

文章编号:1001-5078(2025)03-0466-06

· 光纤传感及光通讯技术 ·

基于聚类神经网络的光纤网络节点异常识别算法

原娇杰, 焦梦甜, 赵杰文

(焦作大学 信息工程学院, 河南 焦作 454003)

摘要: 为了提高光纤网络节点异常的识别准确率与识别速度, 提出了一种基于聚类神经网络的节点异常识别算法。通过聚类计算完成输入数据的预分类, 解决传统分类识别算法容易陷入局部最优的问题。将预分类后的测试数据分组作为输入层, 并将聚类权值和聚类度作为隐藏层的加权系数, 提高异常信号的识别度。实验对光纤网络中 64 个 FBG 节点进行测试, 分别采用温度递变、重物撞击及周期振动模拟异常信号。对比实验结果显示, 三种异常信号均存在的混叠条件下, 本算法的识别准确率为 80.3%、92.8% 和 91.6%, 比不进行预分类的神经网络算法提升了约 20%。在四种测试情况下, 本算法的测试结果最优。对相同数据量测试时, 本算法的速度仅为 SVM 算法的 1/2, 验证了本算法具有更好的时效性。

关键词: 聚类神经网络; 预分类处理; 聚类度; 异常信号识别

中图分类号: TN929.11; TP391.41 文献标识码: A DOI: 10.3969/j.issn.1001-5078.2025.03.022

Fiber optic network node anomaly recognition algorithm based on clustering neural network

YUAN Jiao-jie, JIAO Meng-tian, ZHAO Jie-wen

(School of Information Engineering, Jiaozuo University, Jiaozuo 454003, China)

Abstract: To improve the accuracy and speed of identifying node anomalies in fiber optic networks, a node anomaly identification algorithm based on clustering neural networks is proposed in this paper. Firstly, the preclassification of input data is achieved through clustering calculation, which solves the problem of traditional classification and recognition algorithms easily falling into local optima. Then, the test data grouped after preclassification is used as the input layer, and the clustering weights and clustering degrees are used as the weighting coefficients of the hidden layer to improve the recognition of abnormal signals. Experiments are conducted on 64 FBG nodes in a fiber optic network, and the temperature increment, heavy impact and periodic vibration are used to simulate the anomalous signals, respectively. The results of the comparison experiments show that the recognition accuracy of this algorithm is 80.3%, 92.8%, and 91.6% under the condition of aliasing where all three types of abnormal signals exist, which is an improvement of about 20% over the neural network algorithm without preclassification. Therefore, the present algorithm has the optimal test results in the four test cases. For the same data volume test, the speed of this algorithm is only half of that of SVM algorithm, which verifies that this algorithm has better timeliness.

Keywords: clustering neural network; pre classification processing; cluster degree; abnormal signal recognition

基金项目: 河南省科技攻关计划项目 (No. 172102210390) 资助。

作者简介: 原娇杰 (1981 -), 女, 硕士, 讲师, 主要研究方向为计算机网络、光纤传感、算法研究等。E-mail: yzjj086@163.com

收稿日期: 2024-08-14; 修订日期: 2024-09-07

1 引言

随着光纤网络在居民终端的不断延伸,网络传输速度与数据稳定性成为了用户体验的重要评价标准^[1]。为了保证光纤网络的正常通信,及时发现异常节点并完成实时的网络路径优化就成为了一个新的研究热点^[2]。

光纤网络节点异常可以采用神经网络^[3]、聚类分析^[4]、免疫算法^[5]等多种手段进行检测。神经网络对于非线性问题具有一定的自适应能力。Sidelnikov^[6]将测试范围内多目标信号的频率作为神经元,目标检出率优于 87.1%,但类型识别率较低,主要应用于定性分析。刘云朋^[7]利用深度学习与遗传算法相结合,增强了对异常数据的检出效果,其测试偏差均值为 0.047,相比仅采用遗传算法或聚类算法提高了约 3 倍的精度。秦轶鞏^[8]通过提取异常信息的特征向量实现对不同异常节点的快速检出,该算法主要针对恶劣环境或外部攻击产生的强异常信号。耿德志^[9]采用数据深度挖掘的方式对光纤网络中异常节点进行挖掘,利用不同节点信息熵的特征分离异常节点,其精度可达 99.8%,该方法的优点是识别率高、速度快,但缺点是不能对异常节点类型进行分类。车丽娜^[10]提出了滑动窗口与置信度的异常识别算法,并对比了 2 种不同的异常识别算法的解释结果,该算法识别率优于 98%,具有较好的容错性。梁智洪^[11]设计了一种面向分布式光纤传感网络异常数据的自适应处理算法,主要针对 2 种主要异常类型进行识别,相关性系数优于 0.917。综上所述,现有文献对特定异常类型的识别率可以做的很高,但对多种异常信号的识别与分类仍存在较大困难。

