基于 SVM 算法的 φ-OTDR 分布式光纤扰动传感系统模式识别研究

张俊楠[1],娄淑琴[1*],梁 生[2]

(1. 北京交通大学 电子信息工程学院,北京 100044; 2. 北京交通大学 理学院,北京 100044)

摘 要:针对相位敏感光时域反射计(φ-OTDR)分布式光纤扰动传感系统对扰动事件进行有效判别和识别的问题,提出一种基于支持向量机(SVM)的扰动判别和扰动模式识别的方法。通过提取信号时域和频域的平均值、方差、均方差以及信号功率特征,利用二叉树结构建立基于 SVM 算法的分类器,对扰动进行判别并对扰动模式进行识别。根据传感信号的特征,通过分类器 I 在对有无扰动信号进行判别的基础上,进一步对有扰动信号利用分类器对扰动事件的模式进行识别。通过实验对所提出的方法进行验证,对 600 组实验数据进行扰动判别和模式识别,正确的扰动判别率在 96%以上,漏报率和误报率在 4%以下;正确的模式识别率均在 94%以上。

关键词:分布式光纤扰动传感系统; φ-OTDR; SVM 算法; 扰动判别; 模式识别

中图分类号:TP219 **文献标志码**:A **DOI**:10.3788/IRLA201746.0422003

Study of pattern recognition based on SVM algorithm for φ-OTDR distributed optical fiber disturbance sensing system

Zhang Junnan[1], Lou Shuqin[1*], Liang Sheng[2]

(1. School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China;
2. School of Science, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Currently, phase sensitive optical time-domain reflectometer (φ-OTDR) distributed optical fiber sensing system is difficult to accurately determine current position of disturbance and distinguish the model of disturbance effectively. A method was proposed based on support vector machine (SVM) which can accurately distinguish disturbance and the model of disturbance. With the technique of the binary tree, a categorizer based on SVM was set up by extracting the various signal characteristics of the mean, the variance, the mean square deviation and energy of the time- domain and frequency-domain. Thus the disturbance and disturbance mode can be distinguished. In terms of the sensing signal feature, the categorizer I was determined if the sensing signals was disturbance signals or not firstly. Then, mode of disturbance can be recognized by the following categorizers. Experiments were carried out to validate the proposed method by 600 groups of data. The correct discrimination rate is better than 96%. The rate of missing report and the rate of false positives is less than 4%. The rate

收稿日期:2016-08-05; **修订日期**:2016-09-03

基金项目:国家自然科学基金(61475016,61575016)

作者简介:张俊楠(1990-),男,硕士生,主要从事光纤传感方面的研究。Email:14120040@bjtu.edu.cn

通讯作者:娄淑琴(1965-),女,教授,博士,主要从事光纤传感、特种光纤方面的研究。Email:shqlou@bjtu.edu.cn

of correct pattern recognition is greater than 94%.

Key words: distributed fiber disturbance sensing system; φ-OTDR; SVM algorithm; disturbance discrimination; pattern recognition

0 引 言

分布式光纤扰动传感系统 (DOFS) 以其灵敏度高、结构简单、无需外场供电、检测距离长等诸多优势在周界安防、油气管道、通信光缆监测等诸多领域中获得了广泛的应用[1-8]。其中,基于相位敏感光时域反射计(φ-OTDR)的分布式光纤扰动传感系统凭借其灵敏度高、抗电磁干扰、长距离监测、施工简单、仅用一根光纤以及扰动判别精确稳定和多点同时判别等诸多技术优势已成为目前分布式光纤扰动传感研究领域新的热点。

