

doi:10.3969/j.issn.1003-2029.2020.03.004

# 基于 PCA/LSTM 的海底观测网电力系统 供电海缆故障定位

耿坤, 吕枫\*

(同济大学 海洋与地球科学学院, 上海 200092)

**摘要:**海底观测网电力系统供电海缆易发生绝缘故障, 由于其在海底极端环境下运行, 实现高精度的电缆故障定位, 对于降低其维修成本至关重要。基于海底观测网电能监控系统实时采集的电气测量参数, 本文提出了基于主成分分析 (Principal Components Analysis, PCA) 和长短期记忆 (Long Short-Term Memory, LSTM) 网络的供电海缆高阻故障定位方法, 先使用 PCA 进行数据降维, 再将处理后的数据输入 LSTM 网络进行训练, 捕获多源数据的时间特征, 挖掘电力系统故障特征与电气参数之间的对应关系, 实现海缆故障精确定位。通过在海底观测网原型系统上的实验证明, 该方法经 K 折交叉验证方法 (K-fold Cross Validation, K-CV) 验证, 准确率可达 91% 左右, 故障定位误差约为 0.4 km。

**关键词:**海底观测网; 电力系统; 故障定位; 长短期记忆网络; 主成分分析

**中图分类号:** P751      **文献标志码:** A      **文章编号:** 1003-2029(2020)03-0022-08

海底观测网采用光电复合海缆连接海岸基站和海底主基站, 将电力系统和通信系统从陆地延伸到海底, 实现对大量海底仪器设备的长期供电、双向通信、远程控制和精确授时, 用于观测和记录海底物理、化学、生物和地质等自然环境变化<sup>[1-3]</sup>。

海底观测网运行在极端恶劣环境下, 在运行周期内, 可能因老化、磨损或外力导致供电海缆绝缘故障, 表现为海缆绝缘电阻异常下降。多数情况下海缆的绝缘层是因水树生长而逐渐退化的, 所以在退化初期实现海缆故障精确定位是海底观测网安全运行的保障。在现有的大规模海底观测网的故障定位方法中, 必须将岸基站的电源输出降低至低电压, 或逐步改变岸基站电源的输出电压并采集电气参数, 这都会对整个观测网造成扰动, 且定位精度仅为 1 km 左右<sup>[4-8]</sup>。在陆地电力系统中, 通常采用行波法定位电缆故障, 根据初始波与来自故障点的反

射波之间到达的时间差实现故障定位<sup>[9-11]</sup>, 但由于海底观测网中的中继器和分支器会反射行波, 并且行波通常只能传播数千米, 所以行波法并不适用于海底观测网故障定位。因此, 在海底观测网正常运行过程中, 如何利用电气测量数据进行精准的故障定位是亟需解决的问题。

自 19 世纪 70 年代以来, 针对电力系统故障诊断及定位, 基于数据驱动的方法成为有力工具。数据驱动的故障诊断及预测方法主要有遗传算法、支持向量机、奇异谱分析等。由于海底观测网的电气测量参数众多, 且获得的数据可能存在粗差, 故很难用单一模型来拟合数据曲线以达到理想的定位效果, 而混合方法建模过程复杂, 受先验知识及人为因素影响较大, 不利于实际应用中推广。

针对上述不足, 人工神经网络 (Artificial Neural Network, ANN) 因其无需人工建模, 对数据精度要求

收稿日期: 2020-04-03

基金项目: 上海市科技创新行动计划资助项目 (16DZ1205000); 国家自然科学基金资助项目 (61936014)

作者简介: 耿坤 (1996-), 男, 硕士研究生, 主要从事缆系海底观测网络故障诊断研究。E-mail: kgeng@tongji.edu.cn

通讯作者: 吕枫 (1983-), 男, 博士, 副教授, 博士生导师, 主要从事海底广域观测网、海洋自主智能系统、水下电力电子系统、海底俘能与近场输能技术研究。E-mail: LF@tongji.edu.cn

