

引用本文: 赵怀亮, 杨润平, 赵少华, 等. 基于类内类间距离和 AdaBoost-SCN 的分类识别方法[J]. 光通信技术, 2025, 49(2): 29-33.

基于类内类间距离和 AdaBoost-SCN 的分类识别方法

赵怀亮¹, 杨润平^{1*}, 赵少华¹, 苏润梅¹, 陈林毓¹, 尚秋峰², 姚国珍²

(1. 内蒙古电力(集团)有限责任公司 内蒙古超高压供电分公司, 呼和浩特 010000;

2. 华北电力大学 电子与通信工程系, 河北 保定 071003)

摘要: 为提高输电线路覆冰监测的识别准确率和实时性, 提出了一种基于类内类间距离与自适应增强随机配置网络(AdaBoost-SCN)的分类识别方法。首先, 对相位敏感光时域反射仪相位信号进行类内距离和类间距离的联合评估, 通过评分策略实现全特征向量的降维筛选, 从而提取关键敏感特征; 随后, 采用 AdaBoost-SCN 算法进行覆冰等级分类, 该算法以随机配置网络作为基分类器, 通过迭代优化构建强分类模型。实验结果表明, 该方法在测试集上的平均识别准确率达到 94.7%, 相比传统方法提高了 2%~5%, 计算用时从 0.54 s 降低至 0.32 s。

关键词: 相位敏感光时域反射仪; 光纤传感; 特征选择

中图分类号: TN256 文献标志码: A 文章编号: 1002-5561(2025)02-0029-05

DOI: 10.13921/j.cnki.issn1002-5561.2025.02.006

开放科学(资源服务)标识码(OSID):



Classification and recognition method based on intra-class and inter-class distances and AdaBoost-SCN

ZHAO Huailiang¹, YANG Runping^{1*}, ZHAO Shaohua¹,

SU Runmei¹, CHEN Linyu¹, SHANG Qiufeng², YAO Guozhen²

(1. Inner Mongolia Electric Power (Group) Co., Ltd., Inner Mongolia Ultra High Voltage Power Supply Branch, Hohhot 010000, China;

2. Department of Electronics and Communication Engineering, North China Electric Power University, Baoding Hebei 071003, China)

Abstract: To improve the recognition accuracy and real-time performance of transmission line icing monitoring, this paper proposes a classification and recognition method based on intra-class and inter-class distances and adaptive boosting random configuration network (AdaBoost-SCN). First, the phase signals from phase-sensitive optical time-domain reflectometry are jointly evaluated using intra-class and inter-class distance metrics. A scoring strategy is then applied to reduce the dimensionality of the full feature vector, thereby extracting key sensitive features. Subsequently, the AdaBoost-SCN algorithm is employed for icing severity classification and recognition, where the random configuration network serves as the base classifier, and a strong classification model is constructed through iterative optimization. The experimental results demonstrate that the proposed method achieves an average recognition accuracy of 94.7% on the test set, outperforming traditional methods by 2%~5%, while reducing computation time from 0.54 s to 0.32 s.

Key words: phase-sensitive optical time-domain reflectometry, optical fiber sensing, feature selection

0 引言

相位敏感光时域反射仪(Φ -OTDR)作为一种应用广泛的分布式光传感技术, 具有探测距离长、抗电磁

收稿日期: 2024-09-28。

基金项目: 内蒙古电力(集团)有限责任公司科技项目(2023-5-25)资助。

作者简介: 赵怀亮(1991—), 男, 硕士, 工程师, 毕业于内蒙古大学电子与通信工程专业, 现就职于内蒙古电力集团有限责任公司内蒙古超高压供电分公司, 主要从事电力通信相关工作, 包括网络架构设计、设备选型与配置、通信协议优化等。