φ-OTDR 是由 Tarlor 等人[9-10]在 1993 年首次提出的,它与传统光时域反射计(OTDR)相比,主要区别在于其注入光纤中的光是强相干的, 因此系统输出信号为强干涉性的后向瑞利散射光。目前关于提高 φ-OTDR 系统扰动判别等方面的研究报道较多,但针对 φ-OTDR 系统扰动事件模式识别目前尚无有效的方法, 并且对扰动事件模式分析和识别的研究也相对较少。在扰动判别方面,通常采用阈值判断方式进行扰动判别。阈值算法主要有基于静态阈值的直接平均算法和基于动态阈值的移动平均算法两种。直接平均算法的阈值判定是由传感光纤长度及累加瑞利散射光强曲线周期数所决定的。在实际扰动环境采集阈值时, 信噪比较低, 易导致误报或漏报。移动平均算法,需要设定移动平均次数。平均次数设置小,信噪比低;设置过大,在长距离传输中发生多点同时扰动时, 其运算量大, 实时性和适应性差。在模式分析和识别方面,Tan 等人在 2014 年,以信号强度为识别依据,在对人工挖掘、车辆经过以及挖掘机挖掘三种模式进行了识别研究。但仅以信号强度这单一特征参量为识别依据,较易受到干扰,在复杂环境中应用的识别可靠性较低。2015 年, 本研究组提出了一种基于多特征识别的算法, 虽然大大提高了模式识别率,但该方法与扰动判别相对独立,不易于系统的集成而且计算量大。

文中通过对 φ-OTDR 系统的传感信号进行深入分析。提取传感信号时域和频域的多个特征量作为识别的信号特征, 并运用回归分析的方法建立基于二叉树结构的 SVM 分类器。通过该分类器首先区分扰动和非扰动的传感信号; 然后再对有扰动的传感信号进行进一步的分类,以识别应力破坏(敲击)、浇水(模拟下雨环境)和轻度碾压(模拟非破坏性人为扰动)等扰动事件。该方法突破了单一特征参量目标识别易受到干扰及识别事件类型有限的瓶颈问题, 提高了多种扰动事件识别的准确性,降低了运算的时间复杂度。

1 实验系统及原理

φ-OTDR 实验系统示意图如图 1 所示。超窄带线宽激光器作为光源发出的激光经声光调制器调制成光脉冲,光脉冲通过环形器注入传感光纤。在传感光纤中后向瑞利散射光在脉冲宽度内发生干涉。在干涉光通过环形器被探测器检测后, 经放大后进入数据采集卡进行采集。

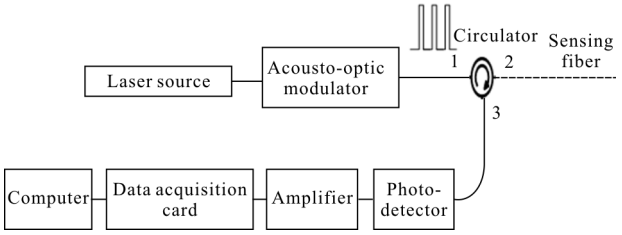


图 1 φ-OTDR 实验系统示意图

Fig.1 Schematic diagram of φ-OTDR experimental setup

当有扰动作用在传感光纤上时,由于弹光效应,受到扰动位置的传输光相位产生变化, 引起相应后向散射光的相位变化。对不同时刻后向瑞利散射光干涉输出光强曲线做差,获得差值曲线。差值曲线上的剧烈变化处,对应的即是扰动事件发生的位置。

对于 φ-OTDR 系统传感信号的分析,为简化分析起见,可只考虑单模光纤中的一重后向瑞利散射,将光纤中的后向瑞利散射看成一维脉冲响应模型[11]。假设脉宽为 W 的相干光波, 在 t=0 时刻注入到传感光纤中,忽略偏振态的影响,则后向瑞利散射的波函数如公式(1)所示。

$$y(t)=\sum_{k=1}^{N}E_s(t-t_k)\alpha(t_k)\exp\left(\alpha\frac{ct_k}{n}\right)\mathrm{rect}\left(\alpha\frac{t-t_k}{W}\right) \quad (1)$$