低等优势在故障诊断及定位研究中发挥了重大的作用。ANN 网络可以用于任何瞬态不稳定性电力系统的故障早期在线诊断,有利于电力系统保护以及提升控制系统的性能<sup>[12-13]</sup>。但海底观测网电气参数中的时间信息不能被 ANN 所提取,为此,循环神经网络(Recurrent Neural Network, RNN)进一步实现了对历史故障数据时间序列信息的提取,可以实现电力系统故障诊断及定位。为了解决传统 RNN 梯度消失问题,人们在此基础上又提出了一种长短期记忆(Long Short-Term Memory, LSTM)网络,其在较长时间范围内提取时间序列特征方面表现良好<sup>[14]</sup>。与普通的 ANN 网络相比, LSTM 网络更适合电力系统故障识别和定位<sup>[15]</sup>。王鑫等<sup>[16]</sup>实现了基于历史故障数据的 LSTM 网络故障时间序列预测, Zhang S 等<sup>[17]</sup>也利用支持向量机及 LSTM 网络实现了陆上电力系统故障定位。而海底观测网电气参数维数大,数据相关度高,精度低,只采用 LSTM 网络的海缆故障定位精度不足。本文提出了一种适用于海底观测网电力系统的基于 PCA/LSTM 网络的供电海缆故障定位方法,该方法充分利用海底观测网电能监控系统提供的电气测量数据,实现了海缆高阻故障的高精度在线定位,并且利用海底观测网原型系统进行了实验验证,最后分析了实验结果准确性。

## 1 海底观测网电能监控系统

国外主要由美国华盛顿大学初步研究了海底观测网电能监控方法<sup>[18]</sup>,国内主要是同济大学和浙江大学等单位研发了电能监控系统相关软硬件<sup>[19]</sup>。同济大学海洋观测与智能系统实验室实现了海底观测网电网拓扑和运行数据的存储、管理和可视化。

海底观测网远程电能监控系统数据库存储有电网参数、实时数据和历史数据。电网参数为各海缆段的长度及其单位长度海缆参数、海底电网的拓扑结构以及所有保护和报警阈值等;实时数据为各开关状态量以及电压值和电流值等模拟量,由控制器测量后通过海底光纤通信系统发送到数据采集系统;历史数据为整个数据库的主体部分<sup>[20]</sup>。

海底设备上传至海岸基地的电力系统状态数据可以实时显示在数据显示界面,同时也存储在数据库中。系统接收到数据后会对数据进行分析和检测,若发现有异常波动或超过阈值的数据,会报警,同时进行故障诊断,判断系统是否发生故障、确定

故障点的位置,并将计算结果反馈给管理员。

海底观测网电能监控系统为实现海缆故障诊断及定位提供了数据支持,图 1 为同济大学海洋观测与智能系统实验室开发的海底观测网电能监控系统软件界面,本文利用数据库中存储的电力系统运行参数,实现海缆的故障诊断。

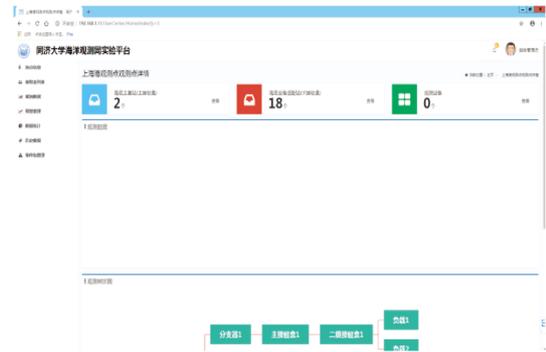


图 1 海底观测网电能监控系统软件界面

## 2 技术原理

### 2.1 LSTM 关键技术介绍

常规 ANN 仅在层与层之间具有连接,层中的单元没有连接,网络不传输时间信息,因此处理时间序列性能较差。而 RNN 则充分利用输入数据的时间信息,其中各隐藏层细胞之间的连接在同一层内形成有向循环<sup>[21]</sup>(如图 2 所示)。

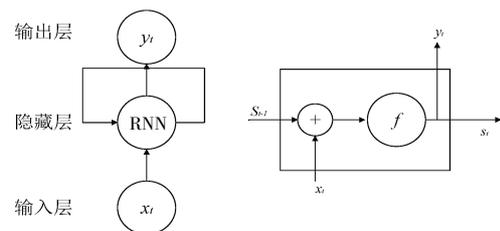


图 2 RNN 网络简图(左)及隐藏层细胞展开结构(右)

