

基于健康画像的光通信设备故障预测算法

王峰¹,李兴华¹,李晓龙¹,刘瑞增²,庄浩涛^{3*},赵永利³

(1.国网宁夏电力有限公司 电力科学研究院,银川 750001;

2.国网宁夏电力公司 石嘴山供电公司,宁夏 石嘴山 753099;3.北京邮电大学,北京 100876)

摘要:针对光通信网设备风险预警无法进行人工分析的问题,提出基于用户画像技术与深度学习算法的设备故障预测算法。基于数据采集与数据增强构建设备健康画像,从含有脏数据且格式不统一的原始数据中抽取与设备故障相关联的标签序列,将序列数据“喂”入深度学习模型,获得高准确度的故障预测结果。仿真结果表明:该算法能够基于设备原始数据构建设备健康画像并实时训练深度学习模型,获得接近100%的设备故障预测准确率;与全标签序列的算法和未进行数据增强的算法相比,所提算法将故障预测准确率分别提升了7.8%和3.3%。

关键词:故障预测;健康画像;光通信设备;深度学习

中图分类号:TP181;TN830 文献标志码:A 文章编号:1002-5561(2021)09-0031-05

DOI:10.13921/j.cnki.issn1002-5561.2021.09.008

开放科学(资源服务)标识码(OSID):



Fault prediction algorithm of optical communication equipment based on health profiles

WANG Feng¹, LI Xinghua¹, LI Xiaolong¹, LIU Ruizeng², ZHUANG Haotao^{3*}, ZHAO Yongli³

(1. Electric Power Research Institute, State Grid Ningxia Electric Power Co., Ltd., Yinchuan 750001, China;

2. Shizuishan Power Supply Company, State Grid Ningxia Electric Power Co., Ltd., Shizuishan Ningxia 753099, China;

3. Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Aiming at the problem that the risk warning of optical communication network equipment cannot be analyzed manually, a failure prediction algorithm of equipment based on user profile and deep learning algorithm is proposed. Based on data collection and data enhancement, the device health profile is constructed, and the label sequence associated with the equipment failure is extracted from the original data containing dirty data and inconsistent format, and the sequence data is fed into the deep learning model to obtain highly accurate failure prediction results. The simulation results show that this algorithm can construct the equipment health profile based on the original equipment data and train the deep learning model in real time, and obtain the accuracy rate of equipment failure prediction close to 100%. Compared with the algorithm of full label sequence and the algorithm without data enhancement, the proposed algorithm improves the accuracy rate of failure prediction by 7.8% and 3.3%.

Key words: failure prediction; health profiles; optical communication equipment; deep learning

0 引言

随着光传输网规模的扩大,传输设备数量快速增

长且设备间依赖关系越发错综复杂。设备一旦发生故障,其影响范围将大规模蔓延。为规避风险,基于机器学习的预测算法在网络预警案例中得到了广泛应用。机器学习算法可发现复杂问题的潜在模式并构建较高准确率的事件预测模型^[1]。例如:文献[2]提出基于双指数平滑和支持向量机的波分复用(WDM)故障预测方法;文献[3]提出具备模型预测控制的体系结构以避免光传输系统中的激光故障;文献[4]通过训练高斯过程分类器预测每条光链路的告警概率。由于机器学习模型基于数据构建,数据集质量的好坏直接影响模型性能(稳定性、准确性等)高低。现在网络所产生的海

收稿日期:2020-11-15。

基金项目:2020年度第二批宁夏自然科学基金项目(2020AAC03487)资助;国网宁夏电力有限公司2020年研究开发项目计划(第一批)(5229DK19004W)资助。

作者简介:王峰(1989—),男,汉族,硕士,高级工程师,曾长期从事电力通信网运行检修、工程建设和技术监督管理工作,主要研究方向为电力信息通信技术、网络与信息安全技术和电力物联网等方面。

*通信作者:庄浩涛(1995—),男,硕士,研究员,主要研究方向为人工智能光网络。



量数据种类繁多,格式不一,无法进行人工分析^[5]。近年来,用户画像(UP)技术在大数据领域得到了广泛应用^[6-7],其核心为特征标签集合,即在具体应用场景下根据对象多个特征定义相应标签,并通过特征标签制定适用于该场景的策略或方案^[8-9]。