式中：t 为时间；$E_s(t)$为激光器光信号的波函数；W 为经过声光调制器调制后脉冲宽度；a 为光纤衰减常数；c 为真空中的光速；n 为光纤的折射率；N 为散射体总数；t_k 和 $\alpha(t_k)$分别为第 k 个散射体的时间延迟和幅值因子，第 k 个散射体距离传感光纤输入端的距离为 $ct_k/2n$。rect(*)为矩形函数如公式(2)所示。

$$\mathrm{rect}(\frac{t}{W})\begin{cases}1,\ 0\leqslant\frac{t}{W}\leqslant 1\\ 0,\ \text{otherwise}\end{cases} \quad (2)$$

当有扰动作用在传感光纤上时，得到的后向瑞利散射的波函数如公式(3)所示。

$$y'(t)=\sum_{k=1}^{N}E_s(t-\tau_k)\alpha(\tau_k)\exp\left(-\alpha\frac{c\tau_k}{n}\right)\mathrm{rect}\left(\frac{t-\tau_k}{W}\right)+\sum_{k=M+1}^{N}E_s(t-\tau_k)\alpha(\tau_k)\exp\left(-\alpha\frac{c\tau_k}{n}\right)\mathrm{rect}\left(\frac{t-\tau_k}{W}\right)+\exp\left\{j\left[f\left(t-\frac{2nz_k}{c}+\frac{nz_k}{c}\right)+f\left(t-\frac{nz_k}{c}\right)\right]\right\} \quad (3)$$

式中：第一项为扰动位置前的后向瑞利散射波的波函数；第二项为扰动位置后的后向瑞利散射波的波函数；M 为扰动位置之前散射体的个数；z_k 为第 k 个散射体距离传感光纤输入端的距离；$f(t)$为因扰动而引入的相位变化函数。由于 $f(t)$的影响，同一位置在不同的时间点的后向散射光强会有较大的变化，需选取扰动位置后向散射值进行差分，对差分曲线进行分析。

因此，由上式可得出信号的不同特征，将这些特征作为判断扰动和模式识别的依据，从而构建基于二叉树结构的 SVM 分类器。

2 基于 SVM 的分类算法

SVM 是在小样本学习的基础上发展起来的统计学习方法，专门用于小样本数据，而且对数据维数不敏感[12]。因此，SVM 方法可以较好地适用于模式识别和分类。

2.1 多分类模型

对于该传感系统不仅要进行扰动的判别，而且要对多种扰动模式进行分类，因此构建的 SVM 分类器要具有多分类的特征。

对于多分类问题，采用二叉树结构进行分类具有反应速度快，易分离，易扩张等优点，有利于将不同扰动事件引起的扰动进行快速区分的特点，而且结构简单，易于实现。文中就是采用多分类问题二叉树结构进行扰动判别和模式分类，将 SVM 与二叉树结构相结合，将多分类问题逐层次地转化为二分类的问题，从而达到扰动判别和模式识别的目的。

在现实中，由于无扰动情况要远远多于扰动情况，文中将满二叉树结构的 SVM 分类器进行优化，将满二叉树结构转化为左偏二叉树结构。优化后二叉树结构的 SVM 分类器如图 2 所示。

对于这种结构，在第一次分类的时候就能对有无扰动进行区分，从而实现有无扰动的判别；而在有扰动的情况下，又分别通过两层分类器进行扰动事件的识别，即通过两次分类确定扰动事件类别。这样既保证了准确的扰动判别，又能准确地区分不同的扰动事件类别。