对于给定的输入端数据,表示为  $\vec{x}$ ,  $\vec{x}=(x_1, x_2, \dots, x_n)$ ,可以通过如下的前向计算公式(1)和式(2)计算出其通过一个标准 RNN 后的隐藏层序列  $\vec{s}_t=(s_1, s_2, \dots, s_n)$ 和输出层序列  $\vec{y}_t=(y_1, y_2, \dots, y_n)$ 。其中,  $W$  表示权重系数矩阵(例如  $W_{xs}$  表示输入层到隐藏层的权重系数矩阵),  $b$  表示各层的偏置向量,  $f$  表示激活函数。

$$s_t=f(W_{ss}s_{t-1}+W_{xs}x_t+b_s) \quad (1)$$

$$y_t=W_{sy}s_t+b_y \quad (2)$$

RNN 隐藏层之间的神经元是互相连接的,因此网络可记住先前的梯度信息,并将其应用于计算当

前输出。相应地,RNN 在训练时序数据时采用反向传播算法更新各系统参数。但在反向传播时,梯度也会呈指数倍数的衰减,这种衰减现象导致 RNN 无法处理数据的长期依赖,故当 RNN 网络层数较大时易产生梯度消失和爆炸问题<sup>[22]</sup>。

LSTM 网络在 RNN 网络中引入了一个长短期记忆功能块,以记住长时间或短时间情况下的梯度值<sup>[23]</sup>。具体来说,RNN 网络的隐藏单元被包含输入、输出和遗忘三个门的 LSTM 功能块所取代,这三个门用于控制数据流入或流出其细胞存储的过程。具体的 LSTM 功能块如图 3 所示。其中, $i, f, o$  分别代表输入门、遗忘门和输出门; $c$  代表细胞状态; $s$  表示隐藏层序列; $W$  表示权重系数矩阵, $b$  表示偏置项; $\tanh$  和  $\sigma$  分别表示正切激活函数和 sigmoid 激活函数。

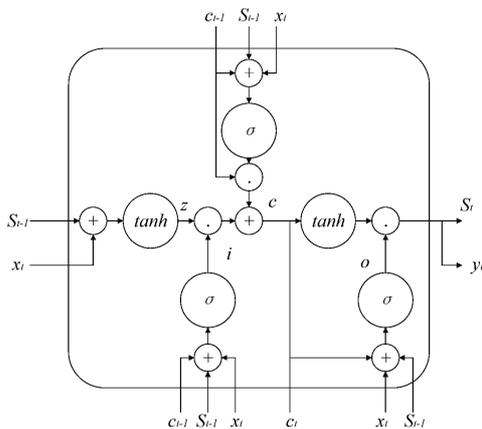


图 3 LSTM 隐藏层展开结构

LSTM 功能块的具体推导公式如下:

$$i_t = \sigma(W_{si}s_{t-1} + W_{ci}c_{t-1} + W_{xi}x_t + b_i) \quad (3)$$

$$f_t = \sigma(W_{sf}s_{t-1} + W_{cf}c_{t-1} + W_{xf}x_t + b_f) \quad (4)$$

$$z_t = \tanh(W_{sc}s_{t-1} + W_{xc}x_t + b_c) \quad (5)$$

$$c_t = i_t z_t + f_t c_{t-1} \quad (6)$$

$$o_t = \sigma(W_{so}s_{t-1} + W_{xo}x_t + W_{co}c_t + b_o) \quad (7)$$

$$s_t = o_t \tanh(c_t) \quad (8)$$

在结构图与公式中可以看出,LSTM 网络可以学习何时让误差值进入或离开 LSTM 功能块,具体原理如下:当输入门的权重取零值时,任何值都不能进入该块,当输出门为零时,该值也不会消失,当两个门都关闭时,该值将被困在 LSTM 单元格中,该值将不会增大或缩小,也不会影响当前时间序列的输出。因此,在向后传播过程中,就可以实现梯度跨时间步长向后传播,而不会发生梯度爆炸和消失。正是因为长期的短期记忆功能块的存在,LSTM 网

络具有学习和训练时间序列长期依赖性的能力。

LSTM 网络在训练数据集时有 3 个步骤:(1)根据前向计算公式计算模型的输出值;(2)根据定义的损失函数计算模型的误差,并利用反向传播算法更新各系统参数;(3)根据相应误差项,计算每个权重的梯度。

## 2.2 故障识别过程

海底观测网海缆故障状态分类识别包括学习训练与诊断测试 2 个过程。

学习训练是一个函数拟合的过程,在一定标准的训练样本上,设计实现智能诊断模型的参数调整,如本文使用的 LSTM 网络模型。智能诊断模型可看作是一个复杂非线性函数,通过不断训练学习样本数据,调整其函数参数,使模型能做到逼近真实的数据函数,从而准确预测或识别待诊断样本。而诊断过程是使用已训练过的模型来预测和识别待诊断样本,诊断过程可以通过识别分类能力来判断模型的性能。因此,模型试验需要分别对学习训练过程和诊断测试过程进行记录和分析。