针对光纤网络数据量大、多类别等特点,提高异常数据识别准确率与速度,本文提出了聚类神经网络的联用模式,通过聚类提高原始信号的数据粘性,再在神经网络反馈中引入特征因子,从而改善少量有效数据与异常数据的区分度。

2 聚类-神经网络模型设计

2.1 数据属性分类依据构建

为了对光纤网络中节点异常进行识别,就需要先对各种正常信号的特征进行分类,再由此类型集合对信号进行判定,当某种节点信号不属于任何一类时则认定其为异常信号。设光纤网络中通信的所有数据为 Q ,异常信息为 A ,异常信息的权值参数为 $c_m(t)$, t 为采样点, $m = 1, 2, \dots, K_m$ 。异常信息的聚类

权值为 $c_n(t)$, $n = 1, 2, \dots, K_n$ 。对 t 个采样数据进行归类计算,则其聚类权值^[12] 可以表示为:

$$c_n(t) = \sum_{m=1}^{K_m} \mu_{mn} c_m(t) + \sum_{n=1}^{K_n} \mu_{nn} c_n(t-1) \quad (1)$$

式中, μ_{mn} 为模糊隶属度, $1 \leq m \leq K_m, 1 \leq n \leq K_n$ 。设在异常信息集合 $A = (a_1, a_2, \dots, a_x)$ 中,任意异常信息的特征值 a_i 的密度指标有:

$$D_i = \sum_{n=1}^{K_n} \exp \left\{ -4 \left(\frac{x_m - x_n}{r} \right)^2 \right\} \quad (2)$$

式中, r 是异常信息 x_n 对应的邻域半径,该半径由所选区域内密度指标决定。当该区域中密度达到最大值 x_{\max} 时,异常信息的密度指标设为 D_{\max} ,则第 i 次的密度指标 D_i 和最大值的比值可以表示为:

$$\frac{D_i}{D_{\max}} = \exp \left\{ -4 \left(\frac{x_m - x_n}{r} \right)^2 \right\} \quad (3)$$

根据上式的关系可知, D_i/D_{\max} 的结果受异常信息权值和对应位置上的邻域半径的影响。其比值越小,聚类算法对初始数据的聚类效果越明显,其数据粘性越高,越有益于异类信息的分离。由于 x_m 和 x_n 是测试获得的,所以作为已知量代入(3)式即可,故邻域半径的设置是决定类别初始分类的关键。在隐含层中,异常聚类度 K 的计算是通过求解任意测试数据点与异常数据维度等相关数据计算得到的。设 f_{ij} 为第 i 个采样点 j 维上的聚类分布,则其异常数据的维度为 l ,总样本 Q 中的数据有 q 个,则异常数据的聚类度^[13] 为:

$$K = \left| \max(f_{ij}) \right| - \left| \sum_{j=1}^m \sum_{i=1}^q f_{ij} \right| \cdot (qf)^{-1} \left| \max(f_{ij}) \right|^{-1} \quad (4)$$

权值参数 c 与聚类度 K 作为参数对隐含层的神经网络分类的权重进行控制,从而使信息中具有相同特征的数据完成类型合并,最终在输出层中输出分类结果,其中,异常数据将被归为一类,从而实现异常数据识别。

2.2 模型设计

整个算法可以分为两个部分组成,一是聚类算法对输入层数据的分类,二是神经网络对分类后数据的交叉与重组。其中输入层数据包括回波波长和节点位置信息。聚类算法的作用是将具有相似属性的回波波长构成新的输入信息集合,并提供神经网络加权计算用的权重系数。而神经网络在隐藏层中利用聚类度与聚类权值将存在异常特征的波长与节