本文提出一种基于光通信设备健康画像的故障预测方法,对设备数据进行处理并生成设备健康画像,用标签序列训练基于深度学习(DL)的故障预测模型,获得高准确率模型对现在网络故障进行实时预测。

1 健康画像构建与故障预测

1.1 DL 算法与整体流程

广泛应用于预测方案的模型主要为时间序列分析模型,例如差分整合移动平均自回归(ARIMA)模型和深度神经网络(DNN)模型^[10]。这些模型在适当调整参数的情况下所有算法都能实现几乎相同的预测准确率。由于 DNN 易于部署且性能稳定,因此本文选择 DNN 对故障做出预测。作为一种经典 DL 算法,DNN 对人脑功能进行简单建模,具有高度的适应性、强大的非线性复杂映射能力以及对大规模并行计算的高性能^[11]。本文采用 DNN 中最为常用的全连接反向传播神经网络(FCNN)构建预警模型。

FCNN 具有输入层、隐藏层和输出层。隐藏层越多的 FCNN 性能越不稳定,容易发生拟合现象^[12]。每一层神经元包含可调参数的激活函数,用于整合输入值并输出到下一层。全连接指相邻层的所有神经元通过权重值完全连接,以使下一层每个神经元可收集上一层所有神经元输出的全部信息,进而提取出更高级别的信息。因此,FCNN 具有强大的能力来查找输入和输出之间的非线性复杂映射。每个神经元中的参数根据损失值以反向传播的方式进行调整,即所有参数的梯度更新是从后层到前层。损失值表示训练集的输出与 FCNN 模型的输出之间的差异。通过迭代,FCNN 可获得稳定而强大的性能,并使预测模型具有较高的准确率和低损失值。

光通信设备故障预测模型整体构建流程如图 1 所示。整个流程主要分为三部分:数据处理、画像构建和故障预测。其中,数据处理包括数据预处理和数据增强;画像构建包括特征标签抽取和健康画像封装;故障预测由 FCNN 构成。

1.2 数据处理

1.2.1 数据采集与预处理

数据采集基于套接字(Socket)实现,采集到的数据依据端口映射投放到 Kafka 组件中对应的 topic 里。原始数据通常包含许多残缺数据、错误数据和重复数据。对脏数据直接分析将得到不可靠结果,因此数据预处理是必要的步骤。预处理进程直接从 Kafka 的 topic 中消费原始数据,数据预处理过程主要包括数据脱敏、数据清洗、数据规约和数据转换 4 个步骤。数据脱敏是对采集到的原始数据中的敏感信息(业务信息、站点位置和设备名称等保密信息)通过哈希算法进行数据的单向转换,实现敏感隐私数据的可靠保护。数据清洗是对脏数据进行冗余清除、缺失填入和错误消除等处理。数据规约是为了获得一个相当小原始数据集,并保持相似的特征。数据转换包括平滑处理、属性构造、编码、规范化和离散化等。

1.2.2 设备数据增强

数据预处理后,数据量急剧下降,无法支撑健康画像的构建。本文引入了数据增强算法用于扩增设备数据量,提高健康画像精度。针对设备相关数据,如日志数据、性能数据等,目前有 2 种有效的增强算法:①通过手动分析文本数据特性,对突变数据进行值抬升处理,对缓变数据进行噪声叠加处理^[13];②基于生成式对抗网络(GANS)的告警数据增强算法。鉴于第一种算法需要人为操作,无法自动化调控抬升值和噪声的量,极易引入人为干扰因素,因而不作为本文的选择。第二种算法由 2 个模型组成:判别模型,用于识别样本是否来自真实数据集;生成模型,用于生成与真实数据类似的数据^[14]。这 2 个模型不断进行博弈式学习,最终得到一个高性能、高可靠的生成模型,用于生成足够量的画像数据。由于第二种算法具有自监督特性,不会引入人为干扰因素,故本文采用第二种算法进行深度学习数据量的扩充,进而保证光通信设备健康画像的高精度。