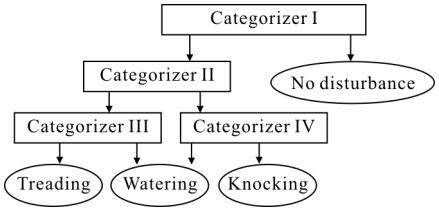


图 2 分类器结构

Fig.2 Categorizer structure

2.2 多特征的提取

对于 φ-OTDR 系统，文中选取时、频域归一化幅度平均值、方差和均方差，归一化幅度能量谱以及时、频域的波形因子作为特征参量。

平均值能够反应组内由于扰动而引起的幅度在时、频域的变化，能够快速地发现由不同扰动事件而造成的散射光强的变化。平均值函数如公式(4)所示。

$$E=\frac{1}{n}\sum_{i=1}^{n}x_i \quad (4)$$

平方差反映组内个体间的离散程度，且运算简单，相对于信号的频率信息可更突出信号的波动情况，反映扰动信号的起伏大小。平方差函数如公式(5)所示。

$$S=\frac{1}{n}\sum_{i=1}^{n}(x_i-EX)^2 \quad (5)$$

式中：EX 表示第 n 组样本的平均值。

均方差是方差的算数平方根，其可以表示一组数据离散程度的最好的指标，当有微弱扰动时，其变化明显，便于观察与计算。

波形因子是均方差与平均值的比值，是信号中的无因次量，能够有效地反应信号短期的变化。

功率谱密度反应了对于不同扰动源(踩压、浇水以及敲击)对相同信号能量的影响。离散信号的功率谱密度函数如公式(6)所示。

$$P=\frac{1}{n}\sum_{i=1}^{N}x_i^2 \tag{6}$$

平均值、方差、均方差和功率谱都能反映信号的特征，对于信号的微弱变化能够有迅速的响应。对于该扰动传感系统，能够及时地判断有无扰动，有效地提高了报警率，降低了漏报率和错报率，而且作为分类特征，能够准确地区分不同扰动事件。

3 实验结果与讨论

3.1 算法建立

实验系统如图 1 所示。传感光纤采用标准单模光纤。线宽为 10 kHz 的超窄带线宽半导体激光器作为光源，其发出 1 550 nm 波长的激光，经过声光调制器可调制为脉冲宽度约为 486 ns 的脉冲信号 (空间分辨率 50 m)，然后再进行放大，传感光纤的总长为 25.05 km。由探测器来接收传感光纤传回的后向散射信号。在探测器接收到传感信号后，经过放大再进入信号处理模块进行处理。信号处理模块包括信号采集和信号处理两部分，信号采集部分接收探测器检测的信号(采集卡采样频率 4 MHz)。信号处理部分使用 LABVIEW 软件进行扰动的判别并将判别信息显示在上位机上，并将得到的扰动信号变化波形进行分析，最终得到扰动信号的模式。

文中分别选取 900、6 500 和 21 500 m 作为实验扰动位置，针对无扰动情况，踩压扰动方式(模拟硬力破坏)，浇水扰动方式(模拟雨天情况)和敲击扰动方式(模拟非硬力破坏)这 4 种情形进行实验。对各种情形分别提取 20 组数据，每组包含 10 个周期的脉冲序列，共计 80 组、800个周期的数据进行分析。

为降低外界对信号的干扰，首先对每组数据进行归一化处理。图 3 为 900 m 处 4 种情况在 10 个周期(1 组数据)内的归一化曲线。

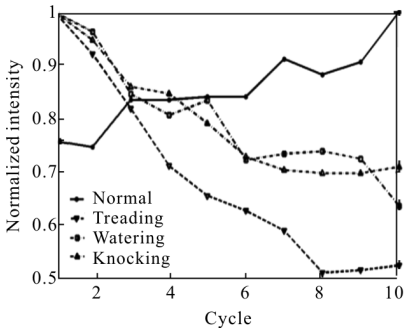
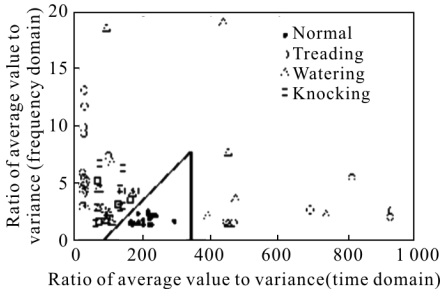


