

光纤入侵信号的特征提取与识别算法

曲洪权, 宫殿君*, 张常年, 王彦平

北方工业大学电子信息学院, 北京 100144

摘要 为了对分布式光纤上的入侵信号类型进行准确识别, 提出了一种基于集合经验模态分解(EEMD)结合随机向量函数链接(RVFL)神经网络的光纤入侵信号的特征提取与识别算法。算法步骤为: 对采集到的光纤入侵信号作预处理操作, 包括最小-最大规范化处理和利用 db3 小波去除信号的低频噪声; 采用 EEMD 方法对入侵信号进行分解, 得到 5 组本征模态函数(IMF); 计算各 IMF 分量的能量占比, 并依据方差分析法筛选出 3 组特征向量; 将特征向量送入 RVFL 神经网络进行训练并对入侵信号进行识别。实验结果显示: 该方法能正确识别不同入侵信号的类型, 具有较高的准确率。

关键词 光纤光学; 光纤预警系统; 特征提取与识别; 集合经验模态分解; 随机向量函数链接神经网络

中图分类号 TN911.6

文献标识码 A

doi: 10.3788/LOP56.130601

Feature Extraction and Recognition Algorithm for Fiber Intrusion Signals

Qu Hongquan, Gong Dianjun*, Zhang Changnian, Wang Yanping

School of Electronic and Information Engineering, North China University of Technology, Beijing 100144, China

Abstract A feature extraction and recognition algorithm for fiber intrusion signals is proposed based on ensemble empirical-mode decomposition (EEMD) coupled with a random vector-function linked (RVFL) neural network to accurately identify the type of intrusion signal on a distributed optical fiber. The proposed algorithm starts with the preprocessing for the collected fiber intrusion signals, including minimum-maximum normalization processing and the removal of low frequency noise using the db3 wavelet. Then, the intrusion signals are decomposed by the EEMD to obtain five groups of intrinsic mode functions (IMF). Subsequently, the energy ratio of each component of the IMF is calculated, and three feature vectors are filtered using the analysis of variance. Finally, the feature vectors are sent into the RVFL neural network to be trained for the completion of the signal recognition. The experimental results validate that the proposed algorithm can accurately distinguish between different intrusion signals with high recognition rate.

Key words fiber optics; fiber pre-warning system; feature extraction and recognition; ensemble empirical mode decomposition; random vector function linked neural network

OCIS codes 060.2370; 150.1135; 070.4340; 060.4250

1 引言

光纤预警系统(OFPS)是一种应用光纤传感的预警系统。OFPS 具有稳定性好、准确率高的特点, 被广泛应用于检测管道泄漏、人为或机械等入侵行为^[1-2]。目前普遍使用马赫-曾德尔光纤干涉(M-Z)和相位敏感光时域反射(Φ -OTDR)两种光纤传感器

采集光纤入侵信号^[3]。由于 M-Z 光纤传感器在实际使用中的需求标准过高, 该系统的使用场景和适用范围都受到限制^[4-5]。基于瑞利散射的 Φ -OTDR 的传感原理是基于光纤中传输光的相位变化, 故其灵敏度极高。 Φ -OTDR 不但能够满足绝大多数场景的使用条件, 而且非常适用于微弱入侵事件的检测^[6-7]。因此本文选择基于 Φ -OTDR 光纤传感器系

收稿日期: 2018-12-12; 修回日期: 2018-12-14; 录用日期: 2019-01-24

基金项目: 国家自然科学基金(61571014); 北京自然科学基金(4172017)

* E-mail: 769353964@qq.com

统对入侵信号进行采集。

OFPS 主要由检测和识别两部分组成。检测部分可以检测入侵信号并对其进行精确定位。目前研究人员已经研究出了很多成熟的检测算法^[8-10]。在检测到入侵信号之后,识别部分可通过一定的算法判断入侵信号的类型,并依据信号的危险程度采取不同的措施。如何判断信号的类型是本系统的难点和重点。因此本文对入侵信号的识别算法重点展开研究,主要包括特征提取与分类两大部分。

针对光纤中入侵信号在时域上区分度不高的特点,通常选择频域方法提取光纤入侵信号的特征。小波多层分解和经验模态分解两种信号分解方式被广泛应用于信号的去噪与特征提取^[11-12]。文献[13-14]中均使用小波函数分解重构的方法提取光纤中光信号的频域特征,但对于如何选择小波基函数的种类、分解层数等没有给出说明和解释。经验模态分解法(EMD)是一种没有基函数表达式、能够自适应地处理信号的方法,非常适合处理绝大多数非线性、非平稳的光纤入侵信号。文献[15]中使用 EMD-AWPP (adaptive wavelet packet processing) 的方法对数据进行时频局部化处理,提高了识别准确率。信号中可能存在间断信号、脉冲干扰和噪声,导致 EMD 分解后的本征模态函数(IMF)分量中包含不同时间尺度的特征成分,降低了特征提取的精度。集合经验模态分解法(EEMD)在 EMD 的基础