学习训练过程与诊断测试过程均包括数据预处理、特征提取、模式识别 3 个部分。

对于海底观测网海缆故障定位而言,电能监控系统提供的电气测量数据保证了输入数据具有丰富的故障特征细节,同时对数据提前进行预处理和分析保障了网络的处理能力,使得 LSTM 网络得以充分发挥作用。

## 3 供电海缆故障定位方法

现有海底观测网络电力系统的拓扑结构可分为链式结构、两端供电结构以及网状结构。本文基于一种双端供电海底观测网原型系统拓扑结构进行试验,其拓扑图如图 4 所示,其中海岸基站 1、2 与远程控制中心相连,电能监控系统位于远程控制中心内。

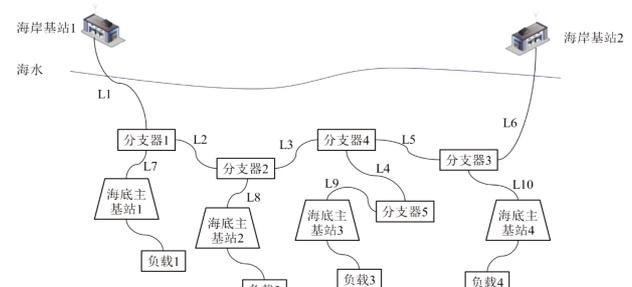


图 4 一种双端供电海底观测网原型系统拓扑结构

### 3.1 实验背景

选用两个直流稳压电源模拟海岸基站电源 (Power Feeding Equipment, PFE) 提供电能, 根据海缆的特性, 本文采用集中参数模型来模拟海缆。将海缆无限微分, 那么每一段海缆都可以视为由电阻、电感、电容组成的, 每段海缆之间是串联关系<sup>[8]</sup>。海缆模型如图 5 所示。

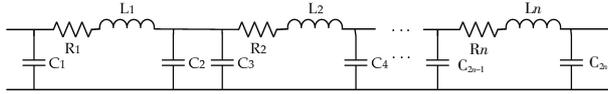


图 5 海缆模型示意图

其中每一段模拟海缆长度为 10 km, 各相应参数如下:  $R_n$  为 10  $\Omega$ ;  $L_n$  为 4 mH;  $C_{2n-1}$ 、 $C_{2n}$  为 1  $\mu$ F。自海岸基站 1 至海岸基站 2 的 5 段主干海缆长度设置为 170 km, 80 km, 10 km, 90 km, 80 km, 分支器与海底主基站之间的海缆长度为 10 km, 分支器 4 与分支器 5 之间的分支海缆长度设置为 100 km。分支器采用实验室自主设计的原型机, 利用恒功率负载模拟观测网的负荷, 其中, 负载 1~负载 4 分别设置为 150 W, 100 W, 50 W, 150 W。海底主基站单元为 400 V / 48 V 的直流变换器, 直流电源提供的电能经变换器后为恒功率负载供电。模拟海岸基站的电源输出电压为 400 V, 实验环境如图 6 所示。



图 6 海底观测网原型系统实验配置

针对图 4 所示拓扑结构, 将海底观测网原型系统海缆故障分为以下几类:

海底观测网在正常运行期间, 很少发生故障, 故障样本数量很少, 因此模型训练容易陷入过度拟合。当迭代次数增加时, 会造成在训练过程中, 网络可能对训练集有很好的拟合度, 但训练集的样本损失值很小, 对验证集的拟合性较差。针对这种情况, 在原型系统实验过程中, 本文通过人为设置接地电阻的方式, 在总长度 430 km 的主干海缆上随机选取故障点, 设置不同的高阻故障阻值模拟海缆故障实验, 进行网络模型训练。海底观测网原型系统高

表 1 故障模式

编码	故障海缆段
$S_1$	海岸基站 1 与分支器 1 之间 L1 段
$S_2$	分支器 1 与分支器 2 之间 L2 段
$S_3$	分支器 2 与分支器 4 之间 L3 段
$S_4$	分支器 5 与分支器 4 之间 L4 段
$S_5$	分支器 4 与分支器 3 之间 L5 段
$S_6$	海岸基站 2 与分支器 3 之间 L6 段
$S_7$	分支器 1 与海底主基站 1 之间 L7 段
$S_8$	分支器 2 与海底主基站 2 之间 L8 段
$S_9$	分支器 5 与海底主基站 3 之间 L9 段
$S_{10}$	分支器 3 与海底主基站 4 之间 L10 段