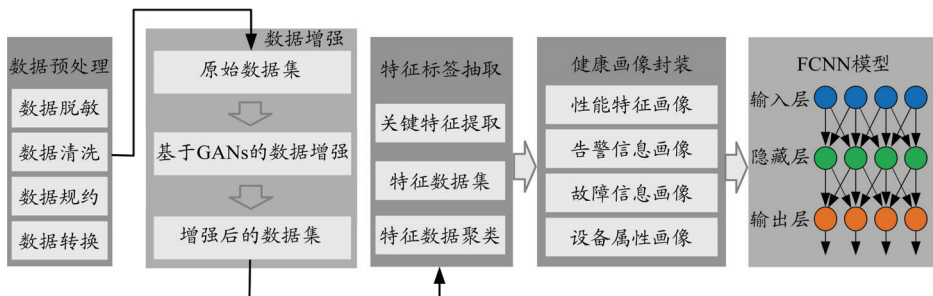


图 1 基于健康画像的光通信设备故障预测流程

1.3 健康画像构建

基于画像技术,本文对反映设备健康状况的运维数据进行挖掘,确定光通信设备健康特征标签并形成健康画像。

标签集合可分为属性集和行为集。属性集表征设备固有特性,如设备信息、功能和级别等。设备属性标签集如图2所示,其中基本标签包括设备全局唯一标识码、类型、厂商信息和机房信息等;功能标签对应光通信设备各功能模块,如主控单元、交叉矩阵等功能标签;级别标签对应设备所在网络级别。行为集主要包含行为标签,用于表示设备运行状态,如性能抖动、告警触发等。行为标签是设备运行状态的时序描述,例如用于表征设备某一功能的性能劣化序列。行为标签主要采用关联分析算法进行挖掘。

基本标签	功能标签	级别标签
全局唯一标识码	主控单元	骨干网级别
类型	交叉矩阵	区域网级别
厂商信息	定时电路	省级节点
机房信息	线路接口	市级节点

图2 光通信设备属性标签集

设备健康画像构建流程如图3所示,本质上,健康画像的构建是对光通信设备诸多特征标签实施再组合与分类应用,具体构建流程主要包括序列标注、属性标签抽取与整合以及行为序列提取与挖掘。标注系统主要基于聚类算法,即把预处理后的数据进行聚类分析,获得每个标签集合中各标签所有数据,并由经验算法或者人工进行验证;设备属性标签主要以全局唯一标识码作为主要标签,同时通过设备文档资料获取设备详细功能标签信息,如告警信息(网元连接

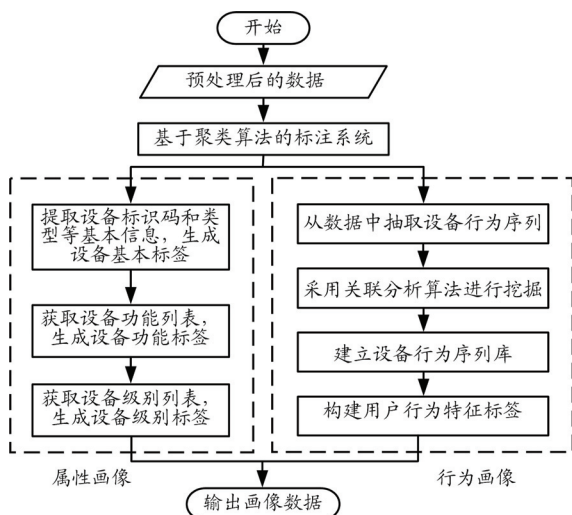


图3 光通信设备健康画像构建流程

失败、ICMP/SNMP 不可达和主备网元发生倒换等),故障信息(板卡故障、设备掉电和网元控制板卡故障等)和事件相关的设施(光纤、产品编号和用户ID)等。设备运行过程中所涉及的各类行为都具有时序性,如一个“功率激增”的告警具有多个时间戳(发生时间、处置时间和恢复时间),因而设备行为标签的提取是在时间维度上进行的。该过程主要采用关联分析算法对性能、告警以及故障进行关联,形成设备行为序列。

1.4 传输设备故障预测模型

采用 FCNN 模型对健康画像中具备时序性的行为序列进行学习,获取设备行为特性。当 FCNN 模型获得较高准确率(模型预测正确的行为序列数与总行为序列的比值)后,即可高效预判设备的行为信息。经过多次调试,本文采用以下 FCNN 模型参数:

①层数与神经元个数。通过多次仿真测试,隐藏层过多的 FCNN 训练过程不稳定,倾向于过度拟合。本文总共设置4层,第一层为33个神经元的输入层,中间2层分别是具有64、128个神经元的隐藏层,最后一层为输出层。其中,第一层输入层中第一个神经元接收故障类型数据,其它32个神经元接收故障发生15 min前的32个离散性能序列。

②激活函数和损失函数。采用整流线性单元(RLU)作为激活函数,均方差函数(MSE)作为损失函数。

③优化器和学习率。采用 Adam 优化算法作为优化器,学习率设置为 2×10^{-3} 。

2 仿真与分析

本文所提故障预测方案的实现主要基于 Spring 和 TensorFlow,数据采集与处理模块、数据增强模块和画像构建模块由 Spring 集成管理,预测模块由 TensorFlow 算法库提供。本次仿真实验通过骨干网采集了覆盖全国20个省的海量设备数据,其中包括几千万组性能数据。

数据经预处理和增强后,本文进行设备健康画像的构建。对其中3类属性标签(节点、连接和板卡)和3类行为标签(历史告警、当前告警和性能标签)关系进行可视化,如图4所示。在设备行为标签抽取中,获得7类故障标签:板卡故障、光缆中断、时钟故障、网元控制板卡故障、设备掉电、线路衰耗过大和基础环境劣化。经过行为标签关联分析,获得性能、告警与故障之间的关联关系。一个设备行为的标签关联关系图如图5所示,其中IN_PWR_LOW告警表示输入功率过低。

本文选择时钟故障和线路衰耗过大作为目标故

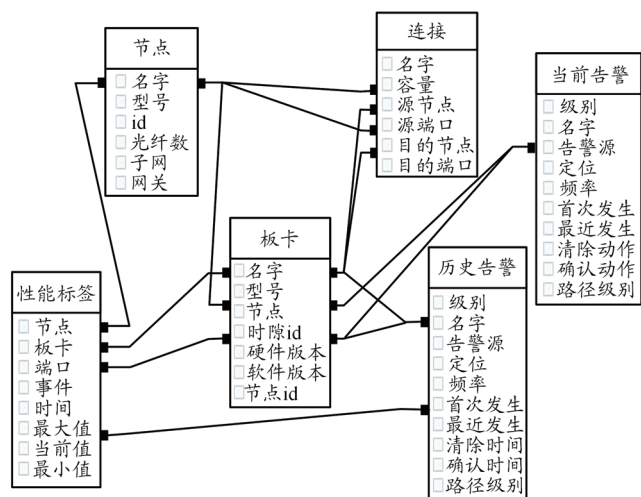


图4 部分设备属性标签与行为标签关系

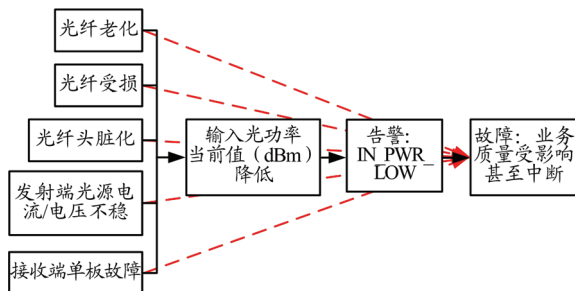


图5 性能、告警与故障之间的关联图

障类型。根据设备健康画像行为标签的关联分析,选定最大光输入功率、泵浦激光工作电流和光输出功率3类最大关联度的性能标签作为 FCNN 的输入。每 15 min 对采集到的数据进行一次存储,因此设定性能序列的第一个离散数据时间戳为故障发生前 15 min。数据预处理后,筛选出 58 组性能标签序列集,记为原数据集(OD-set)。OD-set 中 46 组作为 FCNN 训练集,12 组作为 FCNN 的测试集。数据增强后数据量扩增至 6000 组,记为增强数据集(AD-set)。AD-set 中设 4500 组为训练集,1500 组为测试集。同时,再选定关联度较低的另外 3 类性能标签序列,构建具备 6 个性能特征的数据集,筛选出 116 组序列组成对比数据集,记为随机数据集(FD-set),如表 1 所示。为了评估故障预测准确率,本文采用未进行数据增强的故障预测(ODFP)算法