上,引入了白噪声扰动并对信号作集合平均,从而避免了尺度混合问题,使得分解后的各分量在物理上保持唯一性。因此本文选择 EEMD 作为特征提取方法,将计算所得的各本征模态函数的能量比作为特征向量。随机向量函数链接(RVFL)神经网络是一种结构简单、可随机分配权重和偏置的随机神经网络。RVFL 神经网络具有随机分配的特点,时间消耗极少,数据不易陷入局部极小值。相较于支持向量机(SVM),RVFL 神经网络可实现多类样本的分类,非常适合多类小样本光纤入侵信号的识别^[16]。

2 特征提取与识别

对光纤入侵信号特征的提取及识别算法的整体流程如图 1 所示,其中 E_{sum} 为 IMF 分量的总能量。该流程包括以下步骤:1) 采集所需的光纤入侵信号,通过计算机显示时域信号;2) 对入侵信号进行预处理;3) 对入侵信号进行 EEMD 分解,求取 5 层本征模态函数;4) 分别计算各 IMF 的能量占比 E_1, E_2, \dots, E_5 (第 1 层 IMF 的能量占比表示为 E_1 , 第 2 层 IMF 的能量占比表示为 E_2 , 以此类推),并进行特征筛选,最后确定特征向量;5) 运用 RVFL 神经网络对不同能量占比的入侵信号进行识别。其中, RVFL 模型参数 β 是通过训练部分确定的,即对有标签的入侵信号进行训练,确定模型参数。通过训练参数 β 进行测试数据实验,最后完成对入侵信号的识别分类。

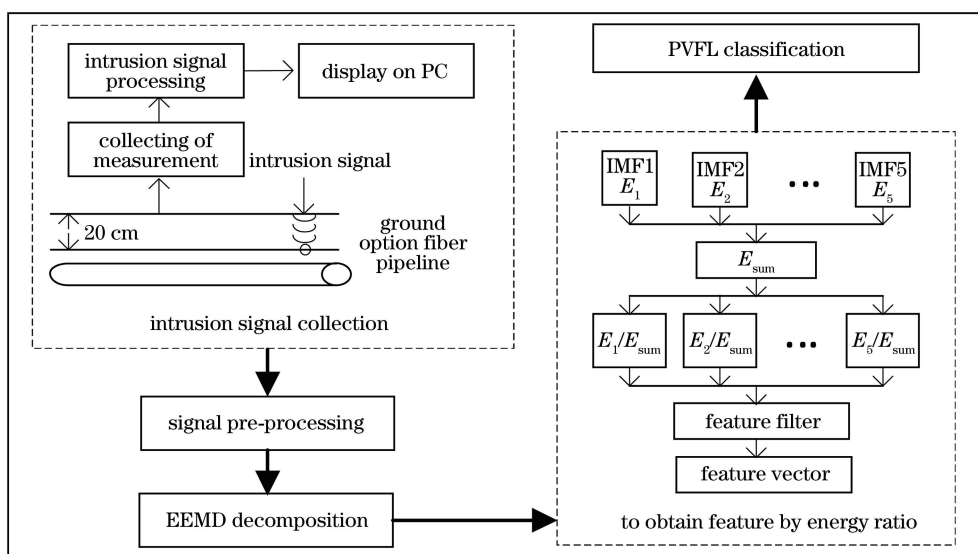


图 1 光纤入侵信号特征提取及识别算法流程图

Fig. 1 Processing flow chart of feature extraction and recognition for fiber intrusion signals

2.1 EEMD 分解与能量比提取特征

EEMD 是对 EMD 的改进算法,该方法解决了 EMD 中出现的模态混叠问题。EEMD 算法的分解

过程以 EMD 为基础^[13],EEMD 的分解步骤如下:

1) 向原始信号 $x(t)$ 中加入随机高斯白噪声序列 $n_i(t)$, 得到新的加噪信号:

$$x_1(t) = x(t) + n_i(t), \quad (1)$$

式中: t 为时间。

2) 对加噪信号 $x_1(t)$ 进行 EMD 分解, 得到 IMF 分量 $c_{ij}(t)$, 其中 $c_{ij}(t)$ 为原始信号经过加入 i 次高斯白噪声后, 由 EEMD 分解所得到的第 j 个 IMF 分量。