图 3 900 m 处 10 个周期的归一化幅度

Fig.3 Normalized amplitude of ten cycles at the position of 900 m

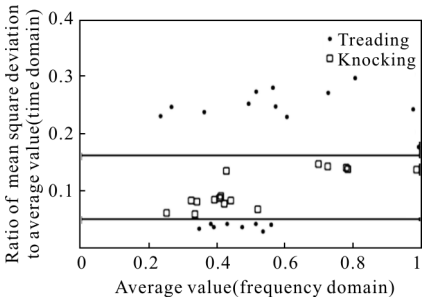
对于无扰动、踩压、浇水和敲击等 4 种情况的 20 组数据，分别做离散傅里叶变换(DFT)，将对应的时间域离散数据转化为频率域的离散数据，在频域范围内提取不同的数字特征，从而更加准确地表示信号的特征。对于时间域和频域的每组数据分别进行平均值、方差、平均值/方差、平均值/均方差、方差/平均值和均方差/平均值的计算，并提取每组信号的能量。将上述 13 个特征分别进行两两组合，并从中提取出最优的分界线以及相应的数字特征。

根据图 2 所示，分类器要提取信号相关特征将信号进行分类，图 4 为 4 个分类器的分类示意图。首先，传感信号通过分类器 Ⅰ，并提取传感信号时间域的平均值/方差和频域的平均值/方差，然后以时间域特征为横坐标，频域特征为纵坐标来确定坐标系，并判断该信号所对应的坐标点是否在由 x轴、垂直于 x 轴的直线和一条斜率 0.029 980 911、截距为−2.432 622 483 的直线所围成三角形区域内。若在该区域内，则说明该传感信号为非扰动信号；若不在，说明该传感信号为扰动信号并对扰动进行报警，再通过下面两层的分类器进行扰动事件源的模



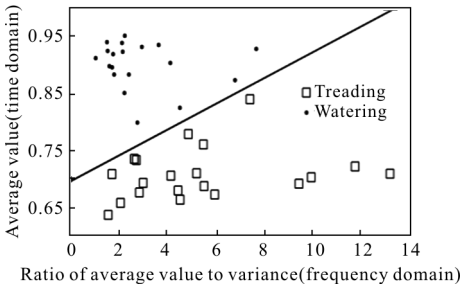
(a) 有扰动与无扰动分界图

(a) Demarcation map of disturbance and un-disturbance



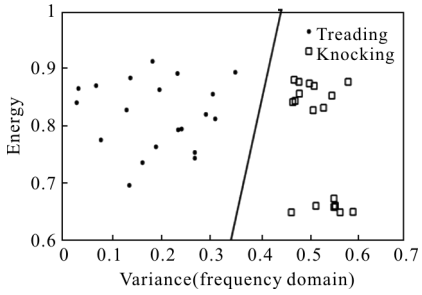
(b) 踩压与敲击分界图

(b) Demarcation map of the treading and the knocking



(c) 踩压与浇水分界图

(c) Demarcation map of the treading and the watering



(d) 浇水与敲击分界图

(d) Demarcation map of the watering and the knocking

图 4 4 个分类器示意图

Fig.4 Schematic diagram of four categorizers

式识别。因此,通过分类器Ⅰ的作用是将传感信号分为扰动信号和非扰动信号两类,然后再将该扰动信号交给分类器Ⅱ进行模式识别的处理，并对扰动信号模式进行报警。

文中针对踩压、浇水和敲击三种事件进行模式识别，并且要对不同的两种模式分别进行区分。因此，先对哪两种模式进行识别是不重要的。在文中中,首先进行踩压和敲击的模式识别。因此,分类器Ⅱ的作用是将踩压和敲击两种扰动事件进行区分。在分类器Ⅱ接收到由分类器Ⅰ交给的扰动信号后,提取信号的时间域波形因子的数值，并判断该信号所对应的数值是否在 0.048 028 与 0.158 629 之间。若波形因子在这区间之内，则表示该传感信号为非踩压式扰动信号，并将该传感信号交给分类器Ⅳ进行踩压和浇水的模式识别;若不在这区间,则该信号为非敲击式扰动信号，并将该传感信号交给分类器Ⅲ进行浇水和敲击的模式识别。