阻故障范围在 1 000~5 000  $\Omega$  之间。

### 3.2 基于 PCA 的数据预处理

海底观测网电力系统的测量数据较多, 电气测量参数之间存在着一定相关性, 增加了故障诊断难度。采用 PCA 方法对变量进行降维, 消除变量间的相关关系, 利用少数几个变量代替原来的多个变量, 同时又基本涵盖原始变量的全部内容。

首先, 剔除样本数据中的畸变数据, 用该点同一时间其他测量数据代替。本文主要涉及两个海岸基站 PFE 输出的电流, 以及四个海底主基站节点的输入电压、电流 10 个变量, 由高精度万用表在图 6 所示海底观测网原型系统各相应节点测得。每组变量分别有 100 组数据, 则构成一个 100 $\times$ 10 阶的数据矩阵, 建立变量的相关系数阵  $R$ , 求  $R$  的特征根  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$  及其对应的单位向量  $e_1, e_2, \dots, e_p$ 。

按照主成分的贡献率确定主成分数量, 其中主

成分的贡献率定义为:  $\frac{\lambda_i}{\sum_{k=1}^p \lambda_k}$ , ( $i=1, 2, \dots, p$ ), 得

到的协方差矩阵特征值、方差贡献率和累计贡献率如表 2 所示 (部分数据因有效数字原因省略未显示)。

由表 2 可知, 前 4 个主成分的累计贡献率已达到 99.99%, 能很好地概括原始变量, 所以选择前 4 个主成分作为输入神经网络预测的输入量。最后, 求出 100 组数据的前 4 个主成分的值。其中前 75 组作为训练样本集, 后 25 组作为测试样本集。

### 3.3 故障诊断识别过程

经过主成分分析后的海底观测网电气测量数据为输入, 将故障类别及发生距离为输出, 各属性值通过非线性变换同时结合 LSTM 时序联系。

表 2 主成分特征值与方差贡献率

编号	特征值	方差贡献率/%	累计贡献率/%
1	442.544 1	89.734 6	89.734 6
2	35.408 0	7.179 0	96.913 6
3	13.694 5	2.776 0	99.689 6
4	1.520 5	0.308 0	99.997 6
5	0.001 6	0.000 3	99.997 9
6	...	...	...
7	...	...	...
8	...	...	...
9	...	...	...
10	...	...	100

由于 LSTM 块构成的网络结构仅能训练获得权重或特征,具有学习和聚类能力,但并不具备对不同类型信号进行分类的能力,需要在深度结构的顶层加入具有判别能力的网络结构,本文选用 *Softmax* 回归模型。*Softmax* 回归模型是逻辑分类模型在分类问题上的推广,在分类问题上,类标签  $y$  可以取 2 个及 2 个以上的值。*Softmax* 回归模型的损失代价函数为:

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k \{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] \quad (9)$$

式中:  $(x^{(i)}, y^{(i)})$  为训练样本;  $m$  为样本个数,  $\theta$  为训练模型参数,使其能够最小化代价函数。由 *Softmax* 分类器诊断电力系统故障类别及发生位置的概率  $p(v_i) (i=1, 2, 3, 4)$ , 并根据最大概率  $\max(p)$  准则确定故障类别及故障定位距离。

基于 LSTM 网络的海底观测网络海缆高阻故障定位架构具体步骤如下:

(1) 将原始的故障文本数据进行特征提取和特征表示,并通过主成分分析后构成的训练集为:  $F_r = \{f_1, f_2, \dots, f_n\}$ 。

(2) 对训练集中的数据进行标准化,处理后的训练集表示为:  $X = \{X_1, X_2, \dots, X_L\}$ , 对应的输出位  $Y = \{Y_1, Y_2, \dots, Y_L\}$ 。

(3) LSTM 网络的初始化。给定初始权值矩阵,初始权值为均匀分布在  $[0, 1]$  之间的随机数,设置合理的最大迭代训练次数和最小误差值,最大迭代次数及最小误差值根据模型训练结果进行微调,找到最优训练网络的参数值。