3) 重复以上两个步骤 n 次, 得到 n 组 IMF 分量 $c_{ij}(t)$ 。

4) 对 $c_{ij}(t)$ 求集总平均结果, 即

$$C_j(t) = \frac{1}{n} \sum_{i=1}^n c_{ij}(t), \quad (2)$$

求得的 $C_j(t)$ 即为最终的 IMF 结果。

一般添加高斯白噪声的幅值为原信号标准差的 10%~40%, 分解次数可根据信号自行设定。以敲击信号为例, 取信号标准差的 10%, 迭代 100 次, 分解 5 层的结果如图 2 所示。由图 2(a) 和图 2(b) 可以看出, EMD 分解中 IMF3 和 IMF4 出现了模态混叠现象, 而 EEMD 分解有效地抑制了模态混叠。

能量是反映不同信号特征的重要物理量, 因此可对光纤入侵信号进行频域分析后提取其能量占比作为信号特征。经过 EEMD 分解后, 得到光纤入侵

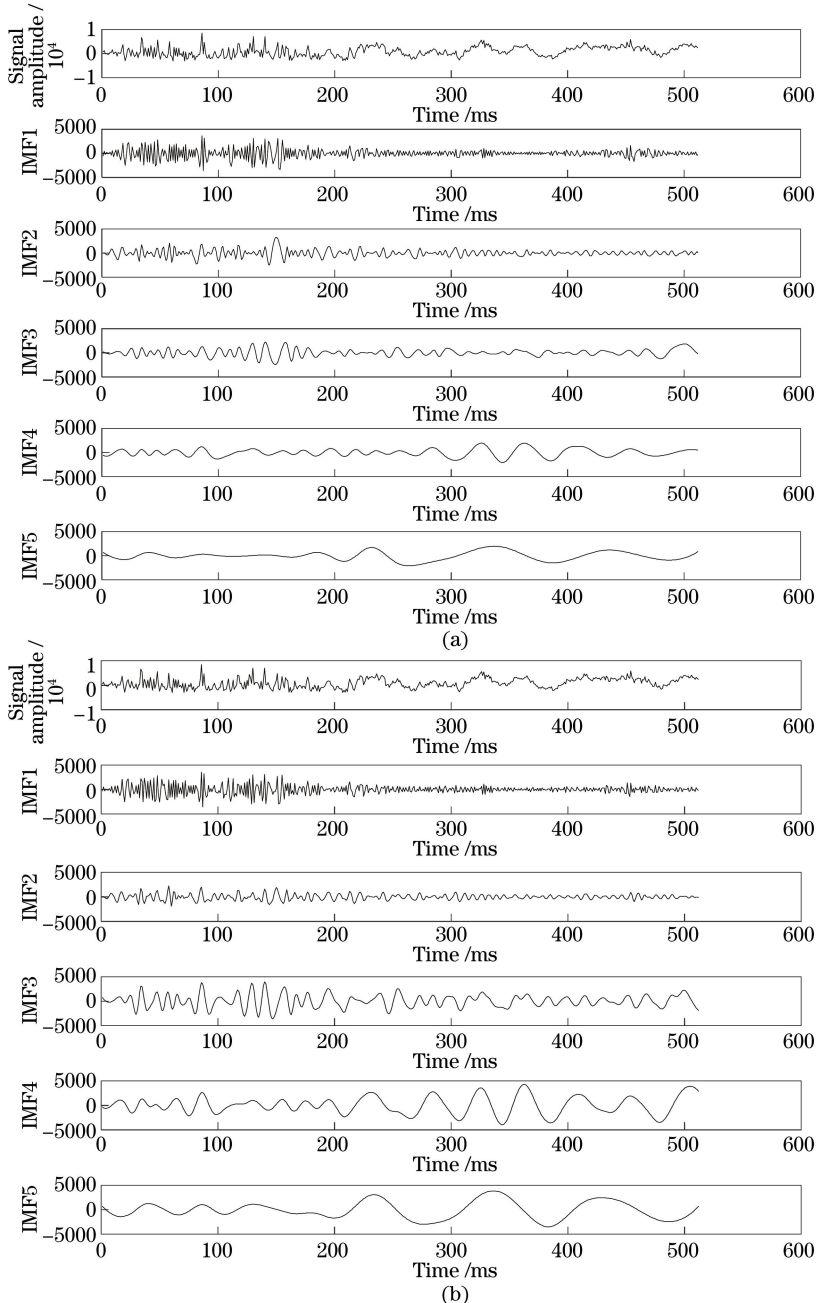


图 2 分解结果。(a) EMD 分解; (b) EEMD 分解

Fig. 2 Decomposition results. (a) EMD decomposition; (b) EEMD decomposition

信号的 5 组 IMF 分量, 每组分量的能量可表示为

$$E_j = \sum_{n=1}^N [C_j(t)]^2, j = 1, 2, \dots, 5, \quad (3)$$

式中: $C_j(n)$ 为第 j 组 IMF 分量; N 为信号的长度。所有 IMF 分量的总能量为

$$E_{\text{sum}} = \sum_{j=1}^5 E_j. \quad (4)$$

因此, 每组 IMF 分量的能量占比为

$$\Omega_j = E_j / E_{\text{sum}}, j = 1, 2, \dots, 5. \quad (5)$$

最后, 特征样本 $\Omega = \{\Omega_1, \Omega_2, \Omega_3, \Omega_4, \Omega_5\}$ 。经过方差分析后, 发现 5 维特征中只有 3 维特征对识别有明显作用。为了减少特征冗余, 忽略其中差异

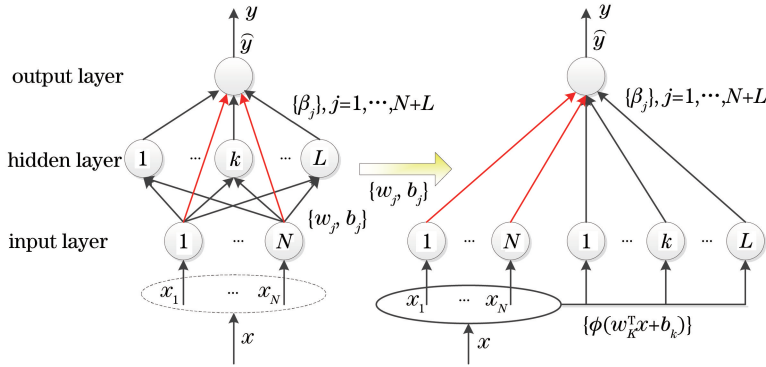


图 3 RVFL 神经网络的结构

Fig. 3 Structure of RVFL neural network