* **通信作者:** 杨润平(1978—), 男, 硕士, 高级工程师, 主要研究方向为电力通信。



干扰、灵敏度高以及无需外置传感器等诸多优点^[1], 能够直接利用光纤复合架空地线(OPGW)中的冗余光纤, 开展沿线振动事件的在线监测。随着 Φ -OTDR 技术不断发展, 其在输电线路监测领域取得了诸多进展。张旭莘等人^[2]对 Φ -OTDR 在周界安防、地震波探测、输电线路监测、水下线缆监测等领域的应用进行了分析与展望。齐蒙等人^[3]针对 Φ -OTDR 在输电线路领域的应用开展了仿真实验。DING Z W 等人^[4]分析并建立了输电线路状态的数学模型, 借助 Φ -OTDR 测得了覆冰与悬垂线固有频率之间的关系, 为输电线路覆冰监测提供了全新的解决方案。同时, 以南瑞为代表的电力公

赵怀亮, 杨润平, 赵少华: 基于类内类间距离和 AdaBoost-SCN 的分类识别方法

司纷纷积极开展分布式光纤传感技术在输电线路覆冰监测方面的应用研究。

目前, 基于机器学习的信号识别是 Φ -OTDR 模式识别的主要方式^[5]。传统的机器学习方法通常分为 2 个阶段: 特征提取与分类识别。在特征提取阶段, 常采用快速傅里叶变换、经验模态分解或小波变换等方法, 从信号中提取时域、频域及时频域特征, 进而构建多维融合特征集。在分类识别阶段, 常用的算法包括支持向量机(SVM)^[6]、K-近邻(KNN)^[7]、决策树(DT)^[8]以及随机配置网络(SCN)^[9]等。传统的机器学习方法通过多域特征提取在一定程度上提高了识别准确率, 然而随着特征数目的增加, 有效特征的选择往往依赖专家经验。于是, 刘帅康等人^[10]提出鲁棒性边界鉴别局部保持投影方法进行特征降维, 并构建集成学习分类器, 将预测精度提升至 80% 以上。马建桥等人^[11]利用相关性和类间差异度进行特征降维, 在确保识别速度的同时, 使识别准确率达到 86.11%, 为特征降维的有效性提供了依据。由于 Φ -OTDR 对外界环境扰动较为敏感, 致使振动信号信息十分丰富, 使用单一模型分类器存在泛化能力差的问题, 而自适应增强(AdaBoost)算法能够在训练过程中学习多个基分类器的优点并减弱其缺点, 从而获得比单分类器更优的分类识别性能。鉴于此, 本文提出一种基于类内类间距离和 AdaBoost-SCN 的分类识别方法。

1 基于类内类间距离和 AdaBoost-SCN 的分类识别方法

本文提出的方法首先通过联合评估类内距离与类间距离, 采用评分策略对全特征向量进行降维筛选, 以保留关键敏感特征。随后, 将这些敏感特征输入分类器, 实现覆冰等级的精准识别。

1.1 基于类内距离和类间距离的特征选择方法

本方法将所提取的多个特征 ($f_1 \sim f_{30}$) 组合成新的特征向量 F , 如式(1)所示。

$$F = \{f_{c,n,k} \mid 1 \leq c \leq C, 1 \leq n \leq N_c, 1 \leq k \leq K\} \quad (1)$$

其中, $f_{c,n,k}$ 为第 c 类下第 n 个样本的第 k 个特征, C 为模式数, N_c 为 c 类模式下的样本数, K 为每个样本特征维数(在本方法中 $K=30$)。

本方法采用类内距离和类间距离对不同维度的特征进行评分, 并设置相应的阈值筛选敏感特征, 减少样本的特征维数 K 。

对于特征向量 F , 其特征降维优化过程按如下步骤进行:

步骤 1: 以 i, j 为累加计数变量, 计算相同类别样本的平均距离 $d_{c,k}$ 。

$$d_{c,k} = \frac{1}{N_c(N_c-1)} \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} |f_{c,i,k} - f_{c,j,k}| \quad (2)$$