(4) 前向计算过程。将  $X$  输入隐层, LSTM 隐层包括  $L$  个按时间顺序排列的 LSTM 记忆细胞,输入经过隐层之后的输出可以表示为:  $P = \{P_1, P_2, \dots, P_L\}$ ,

$P_l = LSTM_f(X_l, C_{l-1}, H_{l-1})$ , 其中  $C_{l-1}, H_{l-1}$  分别为前一个样本隐层的细胞状态和输出,  $LSTM_f$  是 LSTM 网络的前向计算函数。

(5) 误差反向传播。根据相应的误差项,计算每个权重的梯度,通过梯度下降法调整 LSTM 网络的权值和偏置,使网络误差不断减小,实现网络的优化。LSTM 网络的激活函数定义为 *sigmoid* 函数,输出的激活函数为 *tanh* 函数,导数分别为:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (10)$$

$$\sigma'(z) = \sigma(z)(1 - \sigma(z)) \quad (11)$$

式中:  $z$  为输出项。

(6) 当训练次数或误差值满足要求时模型停止训练,将未知测试样本输入建立好的模型中对模型分类准确度进行验证分析。为了衡量本文方法的性能,以故障定位误差值作为性能评价指标进行测试。

### 3.4 实验结果

训练过程中的训练损失值变化曲线,如图 7 所示。经过 140 次迭代,训练过程的损失迅速下降到平滑,趋近于 0。损失值的快速降低证明网络收敛明显,进一步说明了网络模型对故障定位函数拟合程度好,且故障数据对 LSTM 功能块适应能力强。损失值逐渐趋近于 0,达到  $10^{-5}$  的量级,也表示模型对现有故障数据的学习达到了较高的水平,其对于学习过的故障数据拟合结果达到了预期。

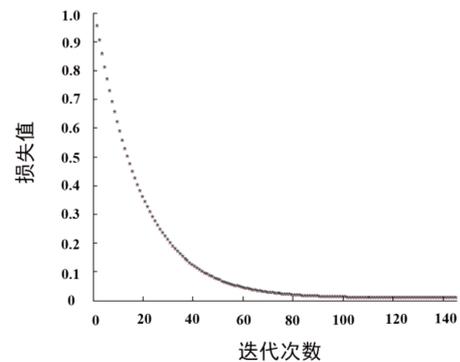


图 7 LSTM 网络模型 Error-Cost 曲线图

当 LSTM 网络模型训练好后,人工输入未使用的 25 个故障数据对模型进行检验,故障类别诊断正确率为 100%。对于故障定位的诊断精度,如图 8 所示,可以看出:当使用未学习的故障数据进行验证时,故障定位误差平均数据为 0.417 km,全部误差在 1 km 以内。试验结果表明,网络模型训练基本可以做到对海底观测网海缆高阻故障的精准定位,可运用到实际海底观测网海缆故障定位过程中。

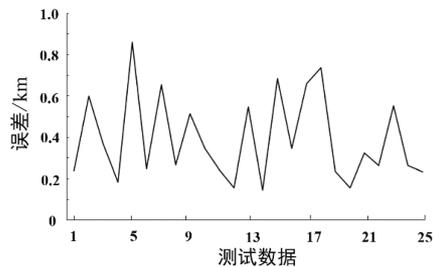


图 8 故障定位误差结果

而对于其他未学习的故障数据的诊断预测效果,需要靠测试准确率来体现,使用 K 折交叉验证方法(K-fold Cross Validation, K-CV)对模型结果进行评估,K-CV 方法是评估训练网络的经典方法。在 K-CV 中,原始样本被随机平均划分为  $K$  个子样本。每个子样本是一个测试集,其余的  $K-1$  个子样本是训练集。这些子样本将实验重复  $K$  次, $K$  个模型的平均性能被认为是最终性能。这种方法的优点是所有的观测值都用于训练和测试,从而避免了过度拟合和缺乏学习的情况,使得模型取得具有说服力的结果。最终结果可以通过准确率观察故障诊断的性能,准确率是正确的诊断样本与所有测试样本之比。如果多次练习的准确性都没有提高,则停止训练。然后,将练习的最佳精度视为实验的结果,重复实验的平均准确度记录为最终结果。

由于重复实验的准确性是稳定的,因此将练习次数设置为 40,优化器选择“RMSprop”优化器,它在 LSTM 网络中具有更好的性能。图 9 显示了基于 5-CV 的 LSTM 网络的故障预测结果。在 K-CV 方法中,经过 40 次训练的最终结果可以看出,5 个训练模型的平均准确度:91.43%。