点位置检出。

输入层中波长变化的程度各不相同,所以分类过程中容易陷入局部最优或者分类错误等问题。故在输入层导入数据前先对初始数据进行聚类分析,从而获得不同类别数据的聚类权值,由此构建有分类倾向的输入层,由此可以大幅减少相关度低的原始数据的匹配计算。同时,根据异常数据特征结构的中心距为隐含层提供不同权重的赋值,从而进一步强化原始信号的分类效果。最终通过输出层完成异常信息的识别。输入层聚类后的神经网络如图 1 所示。

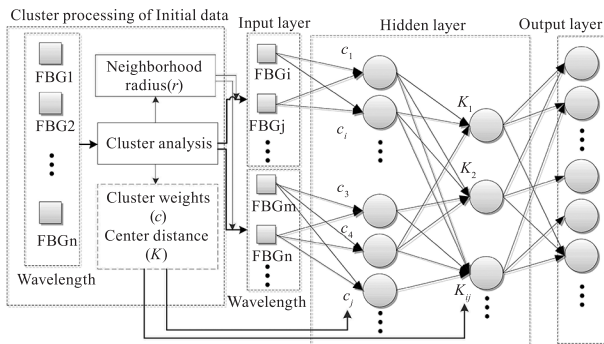


图 1 基于聚类神经网络的数据分类模型

Fig. 1 Data classification model based on clustering neural network

输入数据中核心数据是波长值,并且对应波长的附加信息是该节点的位置信息。首先,波长数据进入聚类算法后,根据邻域半径将原始数据分离成 n 个不同半径区间的集合,由此构建了具有类型特征的输入层。其次,则该聚类后的原始数据由邻域半径 r 构成了神经网络的输入层。该操作会解决同类数据分散造成的局域最优问题。同时,依据数据特征计算得到的聚类权值 c 和中心距 K 将作为隐含层数据赋值的权重系数,从而强化了神经网络在数据选择时的类型属性,提高了分类的准确度。最终,通过输出层完成异常数据的识别与分类。

2.3 异常数据的识别概率

在检测过程中,设测试点是异常数据并判别为异常数据的事件为 TP ,测试点是异常数据但判别为正常数据的事件为 FN ,测试点是正常数据并判别为正常数据的事件为 TN ,测试点是正常数据但判别为异常数据的事件为 FP ^[14]。当采用聚类神经网络算法后,测试点事件的发生概率发生了变化, TP 和 FN 主要受到异常信息的邻域半径 r 的影响,因为半径越小,对信息特征的限定越高,所以其是异常信息的概率越大,同理其判为正常数据的概率越小。而 TN 和 FP 主要受权值参数 c_n 和聚类度 K 的影响,因为权

值和聚类度共同决定了数据的粘性,当数据粘度高时,可以保证在边缘的正常数据依旧会被判定为正常数据,从而提高 TN 和降低 FP ^[15]。由此可知,引入本文算法后的概率可以表示为:

$$P'_{\text{accuracy}} = \frac{r \cdot TP + c_n K \cdot TN}{r \cdot TP + c_n K \cdot TN + (1 - c_n) K^{-1} \cdot FP + r^{-1} \cdot FN} \quad (5)$$

由式可知,通过调节 r 、 c_n 和 K 可以实现对异常数据识别概率的调节,从而增大异常数据被检出的概率及准确率。

3 算法实现

基于聚类-神经网络的异常识别算法主要是通过两部分数据处理实现的,一是图 2 中左则虚线框部分,采用聚类算法对原始波长进行特征分类;二是图 2 中右则虚线框部分,采用神经网络对分类后的输入层进行迭代。