对 $d_{c,k}$ 求平均后得到平均类内距离如式(3)所示。

$$d_k^a = \frac{1}{C} \sum_{c=1}^C \sum_{k=1}^C d_{c,k} \quad (3)$$

步骤 2: 计算 d_k^a 的方差因子 R_k^a 。

$$R_k^a = \frac{\max |d_{c,k}|}{\min |d_{c,k}|} \quad (4)$$

步骤 3: 计算不同类别样本的平均类间距离 d_k^b 。

$$d_k^b = \frac{1}{C(C-1)} \sum_{c=1}^C \sum_{e=1}^C |\mu_{e,k} - \mu_{c,k}| \quad (5)$$

其中, $\mu_{c,k}$ 和 μ_{ek} 分别为第 c 或 e 类中所有第 k 个特征均值, $\mu_{c,k} = \frac{1}{N_c} \sum_{n=1}^{N_c} f_{c,n,k}$, $\mu_{ek} = \frac{1}{N_e} \sum_{n=1}^{N_e} f_{e,n,k}$, N_e 为 e 类模式下的样本数, 且 $c, e \in \{1, 2, \dots, C\}$, $c \neq e$ 。

步骤 4: 定义并计算 d_k^b 的方差因子 R_k^b 。

$$R_k^b = \frac{\max |\mu_{e,k} - \mu_{c,k}|}{\min |\mu_{e,k} - \mu_{c,k}|} \quad (6)$$

步骤 5: 定义并计算补偿因子 δ_k 。

$$\delta_k = \frac{1}{\frac{R_k^a}{\max |R_k^a|} + \frac{R_k^b}{\max |R_k^b|}} \quad (7)$$

步骤 6: 计算平均类间距离 d_k^b 和平均类内距离 d_k^a 的比值 ε_k 并乘以补偿因子 δ_k 。

$$\varepsilon_k = \delta_k \frac{d_k^b}{d_k^a} \quad (8)$$

将 ε_k 归一化得到评价分值 $\Delta \varepsilon_k$ 。

$$\Delta \varepsilon_k = \frac{\varepsilon_k}{\max [\varepsilon_k]} \quad (9)$$

步骤 7: 根据步骤 1~6 可计算各个特征所对应的分数, 分数越高意味着相应的特征越容易区分信号类型。设定阈值为 T (本方法中取 $T=0.6$), 选择 $\Delta \varepsilon_k \geq 0.6$ 的特征作为敏感特征, 得到新的特征向量 $F' = \{f_{c,n,l} \mid 1 \leq c \leq C, 1 \leq n \leq N_c, 1 \leq l \leq L\}$, 其中 $L < K$ 。基于类内类间距离的特征降维筛选流程如图 1 所示。

令 c 类下第 i 个样本降维后的特征子集 $A_i = f_{c,n,l} = (a_i^{(1)}, a_i^{(2)}, \dots, a_i^{(L)})$, 其对应的类别标签为 B_i , 其中 $i = 1, 2, \dots, N_c$, 得到特征数据集 $\{(A_1, B_1), (A_2, B_2), \dots, (A_{N_c}, B_{N_c})\}$ 。

