactions on Industrial Informatics*, DOI: 10.1109/TII.2018.2828811
- 15 Yang Li-Jian, Cao Hui. Deep learning based weld and flange identification in pipeline. *Chinese Journal of Scientific Instrument*, 2018, **39**(2): 193–202
(杨理践, 曹辉. 基于深度学习的管道焊缝法兰组件识别方法. 仪器仪表学报, 2018, **39**(2): 193–202)
- 16 Han Chun-Yu, Huang Chun, Chen Fei, Nan Bing. Surge protection case for Dongying-Linyi parallel oil pipeline. *Oil & Gas Storage and Transportation*, 2008, **27**(2): 53–55
(韩春宇, 黄春, 陈飞, 南兵. 东临复线水击保护实例分析. 油气储运, 2008, **27**(2): 53–55)
- 17 Deng Zhong-Hua, You Dong-Qing, Guo Ye, Li Hong-Jun, Li Yue, Wen Feng, et al. Interruption protection of communication system in Shikong-Lanzhou Crude Oil Pipeline. *Oil & Gas Storage and Transportation*, 2017, **36**(5): 543–547
(邓忠华, 尤冬青, 郭晔, 李洪军, 李岳, 闻峰, 等. 石兰原油管道通信系统中断运行保护. 油气储运, 2017, **36**(5): 543–547)
- 18 He Zhao-Yang, Shang Yi, He Li-Ping, Li Chun, Yin Su-Na. Reasons and solutions of SCADA communication interruption in Mohe-Daqing Crude Oil Pipeline. *Oil & Gas Storage and Transportation*, 2014, **33**(5): 501–504
(何兆洋, 尚义, 何丽萍, 黎春, 殷素娜. 漠大原油管道 SCADA 通讯中断原因及应对措施. 油气储运, 2014, **33**(5): 501–504)
- 19 Bai Z D, Silverstein J W. *Spectral Analysis of Large Dimensional Random Matrices* (Second Edition). New York: Springer-Verlag, 2010.
- 20 Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278–2324
- 21 Zhou Fei-Yan, Jin Lin-Peng, Dong Jun. Review of convolutional neural network. *Chinese Journal of Computers*, 2017, **40**(6): 1229–1251
(周飞燕, 金林鹏, 董军. 卷积神经网络研究综述. 计算机学报, 2017, **40**(6): 1229–1251)

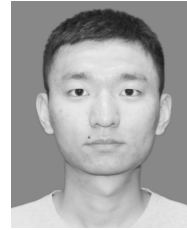


马大中 东北大学信息科学与工程学院副教授. 主要研究方向为故障诊断, 容错控制, 能源管理系统, 分布式发电系统、微网和能源互联网的优化与控制. 本文通信作者.

E-mail: madazhong@ise.neu.edu.cn
(**MA Da-Zhong** Associate professor at the College of Information Science

and Engineering, Northeastern University. His research interest covers fault diagnosis, fault-tolerant control, energy

management systems, and control and optimization of distributed generation systems, microgrids and energy internet. Corresponding author of this paper.)



胡旭光 东北大学信息科学与工程学院博士研究生. 主要研究方向为基于数据驱动的故障诊断, 信息物理系统的建模及优化控制.

E-mail: 1501004@stu.neu.edu.cn

(**HU Xu-Guang** Ph.D. candidate at the College of Information Science and Engineering, Northeastern University.

His research interest covers fault diagnosis based on data-driven, modeling and optimal control of cyber-physical system.)

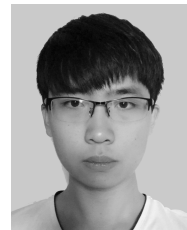


孙秋野 东北大学信息科学与工程学院教授. 主要研究方向为网络控制技术, 分布式控制技术, 分布式优化分析及其在能源互联网, 微网, 配电网等领域相关应用.

E-mail: sunqiuye@mail.neu.edu.cn

(**SUN Qiu-ye** Professor at the College of Information Science and Engineering, Northeastern University. His research interest covers

network control technology, distributed control technology, distributed optimization analysis and various applications in energy internet, microgrid, power distribution network.)



郑君 东北大学信息科学与工程学院硕士研究生. 主要研究方向为基于机器学习的综合能源系统故障检测与诊断.

E-mail: ZJ623928036@163.com

(**ZHENG Jun** Master student at the College of Information Science and Engineering, Northeastern University. His

research interest covers fault detection and diagnosis of integrated energy system based on machine learning.)



王睿 东北大学信息科学与工程学院博士研究生. 2016 年于东北大学获得电气工程及其自动化专业学士学位. 主要研究方向为能源互联网中分布式电源的协同优化及其电磁时间尺度稳定性分析.

E-mail: 1610232@stu.neu.edu.cn

(**WANG Rui** Ph.D. candidate at the College of Information Science and

Engineering, Northeastern University. He received his bachelor degree in electrical engineering and automation from Northeastern University in 2016. His research interest covers collaborative optimization of distributed generation and its stability analysis of electromagnetic timescale in energy internet.)