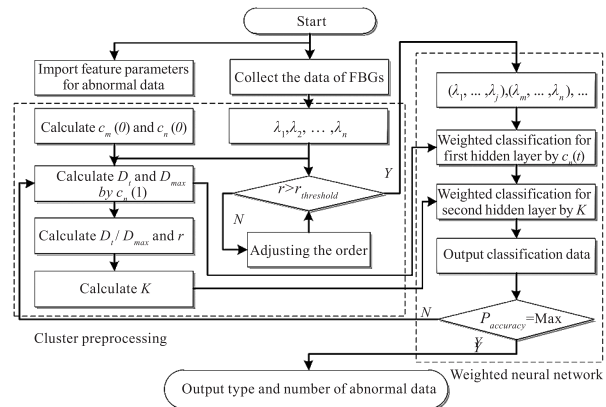


图 2 基于聚类神经网络算法的异常数据识别流程图

Fig. 2 Flow chart of abnormal data recognition based on clustering neural network algorithm

算法的具体实现步骤如下:

- (1) 以采样点 t 标记光纤网络中所有的数据点 Q ,根据已知可能存在的异常数据类型,设置异常数据的权值参数 $c_m(0)$ 和聚类权值 $c_n(0)$ 的初始值;
- (2) 导入光纤传感网络中所有节点的波长数据及对应的节点位置信息;
- (3) 根据聚类权值计算公式完成采样点 t 的聚类权值计算。设置异常信息数据集为 A ,根据异常数据的特征值计算该区域内任意点 t 的密度指标 D_i ,并提取其中的最大值 D_{\max} ;
- (4) 计算 D_i/D_{\max} 的比值,从而确定符合其范围的聚类邻域半径 r ;
- (5) 根据邻域半径 r 对光纤网络中 FBG 的回波数据进行分类,从而达到不同类别回波信号形成各自

的集合,即 $FBG_i-FBG_j, FBG_m-FBG_n, \dots$, 形成输入层;

(6) 将聚类权值 c_n 和聚类度 K 导入隐含层, 成为神经网络数据分类的加权项, 迭代所有的测试数据, 完成数据分类, 从而将异常数据归入一类;

(7) 采用公式(5) 计算 $P'_{accuracy}$, 判断异常数据识别概率是否达到最大值 Max , 如果没有达到则修改聚类算法权值参数, 重新计算聚类度并导入神经网络。如果达到最大值则完成数据输出;

(8) 通过调节 r, c_n 和 K 可以是算法适用于具体的节点异常类型, 并完成最大值寻优, 得到异常数据检出的最优解。

4 实验

系统由 ClarityPlus-C 型可调激光器(波长调谐范围 1528 ~ 1567 nm, 波长精度 ± 1 pm, 输出功率 10 dBm)、光纤耦合器、光纤隔离器、F1-155 型解调仪(解调波段范围 1530 ~ 1565 nm, 波长分辨精度 ± 1 pm)、8 通道 64 个 FBG 测试节点构成。设光纤网络中每个 FBG 需要的带宽为 0.5 nm, 则 64 个 FBG 的带宽为 32 nm, 故本系统选择的激光器带宽 39 nm 完全符合要求且允许一定的冗余。同理, 本系统选择的解调仪波长范围 35 nm 也是与测试实验中 FBG 的个数匹配的。激光器与解调仪的精度均选用 ± 1 pm, 可以直接匹配。激光器波长范围的选择决定了测试系统可以同时测量的 FBG 点位个数, 如果需要更多的测试点, 可以通过增大波长范围或者分时采集的方法实现。同理, 解调仪的波长范围与精度决定了 FBG 的测试范围与最小可识别量, 即为了识别异常信号的种类, 精度采用 ± 1 pm 可以获得不同异常信号造成的光谱分布差异。

通过在节点上施加外力或增温, 使其回波数据产生异常波动, 模拟光纤网络中的异常节点。测试数据分别用局部异常因子算法(LOF, Local Outlier Factor)、支持向量机(SVM, Support Vector Machine, SVM)、人工神经网络(ANN, Artificial Neural Network)和本算法(CNN, Cluster Neural Network)进行分析。

4.1 异常信号提取

本模拟实验中采用 4 种常见的异常信号:A. 测试数据整体缓慢的大幅波长偏移(采用在 FBG1 位置上提供 20 °C 到 80 °C 的缓慢加温模拟);B. 测试数据快速大幅波长偏移(采用在 FBG2 位置上提供敲击冲击, 100 g 砝码 10 cm 高度坠落);C. 测试数据周期性震荡(采用在 FBG3 位置上提供 10 Hz 的