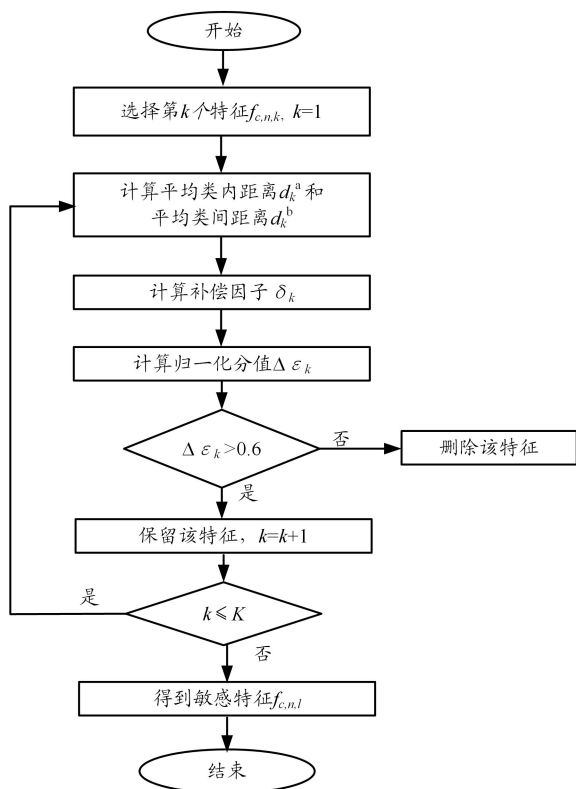


图1 基于类内类间距离的特征降维筛选流程图

$B_N\}$ }, 完成数据集的制作。将数据集划分为训练集 $\{(A_1, B_1), (A_2, B_2), \dots, (A_{N_1}, B_{N_1})\}$ 和测试集 $\{(A_1, B_1), (A_2, B_2), \dots, (A_{N_2}, B_{N_2})\}$, 其中 N_1, N_2 分别为训练集和测试集样本数, 且 $N_1 + N_2 = N$ 。

1.2 AdaBoost-SCN

AdaBoost-SCN 结合了 SCN 和 AdaBoost 的优势。SCN 是一种增量式神经网络, 通过随机配置构建模型, 利用不等式约束更新输入权重, 并采用最小二乘法计算输出权重, 具有快速学习和良好逼近能力。其结构包含输入层、隐藏层(节点数为 M)和输出层, 类似于传统单隐藏层前馈神经网络。AdaBoost 是典型的集成学习算法, 通过调整样本权重训练多个基分类器, 将它们线性组合成强分类模型。在每轮迭代中, 错误分类样本的权重增大, 正确分类样本的权重减小, 从而动态优化分类性能。AdaBoost-SCN 以 SCN 为基分类器(记为 h_p), 借助 AdaBoost 的迭代机制提升整体分类效果, 实现高效的数据分类与预测。分类具体步骤如下:

步骤 1: 对于训练集 $\{(A_1, B_1), (A_2, B_2), \dots, (A_{N_1}, B_{N_1})\}$, 初始化训练集权重 D_1 。

$$D_1 = (1/N_1, 1/N_1, \dots, 1/N_1) \quad (10)$$

步骤 2: 根据设定值, 迭代 p 次, $p=1, 2, \dots, P, P$ 为

总迭代次数, 根据具有权重分布 D_p 的训练集训练 h_p 。

步骤 3: 计算 h_p 的误差率 e_p 。

$$e_p = w_{pi} I[h_p(A_i) \neq B_i] \quad (11)$$

其中, w_{pi} 是迭代 p 次时第 i 个样本所对应的权值, I 是 h_p 对 A_i 的分类结果, 若 $h_p(A_i) \neq B_i$, 则 $I[h_p(A_i) \neq B_i] = 1$, 否则 $I[h_p(A_i) \neq B_i] = 0$ 。

步骤 4: 为防止出现分类错误的样本权重过大的情况, 将当前基分类器的正确情况考虑在内, 计算基分类器 h_p 的系数 η_p 。

$$\eta_p = \frac{1}{2} \ln \frac{1-e_p}{e_p} + F(W_p) \quad (12)$$

其中, $F(x)$ 、 W_p 分别如式(13)、式(14)所示。

$$F(x) = -\ln \left[\frac{1}{0.5(x+1)} - 1 \right] \quad (13)$$

$$W_p = \sum_{i \in I(h_p(A_i) = B_i)} w_{pi} \quad (14)$$

其中, $F(x)$ 在 $(0, 1)$ 内单调递增; W_p 反映分类正确的样本权重分布状态, 其值域为 $(0, 1)$, 值越大表示正确样本权重越大, 即该分类器效果越好。

步骤 5: 根据误差率和基分类器系数, 更新训练集的权重分布 D_{p+1} , 提高分类错误样本的权重。

$$D_{p+1} = \frac{D_p \exp^{[-\eta_p B_i h_p(A_i)]}}{\sum_{i=1}^C D_p \exp^{[-\eta_p B_i h_p(A_i)]}} \quad (15)$$