$$\hat{y} = \sum_{k=1}^L \beta_k \phi(\mathbf{w}_k^T \mathbf{x}_k + b_k), k = 1, 2, \dots, L, \quad (6)$$

式中: \mathbf{w}_k 和 \mathbf{x}_k 是第 k 个节点的权重向量和输入向量; b_k 是输入层至隐层的阈值; β_k 是隐层至输出层的权重; ϕ 为 sigmoid 激活函数; L 为隐层节点数。

网络训练的的目的是得到隐层至输出层的输出权值 $\boldsymbol{\beta}$, 使用标准正则化最小二乘法求取权值 $\boldsymbol{\beta}$, 含有正则化项的平方误差准则函数为

$$E_{\beta} = \min_{\beta} \frac{1}{2} \sum_{i=1}^N [f(x_i) - \sum_{k=1}^L \beta_k \phi(x_i; \mathbf{w}_k, b_k)]^2 + \frac{\lambda}{2} \|\boldsymbol{\beta}\|_2^2 = \min_{\beta} \frac{1}{2} \|\mathbf{Y} - \boldsymbol{\Phi}\boldsymbol{\beta}\|_2^2 + \frac{\lambda}{2} \|\boldsymbol{\beta}\|_2^2, \quad (7)$$

$\lambda > 0,$

式中: λ 为一个常数; $f(x_i)$ 为第 i 个神经元的网络输出; x_i 为第 i 个神经元的输入。

对目标函数求梯度, 得到最终的权值 $\boldsymbol{\beta}$ 为

$$\boldsymbol{\beta} = (\boldsymbol{\Phi}^T \boldsymbol{\Phi} + \lambda \mathbf{I})^{-1} \boldsymbol{\Phi}^T \mathbf{Y}. \quad (8)$$

3 实验与分析

3.1 数据预处理

为了评估该算法的有效性, 本课题组在北京市

较少的两维特征, 将剩余特征作为特征向量送入神经网络进行识别。

2.2 RVFL 神经网络

RVFL 神经网络结构简单、时间消耗少, 非常适合小样本光纤入侵信号的分类与识别。与其他基于梯度下降的神经网络相比, RVFL 神经网络建立了输入层与输出层之间的联系, 输入层与隐层间的权值 w_j 和阈值 b_j 是随机幅值并在学习过程中被固定。RVFL 神经网络在训练中仅仅需要学习隐层与输出层间的权值 β_j , 因此该神经网络学习效率很高。RVFL 神经网络的结构如图 3 所示。