振动源噪声);D. 测试数据同时存在 A、B、C 三种异常干扰的波长偏移(在 FBG4 位置上提供以上三种噪声源), FBG 的响应信号如图 3 所示。

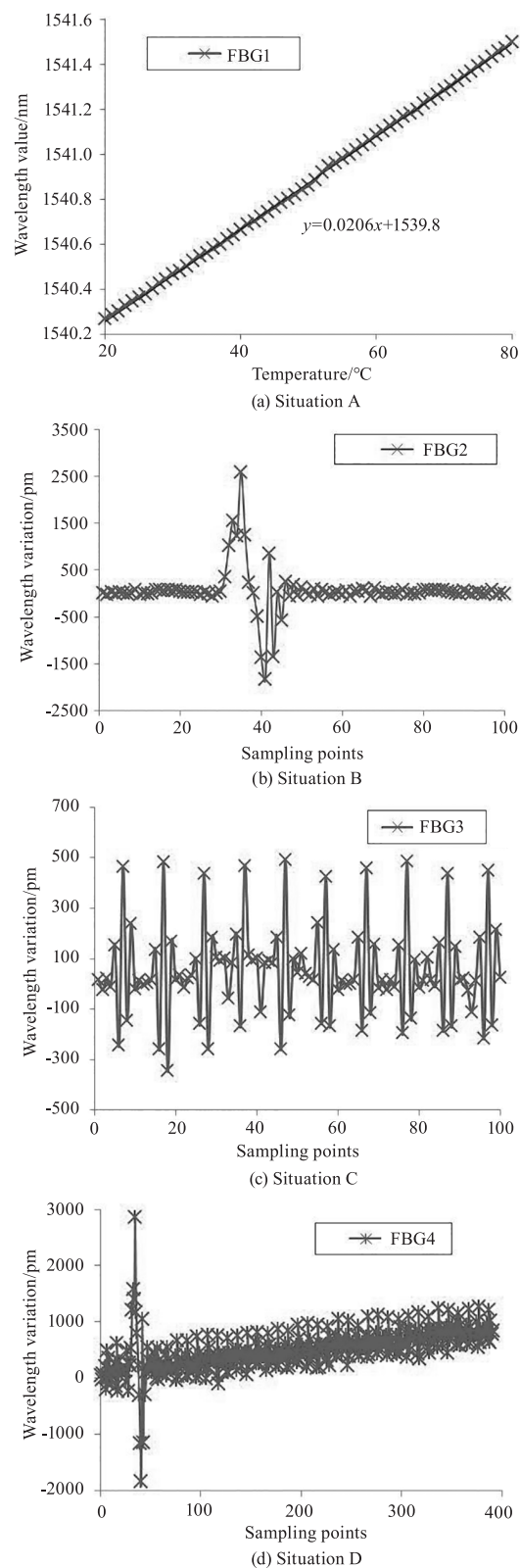


图3 四种异常信号的测试响应曲线
Fig. 3 Test response curves for four types of abnormal signals

如图 3(a) 所示,当测试节点位置的温度发生异常时,模拟过程采用温度从 20 °C 增加至 80 °C,对应位置 FBG 的中心波长发生偏移,从 1540.269 nm 增大至 1541.498 nm,总偏移量为 1.229 nm,线性度可达 0.9998。说明温度的变化可以被有效识别,并且通过其线性关系测试精度优于 ± 1 °C。如图 3(b) 所示,当光纤网络中存在突然性非周期信号干扰时,波长偏移随干扰源时长与强度而变化,相同条件下的冲击干扰测试 10 次后求平均,其最大值和最小值分别为 2584 pm 和 -1826 pm。如图 3(c) 所示,光纤节点上的周期性噪声同样具有明显的波长偏移波动,其振幅差均值为 668 pm。如图 3(d) 所示,三种节点数据异常在混叠情况下,其信号波形将包含所有干扰信号的特征,当分别采用低通滤波、谐波采样及阈值识

别的方式依旧可以完成干扰源的检测。由于篇幅限制,所以图 3(b) 和 (c) 仅给出了两种典型情况,在不同质量、高度、频率等测试中发现,其响应曲线基本一致,故仅展示了此种类型的多次均值结果,并由此完成不同异常类型混叠信号的识别,即图 3(d) 的情况。