步骤 6: 重复步骤 2~5, 直至达到最大的迭代次数 P 次, 训练得到 P 个具有相应系数的基分类器, 将其组合可得到强分类器, 其函数表达式为

$$G(A_i) = \text{sign} \left[\sum_{p=1}^P \eta_p h_p(A_i) \right] \quad (16)$$

强分类器建模过程如图 2 所示。

经过上述步骤 1~6 可以将多个 SCN 分类器组合成一个 AdaBoost-SCN 分类器。在 AdaBoost-SCN 分类器模型中, 每个基分类器会在训练过程中充分发挥其优势, 克服其劣势, 因此该模型相较于单个基分类器而言具有更好的识别效果, 泛化能力更强。

本文采用 AdaBoost-SCN 分类器对输电线路覆冰进行识别, 流程如图 3 所示。

2 现场数据验证

2.1 数据获取

本文用于输电线路覆冰监测的 Φ -OTDR 系统结构如图 4 所示。系统采用输出功率为 13 dBm、中心波

赵怀亮, 杨润平, 赵少华: 基于类内类间距离和 AdaBoost-SCN 的分类识别方法

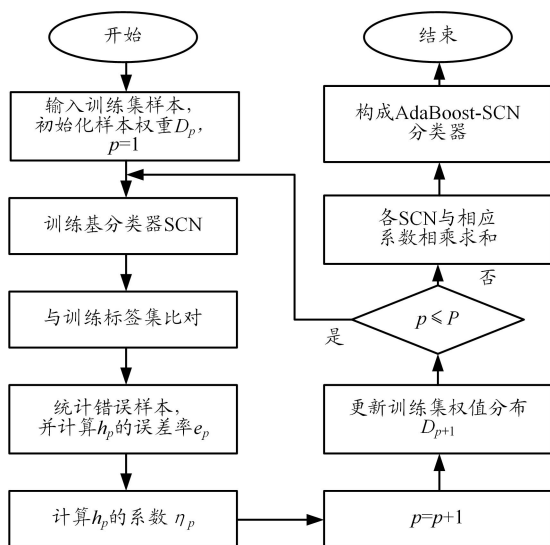


图2 强分类器建模过程

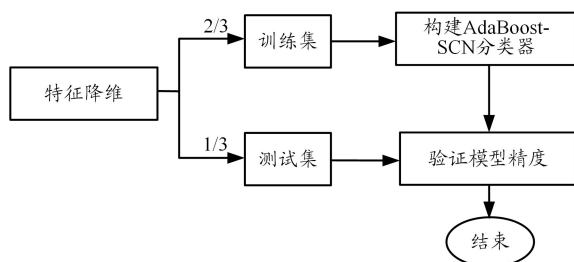
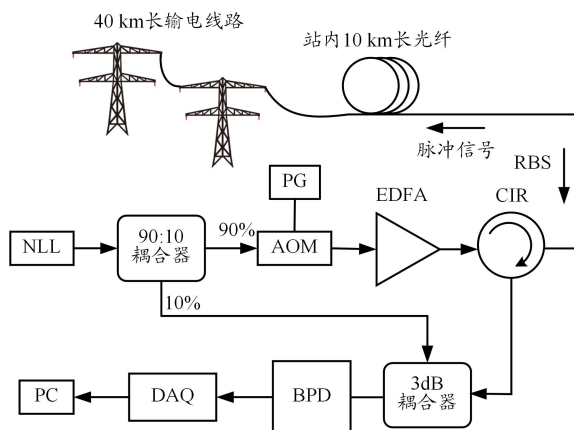


图3 输电线路覆冰识别流程

图4 Φ -OTDR系统结构

长为 1 550 nm 的窄线宽激光器(NLL)。脉冲采样频率为 1.25 kHz, NLL 发出的光经过 90:10 耦合器后分为 2 路: 一路(90%)作为探测光, 另一路(10%)作为本地参考光。探测光经由声光调制器(AOM)调制后产生 50 MHz 频移的脉冲光输出, 其脉宽为 100 ns。脉冲发生器(PG)直接驱动 AOM 工作, 调制后的光信号经掺铒光纤放大器(EDFA)进行功率放大后, 通过光纤环形器(CIR)耦合注入至 OPGW 内的传感光纤中。光纤中产生的后向瑞利散射(RBS)光经 CIR 后在 3 dB 耦合器中与本地