表 1 AD-set, OD-set 和 FD-set 数据量

数据集	训练集数据/组	测试集数据/组
OD-set	46	12
AD-set	4500	1500
FD-set	92	24

和全标签序列的故障预测(FDFP)算法这 2 种算法为对比基准。本文所提方案基于 AD-set,记为数据增强式故障预测(ADFP)算法。

根据 1.4 节,配置参数相同的 3 个 FCNN 模型并进行模型训练和测试。ODFP、ADFP 和 FDFP 3 种算法准确率对比情况如图 6 所示。可以看出,随着迭代次数的增加,3 条准确率曲线都呈现上升趋势。FCNN 模型能够从设备标签数据中学习到设备性能值与故障之间的关系。该模型所学到的规律体现为模型中不断调整的神经元参数,比如权值和偏置。当模型相关指标(例如准确率和损失值)得到收敛,模型中所有神经元参数的变化趋于稳定,也即所学习到的性能与故障之间的关联关系得到明确。ADFP 算法曲线呈现高斜率快收敛的特点,意味着 FCNN 模型具备最强的学习性能。ADFP 算法准确率稳定保持在 99.9%左右,ODFP 算法准确率收敛于 96.6%,而 FDFP 算法收敛于 92.1%。ODFP 算法准确率比 FDFP 算法低了 4.5%,说明引入过多无关特征标签会影响模型对性能标签与故障标签之间关联信息的学习。ADFP 算法准确率比 ODFP 算法高出 3.3%,说明增强后的数据有助于模型学习到更多关联信息。虽然 ADFP 算法比 FDFP 算法仅高出约 7.8%,但这部分提升对于大型光传输网及传输设备的稳定运行具有重要意义。ODFP、ADFP 和 FDFP 3 种算法损失值对比如图 7 所示。3 种算法损失值曲线均随着迭代次数的增加而逐渐下降,最终达到收敛,说明 3 种算法均能从 3 个训练集中挖掘到信息。FDFP 算法的损失值曲线收敛值小于 ODFP 算法,但 FDFP 算法准确率却低于 ODFP 算法,这是因为 FDFP 算法数据集中引入了无关信息,对模型的学习性能造成影响,即导致过拟合结果。ADFP 算法的损失值曲线较为平滑且下降速度最快,最后几乎收敛于 0。

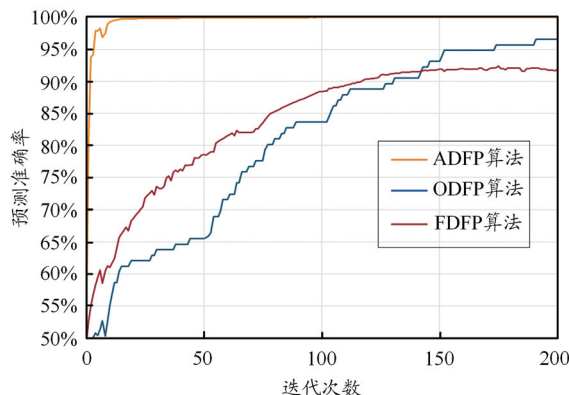


图6 ODFP、ADFP 和 FDFP 3 种算法准确率对比

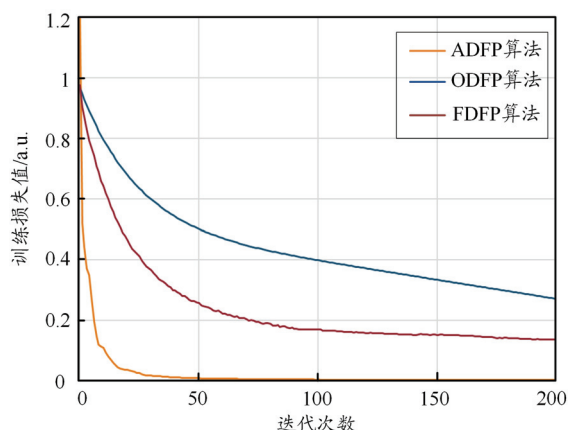


图7 ODFP、ADFP 和 FDFP 3种算法损失值对比

相较于 ODFP 算法和 FDFP 算法,ADFP 算法具有最快的收敛速度和学习稳定性。

从 ODFP 算法和 FDFP 算法的仿真结果对比中可发现:在设备健康画像构建过程中,设备行为序列得到了有效关联,防止模型出现过拟合现象。相较于 ODFP 算法,ADFP 算法具有高准确率和稳定性,表明数据增强算法改善了数据集的质量,防止模型出现欠拟合现象,这也侧面说明优质的数据在预测模型训练过程中的重要性。