当分类器Ⅲ接收到由分类器Ⅱ交给的非敲击式传感信号后，首先要提取传感信号的频域平均值/方差和时间域平均值,然后以频域特征为横坐标,时间域特征为纵坐标建立坐标系,并判断该传感信号所对应的坐标点与公式 (7) 所示直线的位置关系。若传感信号所对应的坐标点在该直线的下方,则说明该信号为浇水扰动信号;若坐标点在直线的上方则为踩压扰动信号。因此,分类器Ⅲ的作用是对踩压扰动信号和浇水扰动信号进行最后的识别。

$$y=0.022\,837\,091\,6x+0.696\,184\,378\,0 \tag{7}$$

当分类器Ⅳ接收到分类器Ⅱ交给的非踩压式扰动信号,首先提取功率谱密度和频域的方差,然后以频域特征为横坐标，功率谱密度为纵坐标来建立坐标系，并判断该传感信号所对应的坐标点与公式(8)所示直线的位置关系。若传感信号对应的坐标点在该直线的左面则说明该信号为浇水扰动信号；若在直线的右面则为敲击扰动信号。因此，分类器Ⅳ的作用是对敲击扰动信号和浇水扰动信号进行最后的识别。

$$y=3.868\,524\,862\,6x-0.718\,101\,093\,7 \tag{8}$$

3.2 实验结果验证

对于以上结果在实际现场进行了验证与结果分析。进行实验 1 验证有无扰动情况。分别在 500 m, 1 400 m 和 10 000 m 处各进行了 100 组、共计 300 组的实验,来检验 SVM 算法对扰动判别的正确性。实验结果如表 1 所示。

表1 无扰动实验数据

Tab.1 Undisturbed experimental data

Distance/m	Sample number	Correct number	Correct rate	False number	False rate
500	100	99	99%	1	1%
1400	100	97	97%	3	3%
10000	100	96	96%	4	4%

在 500 m 处进行实验 2。分别对踩压、浇水和敲击这 3 种扰动模式各进行 100 组实验,来检验 SVM 算法对这 3 种模式识别的正确性。实验结果如表 2 所示。

表2 SVM 算法模式识别实验数据

Tab.2 Experimental data of pattern recognition based on SVM algorithm

Model	Sample number	Correct number	Correct rate	False number	False rate
Treading	100	94	94%	6	6%
Watering	100	96	96%	4	4%
Knocking	100	95	95%	5	5%

通过表 1、2可以看出,虽然随着距离的增加,误报率在上升,但总体的误报率控制在 4%以下,在所接受的范围之内, 并且该算法能够准确地区分不同扰动源的扰动,保证识别的正确率在 94%以上。

3.3 对比实验

将 SVM 算法与本研究组所提出的基于多特征识别算法进行对比。在实验 2 的条件下,利用多特征识别算法进行实验 3,实验结果如表 3 所示。

表 3 多特征识别算法模式识别实验数据

Tab.3 Experimental data of pattern recognition based on multi-feature recognition algorithm

Model	Sample number	Correct number	Correct rate	False number	False rate
Treading	100	92	92%	8	8%
Watering	100	90	90%	10	10%
Knocking	100	86	86%	14	14%

通过表 2 和表3的对比,可以明显地看出,在相同的情况下,对于踩压、浇水和敲击 3 种扰动事件,SVM 算法有着更高的识别率。对于多特征识别算法来说,其识别的准确性与提取信号特征个数的多少有关, 提取的信号特征越多,其识别越准确,但随着提取特征数量的增加,运算的时间复杂度也在随之增大,实时性和适应性变差;但对于 SVM 分类算法而言,其识别的准确性不依赖于所提取信号特征的个数,且其运算简单,时间复杂度低,实时性好。