网络的输出可表示为

门头沟区上苇甸村试验田采集了敲击、过车和小跑 3 类光纤入侵信号, 光缆埋敷在试验田地地表下 20 cm 处。收集后截取实验样本, 每类入侵信号共采集 100 个样本, 每份样本采集时长为 512 ms。为了减弱噪声和幅度干扰, 对样本进行最小-最大规范化处理, 并采用阶数为 3 的紧支集正交小波 db3 小波去除入侵信号的低频噪声, 处理结果如图 4 所示。

3.2 EEMD 分解能量比结果与特征筛选

使用 EEMD 分解法分别处理敲击、过车和小跑 3 类光纤入侵信号, 得到 3 种信号的 5 组 IMF, 并计算其各自的能量占比, 结果如图 5 所示。经过观察, 得到的 5 维特征中只有 3 维特征具有显著区别, 其他两维均为冗余特征, 因此使用方差分析法对所有数据进行验证, 方差分析法输出的 p 值为衡量维度之间差异大小的指标。结果发现: 敲击信号经方差分析后 $p = 1.55 \times 10^{-190} < 0.01$, 第 1 维特征相较于其他 4 维特征差异显著; 过车信号经方差分析后 $p = 3.41 \times 10^{-170} < 0.01$, 第 4 维特征相较于其他 4 维差异显著, 小跑信号经方差分析后 $p = 1.38 \times 10^{-203} < 0.01$, 第 5 维相较于其他 4 维差异显著。实验结果如图 6 所示, 其中横坐标为特征维度, 纵坐标

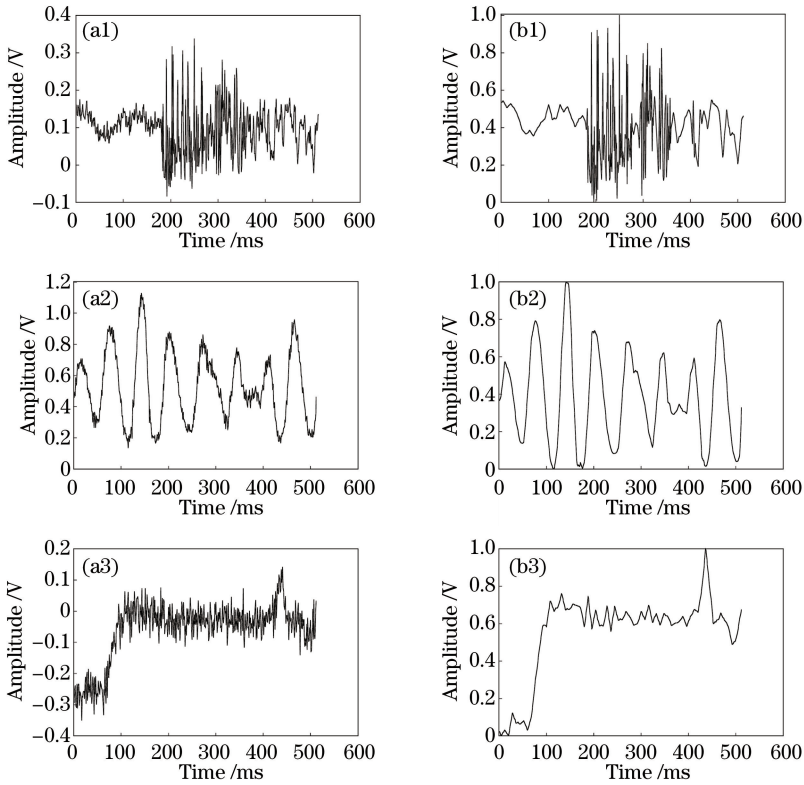


图 4 原始入侵信号预处理。(a1)(b1)敲击信号;(a2)(b2)过车信号;(a3)(b3)小跑信号

Fig. 4 Preprocessing for original intrusion signals. (a1)(b1) Tapping signal; (a2)(b2) vehicle signal; (a3)(b3) running signal