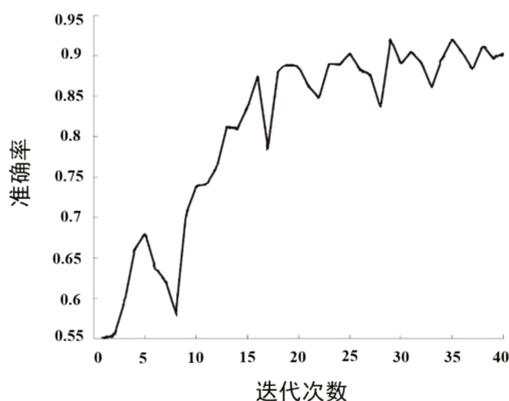


图 9 基于 5-CV 的 LSTM 网络的故障预测结果

为进一步验证本文所提出方法的优越性,基于原始数据,本文分别使用不同海底观测网海缆故障定位方法进行实验验证,表 3 显示了比较结果。

反向传播神经网络 (Back Propagation Neural Network, BPNN) 与 RNN 网络并不具备从互相连接的隐藏层单元中提取时间信息的能力,其准确率分别降低 23.2%及 8.9%,BPNN 方法误差值波动范围较大,RNN 方法误差值稳定在 2 km 以内。本文所提方法较使用原始数据直接输入到 LSTM 网络做法而言,结果准确率提高 1.8%,并且直接套用 LSTM 网络,所得结果误差值数据方差较大,数据集中度差。因此,使用基于 PCA/LSTM 网络的海缆故障定位方法,对于确保海底观测网的可靠性和稳定性是一项值得注意的改进。

表 3 不同诊断方法效果比较

	准确率	误差值
BPNN	68.2%	<5 km
RNN	82.5%	<2 km
LSTM	89.6%	<1 km
PCA/LSTM	91.4%	<1 km

## 4 结论

为了实现在海底观测网电力系统正常运行过程中,对海缆高阻故障的高精度在线定位,本文提出了基于 PCA/LSTM 网络的海缆故障定位方法。

该方法充分利用海底观测网电气测量数据,利用 PCA 消除变量相关性,再将处理后的数据输入 LSTM 网络进行训练,从而挖掘出海底观测网输配电中的海缆高阻故障特征与电气参数之间的对应关系,实现海缆故障精确定位。海底观测网原型系统实验表明该方法能够准确分类高阻故障位置,并且其定位准确率可达 91%左右,误差维持在 0.4 km 左右,相比较其他智能故障诊断方法而言,该方法在准确率及误差方面均有显著的性能提升,这对提高海底观测网的运行稳定性具有重要的意义。下一步将重点挖掘 LSTM 网络模型的故障预测能力,确保海底观测网故障的早期预防及处理。