4 结 论

针对 φ-OTDR 分布式光纤扰动传感系统预警阈值采集问题, 文中提出一种基于 SVM 算法的扰动判别和模式识别方法。通过对实验室 80 组数据的处理, 建立了基于二叉树结构的 SVM 算法分类器,并在实际现场通过对 600 组实验数据进行扰动判别和模式识别的检验,其正确扰动判别率在 96%以上, 而漏报率和误报率均在 4%以下; 正确模式识别率在 94%以上。该方法在进行准确地判别的同时, 还能够准确地分辨出不同扰动事件的模式,实现了对于不同扰动事件的识别, 提高了扰动报警的正确率。文中所提出的 SVM 算法具有较高的实时性、适应性以及模式识别的准确性,可为 φ-OTDR 分布式光纤扰动传感系统性能的进一步提高提供有益参考。

参考文献:

[1] Li X, Sun Q, Wo J. Hybrid TDM/WDM-based fiber-optic sensor network for perimeter intrusion detection [J]. ***Journal of Lightwave Technology***, 2012, 30(8): 1113-1120. (in Chinese)

[2] Wang Siyuan, Lou Shuqin, Liang Sheng, et al. Pattern recognition method for distributed disturbance sensing system based on MZ interferometer[J]. ***Infrared and Laser Engineering***, 2014, 43(8): 2613-2618. (in Chinese)

[3] Zhang Chunxi, Zhong Xiang, Li Lijing, et al. Long-distance intrusion detection system based on phase sensitive light time domain reflectometer [J]. ***Infrared and Laser Engineering***, 2015, 44(2): 742-746.

[4] Wang Peng, Lou Shuqin, Liang Sheng, et al. Threshold algorithm for selective average φ-OTDR distributed fiber perturbation sensing system[J]. ***Infrared and Laser Engineering***, 2016, 45(3): 1003-1007. (in Chinese)

[5] Lin Wentai, Liang Sheng, Lou Shuqin, et al. A new type of optical fiber distributed vibration sensing system with low false alarm rate [J]. ***Infrared and Laser Engineering***, 2016, 45(4): 1845-1848. (in Chinese)

[6] Liu Jianxia. Research on φ-OTDR distributed optical fiber sensing monitoring technology [J]. ***Progress in Laser and Optoelectronics***, 2013, 50(8): 080021. (in Chinese)

[7] Wang He, Sun Qizhen, Li Xiaolei. Research progress of interferometric distributed fiber vibration sensing technology [J]. ***Progress in Laser and Optoelectronics***, 2013, 50 (2):

020004. (in Chinese)

[8] Li Qin, Zhang Chunxi, Li Lijing. Effect of frequency drift of laser on the accuracy of phase-sensitive optical time-domain reflectometer disturbance [J]. ***Chinese Journal of Lasers***, 2014, 41(3): 0305003. (in Chinese)

[9] Zhao Enming, Li Entao, Teng Pingping, et al. Integrated nitrite microfluidic fluorescence sensor based on surface open fiber[J]. ***Optical Precision Engineering***, 2015, 23(8): 2158-2163. (in Chinese)

[10] Xu Ning, Dai Ming. Distributed optical fiber temperature and pressure sensor design [J]. ***Chinese Optics***, 2015, 8(4): 629-635. (in Chinese)

[11] Wang Jie, Jia Xinhong. Phase-sensitive time-domain reflectometry based on bidirectional Raman amplification [J]. ***Acta Physica Sinica***, 2013, 62(4): 044212.

[12] Fan Qi. Study on the toxicity of mushrooms based on support vector machine [J]. ***Chinese Agricultural Science Bulletin***, 2015, 31(19): 232-236. (in Chinese)