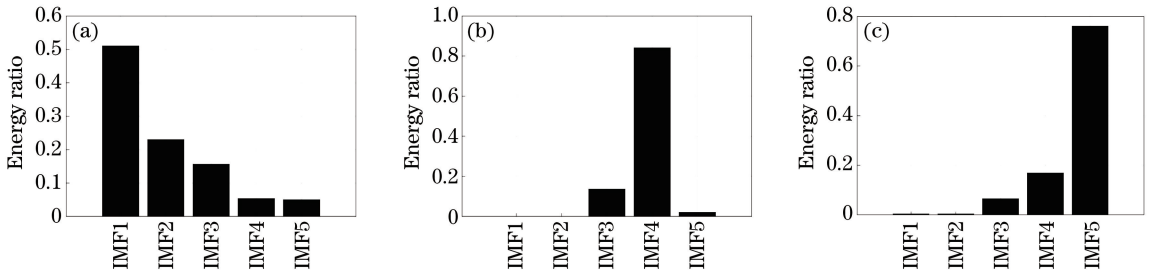


图 5 能量比结果。(a)敲击信号;(b)过车信号;(c)小跑信号

Fig. 5 Results of energy ratio. (a) Tapping signal; (b) vehicle signal; (c) running signal

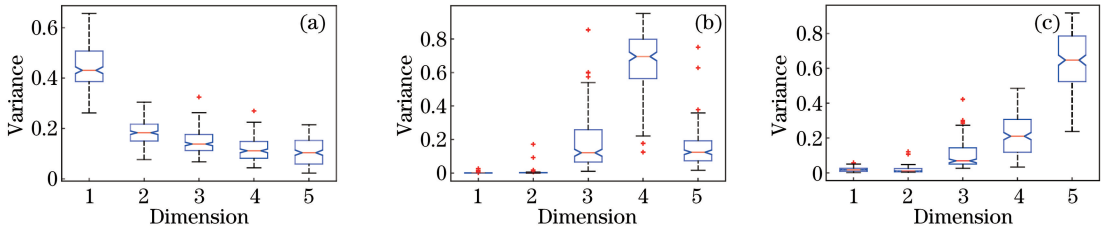


图 6 方差分析结果。(a)敲击信号;(b)过车信号;(c)小跑信号

Fig. 6 Variance analysis results. (a) Tapping signal; (b) vehicle signal; (c) running signal

为特征向量的数值大小。因此选择第 1,4,5 维特征作为特征向量送入神经网络,图 7 为所选特征向量的可视化结果。

箱线图盒两端为该维度的四分位数,盒长度为四分位数极差(IQR),盒中线为中位数,盒外两条虚线延伸到最小和最大观测值,观测值不超过 IQR 值

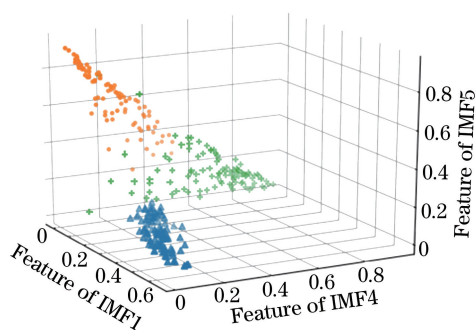


图7 特征向量可视化结果

Fig. 7 Result of feature vector visualization

的1.5倍,超过部分由离群加号表示。通过箱线图,可直观地表现出数据分布的散度,便于提取有效的特征。

由图7可知,第1维特征对于识别敲击信号(三角表示)起重要作用;第4维特征对于识别过车信号(十字表示)起重要作用;第5维特征对于识别小跑信号(圆圈表示)起重要作用。经过方差分析筛选出的3维特征向量可有效地表征敲击、过车和小跑光纤入侵信号并实现对其正确分类。

4 RVFL神经网络识别结果分析

4.1 模型参数选择

对于RVFL神经网络模型,通过大量实验研究发现,其隐层节点数为100时误差最低。网络中未确定的参数 λ, ω, b 的取值可通过对比实验确定,考虑不同 λ, ω, b 对识别误差的影响,其中 ω, b 分别为权值和偏置。3种信号标准误差的平均值如表1所示。

表1 不同的 λ, ω, b 对模型误差的分析Table 1 Model error analysis under different λ, ω , and b

λ	$\omega, b \in [-2, 2]$		$\omega, b \in [-100, 100]$		$\omega, b \in [-200, 200]$		$\omega, b \in [-400, 400]$	
	ω	b	ω	b	ω	b	ω	b
0.005	0.278	0.165	0.198	0.323				
0.050	0.292	0.176	0.202	0.210				
0.500	0.297	0.162	0.172	0.176				
5.000	0.312	0.186	0.189	0.203				

由表1可知,当 λ 取值为0.500,权值 ω 和偏置 b 的取值范围均为 $[-100, 100]$ 时,识别结果的误差率为0.162,误差率达到最低。

4.2 识别结果分析

在使用RVFL神经网络进行训练时,设置输入层与隐层之间的权值 ω 和偏置 b 的取值范围均为 $[-100, +100]$, λ 值为0.500,隐含层节点数为100。敲击、过车和小跑3类光纤入侵信号的训练样本各