## 参考文献:

- [1] Samuel P, Victor K. Observing the oceans[M]. Remote sensing and global environmental change, 2013.
- [2] Lyu F, Zhou H Y, Peng X T, et al. Technical preparation and prototype development for long-term cabled seafloor observatories in Chinese marginal seas [M]. Favali P, Beranzoli L, Santis A. Berlin. Seafloor observatories: A new vision of the Earth from the Abyss. Springer, 2015: 503-530.
- [3] 吕枫,周怀阳. 缆系海底科学观测网研究进展[J]. 工程研究—跨学科视野中的工程, 2016, 8(2): 139-154.
- [4] Yokobiki T, Iwase R, Takahashi Y, et al. Underwater repair method for shunt fault on submarine cable[C]//Bremmer, Jan N. Underwater Technology Symposium. Tokyo: IEEE, 2013: 1-4.
- [5] Liu C C, Schneider K, Kirkham H, et al. State estimation for the NEPTUNE power system [C]// Dallas, TX, USA. Transmission & Distribution Conference & Exposition. USA: IEEE, 2004: 748-754.
- [6] Lu S, El-Sharkawi M A. NEPTUNE power system: Detection and location of switch malfunctions and high impedance faults[C]//Jeong, Jinsoo. IEEE International Symposium on Industrial Electronics. Canada: IEEE, 2007: 1960-1965.
- [7] Schneider K, Liu C C, McGinnis T, et al. Real-time control and protection of the NEPTUNE power system [C]// Biloxi, MI. OCEANS '02 MTS. USA: IEEE, 2002: 1799-1805.
- [8] Yang F, Lyu F. A novel fault location approach for scientific cabled seafloor observatories [J]. Journal of Marine Science and Engineering, 2020, 8(3):190.
- [9] Nanayakkara K, Rajapakse A, Wachal R. Location of DC line faults in conventional HVDC systems with segments of cables and overhead lines using terminal measurements[J]. Power Delivery IEEE Transactions on, 2012, 27(1): 279-288.
- [10] Namdari F, Salehi M. A new algorithm for wide area fault location in power system [C]// 2015 30th International Power System Conference (PSC). Tehran: IEEE, 2015: 363-370.
- [11] Song G B, Chu X, Cai X L, et al. A fault-location method for VSC-HVDC transmission lines based on natural frequency of current[J]. International journal of electrical power & Energy systems, 2014: 63.
- [12] Tao Y, Zheng J, Wang T, et al. A state and fault prediction method based on RBF neural networks [C]// 2016 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO). Shanghai: IEEE, 2016: 221-225.
- [13] Rakhshani E, Sariri I, Rouzbehi K. Application of data mining on fault detection and prediction in Boiler of power plant using artificial neural network[C]//Energy and Electrical Drives. International Conference on Power Engineering. Lisbon: IEEE, 2009: 473-478.
- [14] Srivastava N, Mansimov E, Salakhutdinov R. Unsupervised learning of video representations using LSTMs[C]//International Conference on Machine Learning. Proceedings of the 32nd International Conference on Machine Learning, Lille: PMLR, 2015, 37: 843-852.
- [15] Chen K, Huo Q. Training deep bidirectional LSTM acoustic model for LVCSR by a context-sensitive-chunk BPTT approach[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2017, 24(7): 1185-1193.
- [16] 王鑫, 吴际, 刘超, 等. 基于 LSTM 循环神经网络的故障时间序列预测[J]. 北京航空航天大学学报, 2018(4): 772-784.
- [17] Zhang S, Wang Y, Liu M, et al. Data-based line trip fault prediction in power systems using LSTM networks and SVM [J]. IEEE Access, 2017: 7675-7686.
- [18] El-Sharkawi M A, Upadhye A, Lu S, et al. North east pacific time-integrated undersea networked experiments (NEPTUNE): cable switching and protection[J]. IEEE Journal of Oceanic Engineering, 2005, 30(1): 232-240.
- [19] 卢汉良, 李德骏, 杨灿军, 等. 深海海底观测网络远程电力监控系统研究[J]. 传感技术学报, 2011, 24(4): 564-569.
- [20] 吕枫, 周怀阳, 岳继光, 等. 缆系海底观测网电力系统结构与拓扑可靠性[J]. 同济大学学报: 自然科学版, 2013, 42(10): 1604-1610.
- [21] Tran V T, Nguyen K H, Bui D H. A Vietnamese language model based on Recurrent Neural Network [C]// 2016 Eighth International Conference on Knowledge and Systems Engineering (KSE). Hanoi: IEEE, 2016: 274-278.
- [22] Greff K, Srivastava R K, Koutník J, et al. LSTM: A search space odyssey [J]. IEEE Transactions on Neural Networks & Learning Systems, 2015, 28(10): 2222-2232.
- [23] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

## PCA/LSTM–Based Submarine Cable Fault Location Approach for Seafloor Observatory Network Power Systems

GENG Kun , LYU Feng

*Tongji University, School of Ocean and Earth Science, Shanghai, 200092, China*

**Abstract:** The submarine cables of seafloor observatory networks are prone to insulation failure. Since the cables operate in extreme subsea environments, achieving high–precision cable fault location is critical to reducing the maintenance costs. Based on the real–time electrical measurement parameters collected by the power monitoring system of submarine observatory networks, the Principal Components Analysis (PCA) and Long Short–Term Memory (LSTM) network is used in the submarine cable high–impedance fault location method proposed in this paper. The PCA is used for data dimensionality reduction, and then the processed data is put into the LSTM network for training. The LSTM network is applied to capture the time characteristics of multi–source data, mining the correspondence between cable fault characteristics and electrical system parameters to achieve accurate location of submarine cable faults. On the prototype system of the East China Sea seafloor observatories, the proposed approach is verified by K–fold Cross Validation (K–CV), and the experiments prove that the accuracy rate is about 91% and the fault location error is about 0.4 km.

**Key words:** seafloor observatory networks; power system; fault location; long short–term memory; principal component analysis