4.2 监测结果对比

分别采用四种算法对光纤网络中的异常数据进行检测,采样频率 100 Hz,单次监测获取 1 min 的测试数据,每种情况测试 5 次取平均值,a,b,c,d 四种情况分别对应四种异常数据,共存在 7 种组合形式,组合方式以表 1 中“Situation”表示。获得不同异常信号混叠条件下,节点异常的识别准确率,结果如表 1 所示。

表 1 不同测试条件下四种算法对三个异常事件的测试结果均值

Tab. 1 Mean test results for three abnormal events by four algorithms in different testing conditions

Comparison parameters		Situation						
		a	b	c	a,b	a,c	b,c	a,b,c
$P'_{\text{accuracy}}/\%$ (for a)	LOF	97.7	-	-	42.5	43.6	-	23.5
	SVM	98.8	-	-	78.7	79.5	-	75.8
	ANN	98.4	-	-	59.8	62.8	-	61.4
	CNN	98.3	-	-	79.2	83.5	-	80.3
$P'_{\text{accuracy}}/\%$ (for b)	LOF	-	99.5	-	85.6	-	84.9	81.8
	SVM	-	99.8	-	97.2	-	91.5	88.6
	ANN	-	99.8	-	96.3	-	85.6	72.5
	CNN	-	99.8	-	97.6	-	94.7	92.8
$P'_{\text{accuracy}}/\%$ (for c)	LOF	-	-	97.9	-	61.3	79.8	53.6
	SVM	-	-	99.2	-	92.2	90.1	89.4
	ANN	-	-	98.7	-	84.2	73.6	71.2
	CNN	-	-	98.8	-	93.8	91.9	91.6

由表 1 可知,当仅存在单一类型节点异常时,四种算法的检出概率接近。但当存在两种及以上干扰源时,LOF 的检测概率会大幅降低。以最复杂的三种干扰源同时存在的情况 D 为例,CNN 效果最优,分别为 80.3 %、92.8 % 和 91.6 %,相比仅采用 ANN 的识别率提升了约 20 %,略高于 SVM 算法。验证了算法的可行性。

4.3 耗时分析

除了算法识别准确率外,算法的时效性对网络异常数据检测同样重要,才采用相同的测试数据源的基础上,对比四种算法的耗时关系如图 4 所示。

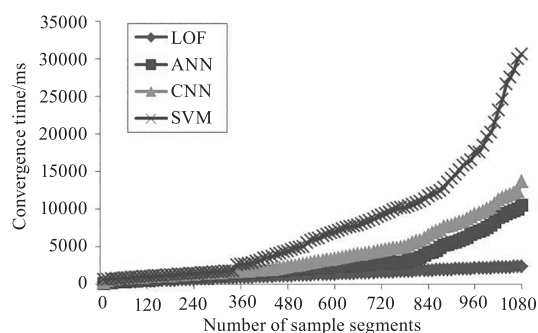


图 4 随数据总量增加的三种算法的收敛时间

Fig. 4 Convergence times of three algorithms with increasing total data volume

LOF 算法最快,SVM 算法最费时,CNN 比 ANN 略

微费时。以 500 个数据段和 1000 个数据段为例, CNN 算法的耗时分别为 2497 ms 和 10245 ms, 而识别准确率相似的 SVM 算法则需要 4825 ms 和 20154 ms。综合识别率与收敛速度, CNN 更具优势。

5 结论

本文设计了一种聚类神经网络的节点异常识别算法。通过对输出的节点数据进行预分类处理, 从而增加了相似数据的数据粘性, 降低了局部最优的风险。同时, 采用聚类权值对隐藏层进行加权补偿, 从而强化了异常信号的特征表达, 提高了其可识别概率。实验采用的 3 种异常中心波长波动模拟节点异常状态, 并通过与三种常见算法的对比, 验证了本算法的可行性与优势。

参考文献:

- [1] Peng Z Q, Jian J N, Wen H Q, et al. Distributed fiber sensor and machine learning data analytics for pipeline protection against extrinsic intrusions and intrinsic corruptions [J]. *Optics Express*, 2020, 28(19): 27277 - 27292.
- [2] Jiang J P, Liu F, Wang H H, et al. Lateral positioning of vibration source for underground pipeline monitoring based on ultra-weak fiber Bragg grating sensing array [J]. *Measurement*, 2021, 172: 108892.
- [3] Di Sante R. Fibre optic sensors for structural health monitoring of aircraft composite structures: recent advances and applications [J]. *Sensors*, 2015, 15(8): 18666 - 18713.
- [4] Wu H J, QAINY, Zhang W, et al. Feature extraction and identification in distributed optical-fiber vibration sensing system for oil pipeline safety monitoring [J]. *Photonic Sensors*, 2017, 7(4): 305 - 310.
- [5] Ren L, Jiang T, Jia Z G, et al. Pipeline corrosion and leakage monitoring base optical fiber sensing technology [J]. *Measurement*, 2018, 122: 57 - 65.
- [6] Sidelnikov O, Redyuk A, Sygletos S. Qualization performance and complexity analysis of dynamic deep neural networks in long hal transmission systems [J]. *Optics Express*, 2018, 26(25): 32765 - 32776.
- [7] Liu Yunpeng, Huo Xiaoli, Liu Zhichao. Optical fiber network abnormal data detection algorithm based on deep learning [J]. *Infrared and Laser Engineering*, 2021, 50(6): 280 - 285. (in Chinese)
刘云朋, 霍晓丽, 刘智超. 基于深度学习的光纤网络异常数据检测算法 [J]. *红外与激光工程*, 2021, 50(6): 280 - 285.
- [8] Qin Yihui, Ma Tao. Identification method of node abnormal state in distributed optical fiber strain sensing network [J]. *Laser Journal*, 2022, 43(6): 211 - 215. (in Chinese)
秦轶翬, 马涛. 分布式光纤应变传感网络节点异常状态识别方法 [J]. *激光杂志*, 2022, 43(6): 211 - 215.
- [9] Geng Dezhi, Gong Haixiao. Research on deep mining method of abnormal node data in large-scale optical fiber network [J]. *Laser Journal*, 2023, 44(4): 124 - 128. (in Chinese)
耿德志, 宫海晓. 大规模光纤网络异常节点数据深度挖掘方法研究 [J]. *激光杂志*, 2023, 44(4): 124 - 128.
- [10] Che Lina, Ren Xiuli. Anomaly detection algorithm based on sliding windows and confidence for WSN [J]. *Chinese Journal of Sensors and Actuators*, 2023, 36(11): 1801 - 1807. (in Chinese)
车丽娜, 任秀丽. 基于滑动窗口和置信度的无线传感器网络异常检测算法 [J]. *传感技术学报*, 2023, 36(11): 1801 - 1807.
- [11] Liang Zhihong, Deng Kaiwen, Ma Yunlong, et al. An adaptive post-processing algorithm for strain reading anomalies in distributed optical fiber sensors [J]. *Acta Optica Sinica*, 2024, 44(1): 343 - 354. (in Chinese)
梁智洪, 邓凯文, 马云龙, 等. 分布式光纤传感器应变读数异常的自适应后处理算法 [J]. *光学学报*, 2024, 44(1): 343 - 354.
- [12] Jiang Hong, Zeng Qinglong, Li Jiacheng, et al. Intrusion event identification for fiber optic perimeter security based on VMD and MFE [J]. *Laser & Infrared*, 2023, 53(7): 1073 - 1080. (in Chinese)
江虹, 曾庆龙, 李家成. 基于 VMD 与 MFE 的光纤周界安防入侵事件识别 [J]. *激光与红外*, 2023, 53(7): 1073 - 1080.
- [13] Li T D, Fan C Z, Li H, et al. Nonintrusive distributed flow rate sensing system based on flow-induced vibrations detection [J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 7001808.
- [14] Chi Q, Luo Y, Ren J, et al. Review on fiber-optic sensing in health monitoring of power grids [J]. *Optical Engineering*, 2019, 58(7): 072007 - 072011.
- [15] Du Y K, Shang Y, Wang C, et al. Intrusive and non-intrusive microflow measurement based on distributed optical fiber acoustic sensing [J]. *Measurement*, 2023, 210: 112513.