150个,测试样本同样各150个。其中,训练、测试样本中敲击信号标签设置为0,过车信号标签设置为1,小跑信号标签设置为2。在训练后构建的模型中输入测试样本,得到的信号分类结果如图8所示。

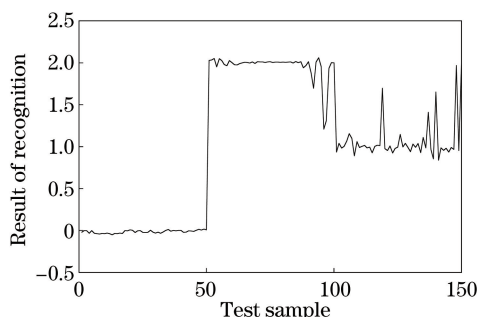


图8 测试样本识别结果

Fig. 8 Recognition result of test sample

图8的横坐标为测试样本数量,纵坐标为RVFL神经网络的输出值。根据输出结果设置识别阈值:识别结果大于-0.5且小于0.5,识别为敲击信号;识别结果大于0.5且小于1.5,识别为小跑信号;识别结果大于1.5且小于2.5,识别为过车信号。通过上述方法计算得到敲击、过车和小跑3类光纤入侵信号的识别率分别为100%、98%、92%,从而得到入侵信号的识别率为96.7%。敲击信号特征明显,最容易识别。过车信号和小跑信号有极个别样本混淆,但仍达到了较高的识别准确率。由此可见,通过EEMD分解、计算能量比并使用RVFL神经网络对光纤入侵信号的类型进行识别的方案是有效可靠的。

4.3 对比实验分析

将上述训练样本、测试样本输入常用的样本分类器BP(back propagation)神经网络,再次对3类光纤入侵信号进行分类识别。将BP神经网络所得结果与RVFL神经网络的结果进行对比,结果如表2所示。

表2 识别结果对比

Table 2 Comparison of recognition results

Recognition method	Recognition result /%
RVFL neural network	96.7
BP neural network	94.7

由表2可知,采用所提特征提取算法结合RVFL神经网络可以有效地提取敲击、过车和小跑光纤入侵信号的特征,并具有较高的识别准确率。

5 结论

提出一种基于EEMD-RVFL神经网络的光纤

入侵信号特征提取与识别算法,该算法通过 Φ -OTDR 采集光纤入侵信号,并对其作进一步处理。处理步骤为:对预处理后的光纤入侵信号作 EEMD 分解得到 5 组 IMF;分别计算能量比得到 5 维特征,筛选其中特征显著的 3 维特征作为特征向量,完成特征提取工作;将训练样本送入 RVFL 神经网络,搭建模型并用测试样本进行测试,检测分类识别效果。实验结果显示,对于敲击、过车和小跑 3 类光纤入侵信号,识别准确率达到 96.7%,验证了所提算法的可行性与有效性。