光进行拍频, 然后进入平衡光电探测器(BPD)转换为电信号。转换后的信号经过采样率为 250 MSa/s 的数据采集卡(DAQ)转换为数字信号。最后, 计算机(PC)对接收到的数字信号进行滤波、解调和分析, 从而获取 RBS 光的相位信息。

Φ -OTDR 系统安装在某超高压变电站内, 通过 OPGW 内的光纤实现了对 50 km 输电线路的分布式监测, 其中包括 10 km 站内光缆和 40 km 站外输电线。本文采集了 2023 年 12 月 13 日至 2023 年 12 月 31 日的的数据, 并结合现场观测结果对覆冰情况进行分类: 12 月 13 日至 12 月 14 日覆冰厚度为 0~5 mm, 定义为 1 级覆冰; 12 月 15 日至 12 月 21 日覆冰厚度为 5~10 mm, 定义为 2 级覆冰; 而 12 月 26 日至 12 月 29 日线路无覆冰。

本文针对上述 3 种覆冰工况, 提取了不同距离点的数据特征, 经过初步特征分析, 筛选出 30 维能够有效反映覆冰工况的特征。每种工况分别提取了 300 组样本数据, 总计获得 900 组样本数据。这些数据按照 2:1 的比例划分为训练集和测试集, 用于后续分析与建模。

此外, 本文对信号标签进行了定义: 无覆冰信号标记为 1, 1 级覆冰信号标记为 2, 2 级覆冰信号标记为 3, 以便于区分和处理不同工况下的数据特征。

2.2 方法的有效性评估

本文为验证方法的有效性, 计算了提取的 30 维特征向量的类内和类间间距, 并依据评分策略进行筛选。识别准确率、特征维数与阈值之间的关系曲线如图 5 所示。可以看出, 识别准确率随着阈值的增大而提升, 并在阈值为 0.6 时达到拐点, 此时识别准确率达到 94.7%。经特征优化筛选后, 特征维度由初始 30 维有效压缩至 13 维, 在维持最佳识别性能的同时显著提升了系统实时性, 充分验证了方法的有效性。因此, 本文选取阈值 $T=0.6$ 。当阈值 $T=0.6$ 时, 最终得到的 13 维敏感特征包括: 时域信号中的能量和样本熵(2 种特征); 频域信号中的最大值、峰峰值、能量以及频谱幅值样本方差(4 种特征); 时频域中的 IMF_2 、 IMF_3 、 IMF_4 、 IMF_5 、 IMF_6 、 IMF_7 和 IMF_8 小波特征尺度熵(7 种特征)。

本文将未降维与降维后的特征集分别输入到 AdaBoost-SCN 分类器中进行性能评估, 结果如表 1 所示。通过对比准确率与计算时间, 可以看出本文提出的方法不仅能够显著降低特征维数, 还有效提高了识别准确率, 从而在识别率和实时性之间实现了良好的平衡。

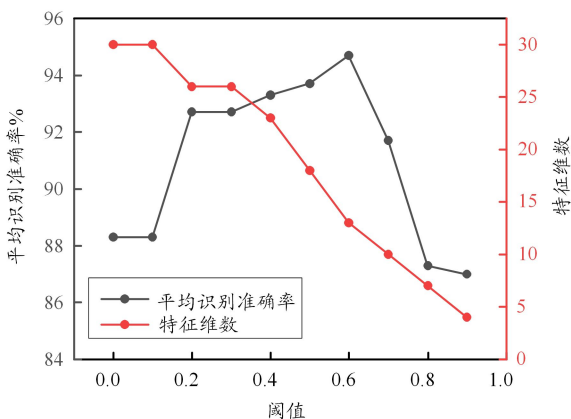


图 5 平均识别准确率、特征维数与阈值的关系

表 1 2 种特征选择指标对比

方法	特征维数/维	平均识别准确率/%	计算用时/s
未降维	30	88.3	0.54
降维后	13	94.7	0.32

2.3 分类器评估

在分类器参数优化方面,通过遍历法确定了 SCN 的最佳隐藏层节点数为 4,迭代次数为 4;对于 SVM,核函数选用径向基函数(RBF),并通过网格搜索法进行参数调优,最终确定最佳惩罚因子为 10,核函数参数为 0.01。