3 结束语

本文针对光传输网中设备预警问题,提出了一种基于健康画像的光通信设备故障预测方法,同时采用数据增强算法增量了原始数据集。仿真实验结果表明:与传统 FCNN 预测模型相比,基于健康画像的 FCNN 预测模型具有 100% 的故障解决准确率,这有助于获取高准确率的设备预警信息以规避光通信设备运行风险,从而提高光传输网的稳健性。

参考文献:

- [1] ZHAO Y, YAN B, LI Z, et al. Coordination between control layer AI and on-board AI in optical transport networks [Invited][J]. Journal of Optical Communications and Networking, 2020, 12(1): 49–57.
- [2] WANG Z, ZHANG M, WANG D, et al. Failure prediction using machine learning and time series in optical network [J]. Optics Express, 2017, 25(16): 18553–18565.

- [3] BARLETTA L, GIUSTI A, ROTTONDI C, et al. QoT estimation for unestablished lighpaths using machine learning [C]// Optical Fiber Communications Conference&Exhibition, March 19–23, 2017, Los Angeles, CA, USA. IEEE, 2017: 1–3.
- [4] PANAYIOTOU T, CHATZIS S P, ELLINAS G. Leveraging statistical machine learning to address failure localization in optical networks[J]. IEEE/OSA Journal of Optical Communications and Networking, 2018, 10(3): 162–173.
- [5] WANG F, ZHAO Y L, WANG W, et al. Machine learning with service classification for detecting control plane intrusions in software defined optical networks [C]//23rd Opto-Electronics and Communications Conference, July 2–6, 2018, Jeju Island, Korea (South). Piscataway: IEEE, 2019: 1–2.
- [6] LEGG P A, BUCKLEY O, GOLDSMITH M, et al. Automated insider threat detection system using user and role-based profile assessment [J]. IEEE Systems Journal, 2017, 11(2): 503–512.
- [7] BHARAT K, LAWRENCE S, SAHAMI M. Generating user information for use in targeted advertising: 9235849[P]. 2016-01-12.
- [8] ELALLIOUI Y, BEQQALI O E. User profile ontology for the personalization approach [J]. International Journal of Computer Applications, 2013, 41(4): 31–40.
- [9] OUAFTOUH S, ZELLOU A, IDRI A. User profile model: A user dimension based classification[C]//International Conference on Intelligent Systems: Theories & Applications, October 20–21, 2015, Rabat, Morocco. Piscataway: IEEE, 2015: 1–5.
- [10] WANG L, ZENG Y, CHEN T. Back propagation neural network with adaptive differential evolution algorithm for time series forecasting [J]. Expert Systems with Applications, 2015, 42(2): 855–863.
- [11] TICKNOR J L. A bayesian regularized artificial neural network for stock market forecasting [J]. Expert Systems with Applications, 2013, 40(14): 5501–5506.
- [12] YAN B, ZHAO Y, RAHMAN S, et al. Dirty-data-based alarm prediction in self-optimizing large-scale optical networks[J]. Optics express, 2019, 27(8): 10631–10643.
- [13] CUI L, ZHAO Y, YAN B, et al. Deep-learning-based failure prediction with data augmentation in optical transport networks[C]//17th International Conference on Optical Communications and Networks (ICOON2018), November 16–19, 2018, Zhuhai, China. Zhuhai: International Society for Optics and Photonics, 2019: 110482–110488.
- [14] ZHUANG H, ZHAO Y, YU X, et al. Machine-learning-based alarm prediction with GANs-based self-optimizing data augmentation in large-scale optical transport networks[C]//2020 International Conference on Computing, Networking and Communications (ICNC), February 17–20, 2020, Big Island, USA. Piscataway: IEEE, 2020: 294–298.