参 考 文 献

- [1] Bi F K, Feng C, Qu H Q, *et al.* Harmful intrusion detection algorithm of optical fiber pre-warning system based on correlation of orthogonal polarization signals[J]. *Photonic Sensors*, 2017, 7(3): 226-233.
- [2] Liang W, Lu L L, Zhang L B. Coupling relations and early-warning for "equipment chain" in long-distance pipeline [J]. *Mechanical Systems and Signal Processing*, 2013, 41(1/2): 335-347.
- [3] Yang Y, Feng H, Wang Z H, *et al.* Application and development of distributed optical fiber sensing technology in pipeline detection [J]. *Electro-Optic Technology Application*, 2016, 31(6): 1-9, 76.
杨洋, 封皓, 王宗和, 等. 光纤传感技术在管道检测中的应用与发展[J]. *光电技术应用*, 2016, 31(6): 1-9, 76.
- [4] An Y, Jin S J, Feng X, *et al.* Optical fiber pipeline security pre-warning system based on coherent Rayleigh scattering [J]. *Journal of Tianjin University*, 2015, 48(1): 70-75.
安阳, 靳世久, 冯欣, 等. 基于相干瑞利散射的管道安全光纤预警系统[J]. *天津大学学报*, 2015, 48(1): 70-75.
- [5] Sheng Z Y, Zhang X Y, Wang Y P, *et al.* Feature extraction and linear classification for fiber vibration signals[J]. *Journal of Optoelectronics • Laser*, 2018, 29(7): 760-768.
盛智勇, 张新燕, 王彦平, 等. 光纤振动信号特征提取及线性分类方法[J]. *光电子 • 激光*, 2018, 29(7): 760-768.
- [6] Shang Y, Wang C, Wang C, *et al.* Distributed vibration sensing of perimeter security based on space difference of Rayleigh backscattering[J]. *Infrared and Laser Engineering*, 2018, 47(5): 0522001.
尚盈, 王晨, 王昌, 等. 采用后向瑞利散射空间差分的周界安防分布式振动监测[J]. *红外与激光工程*, 2018, 47(5): 0522001.
- [7] Pang F F, Liu H H, Wang T Y. A review of distributed fiber sensors based on phase-sensitive optical time domain reflectometer [J]. *Journal of Nanjing University of Information Science & Technology (Natural Science Edition)*, 2017, 9(2): 130-136.
庞拂飞, 刘奂奂, 王廷云. 相位敏感光时域反射光纤传感技术的研究综述[J]. *南京信息工程大学学报(自然科学版)*, 2017, 9(2): 130-136.
- [8] Qu H Q, Ren X C, Bi F K, *et al.* Two-level detection algorithm of two-dimensional for vibration signals detected by optical fiber [J]. *Acta Optica Sinica*, 2015, 35(10): 1006004.
曲洪权, 任学丛, 毕福昆, 等. 光纤振动信号的二维二级检测算法[J]. *光学学报*, 2015, 35(10): 1006004.
- [9] Qu H Q, Wang T Q, Bi F K, *et al.* Harmful intrusion detection method based on two-dimensional K-S test in reconfigured background for optical fiber pre-warning system[J]. *Journal of Jishou University (Natural Science Edition)*, 2017, 38(1): 19-23.
曲洪权, 王天琦, 毕福昆, 等. 基于重构背景二维 K-S 检验的有害入侵光纤预警[J]. *吉首大学学报(自然科学版)*, 2017, 38(1): 19-23.
- [10] Qu H Q, Zheng T, Bi F K, *et al.* Vibration detection method for optical fibre pre-warning system [J]. *IET Signal Processing*, 2016, 10(6): 692-698.
- [11] Sha Y Y, Xi L X, Zhang X G, *et al.* Polarization mode dispersion measurement based on wavelet threshold denoising [J]. *Chinese Journal of Lasers*, 2018, 45(11): 1106006.
沙宇洋, 席丽霞, 张晓光, 等. 基于小波阈值去噪的偏振模色散测量[J]. *中国激光*, 2018, 45(11): 1106006.
- [12] Pan P, Xi L X, Zhang X G, *et al.* Experimental research on polarization mode dispersion measurement based on empirical mode decomposition [J]. *Chinese Journal of Lasers*, 2018, 45(1): 0106002.
潘潘, 席丽霞, 张晓光, 等. 基于经验模态分解的偏振模色散测量实验研究[J]. *中国激光*, 2018, 45(1): 0106002.
- [13] Qu H Q, Wang X X, Bi F K, *et al.* Optical fiber vibration recognition based on wavelet reconstruction and time-space features [J]. *Journal of Jishou University (Natural Science Edition)*, 2017, 38(2): 36-41.

- 曲洪权, 王笑笑, 毕福昆, 等. 基于小波重构与时空二维特征的光纤振动识别[J]. 吉首大学学报(自然科学版), 2017, 38(2): 36-41.
- [14] Wang J P, Hao Z, Zhu C H. Research on vibration signal recognition of optical fiber perimeter based on phase space reconstruction [J]. Journal of Hefei University of Technology (Natural Science), 2017, 40(5): 643-648.
- 王建平, 郝钊, 朱程辉. 基于相空间重构的光纤周界信号识别算法研究[J]. 合肥工业大学学报(自然科学版), 2017, 40(5): 643-648.
- [15] Zhang Y J, Liu W Z, Fu X H, *et al.* An extraction and recognition method of the distributed optical fiber vibration signal based on EMD-AWPP and HOSA-SVM algorithm [J]. Spectroscopy and Spectral Analysis, 2016, 36(2): 577-582.
- 张燕君, 刘文哲, 付兴虎, 等. 基于 EMD-AWPP 和 HOSA-SVM 算法的分布式光纤振动入侵信号的特征提取与识别[J]. 光谱学与光谱分析, 2016, 36(2): 577-582.
- [16] Igel'nik B, Pao Y H. Stochastic choice of basis functions in adaptive function approximation and the functional-link net[J]. IEEE Transactions on Neural Networks, 1995, 6(6): 1320-1329.