本文将经过特征选择后提取的敏感特征作为分类器输入,用于测试所提出方法相较于 SCN 和 AdaBoost-SVM 的优越性。利用 300 组测试样本对分类器进行评估,得到各分类器的评价指标如表 2 所

表 2 各分类器评价指标

分类器	类别	识别准确率/%	平均识别准确率/%
SCN	无覆冰	86	92.7
	1 级覆冰	93	
	2 级覆冰	99	
AdaBoost-SVM	无覆冰	86	89.7
	1 级覆冰	91	
	2 级覆冰	92	
AdaBoost-SCN	无覆冰	89	94.7
	1 级覆冰	98	
	2 级覆冰	97	

示。可以看出,AdaBoost-SCN 分类器在各类别覆冰工况识别准确率上均有所提升,其平均识别准确率达到 94.7%,相比其它分类器提高了 2%~5%。这表明本文提出的方法在分类性能上具有显著优势,能够更有效地完成覆冰工况的识别任务。

3 结束语

本文针对不同覆冰工况下的 Φ -OTDR 相位信号,提出了类内和类间的评分策略,实现了有效的特征降维,成功保留了对覆冰识别敏感的关键特征。基于迭代评估原则,本文采用 AdaBoost-SCN 分类器对覆冰等级进行识别,取得了显著的效果:平均识别准确率达到 94.7%,计算用时由 0.54 s 降低至 0.32 s,不仅提高了识别准确率,还大幅提升了识别速度。这一结果表明,本文提出的方法在特征提取与分类识别方面具有较高的实用价值。

参考文献:

- [1] 黄麟景,周霄,樊昕昱,等.多机理融合分布式光纤传感技术研究[J].光学学报,2024,44(1):161-172.
- [2] 张旭革,丁哲文,洪瑞,等.相位敏感光时域反射分布式光纤传感技术[J].光学学报,2021,41(1):100-114.
- [3] 齐蒙,党华,张珂,等.基于 Φ -OTDR 的光纤分布式振动传感器的研究及改进[J].光通信技术,2018,42(7):38-41.
- [4] DING Z W, ZHANG X P, ZOU N M, et al. Phi-OTDR based on-line monitoring of overhead power transmission line [J]. Journal of Lightwave Technology, 2021, 39(15): 5163-5172.
- [5] 赵丽娟,魏迎健,徐志钮.基于 Φ -OTDR 的振动事件识别分类器研究进展[J].光通信技术,2023,47(2):1-5.
- [6] LIU T, LI H, HE T, et al. Ultra-high resolution strain sensor network assisted with an LS-SVM based hysteresis model [J]. Opto-Electronic Advances, 2021, 4(5): 5-15.
- [7] LIU S, YU F, HONG R, et al. Advances in phase-sensitive optical time-domain reflectometry [J]. Opto-Electronic Advances, 2022, 5(3): 28-55.
- [8] HUANG Y, ZHAO H, ZHAO X, et al. Pattern recognition using self-reference feature extraction for Φ -OTDR [J]. Applied Optics, 2022, 61(35): 10507-10518.
- [9] 吉建娇,王殿辉.基于区间边界探测的进化随机配置网络[J].控制理论与应用,2023,41(10):1913-1922.
- [10] 刘帅康,曹伟,管志强,等.基于 RMDLPP 的雷达空中目标分类[J].系统工程与电子技术,2024,46(4):1220-1228.
- [11] 马建桥,杨广泽,曹盘盘,等.基于相关性和类间差异度的放电声信号特征量选择与降维[J].高电压技术,2023,49(3):1194-1204.