第 37 卷 第 5 期
2022 年 10 月

天津科技大学学报
Journal of Tianjin University of Science & Technology

Vol. 37 No. 5
Oct. 2022

DOI:10.13364/j.issn.1672-6510.20220036

基于 W-ReLU 的设备多工况状态异常检测方法

张翼英[1]，王鹏凯[1]，柳依阳[1]，武延年[2]，郭晓艳[3]
（1. 天津科技大学人工智能学院，天津 300457；2. 深圳市国电科技通信有限公司，深圳 518109；
3. 国网天津市电力公司信息通信公司，天津 300000）

摘 要：针对光通信设备在多工况条件下由于健康状态的不确定性导致难以及时发现运行异常的问题，对基于监控日志数据量化分析的设备异常检测方法进行研究，提出了基于加权线性修正函数（W-ReLU）的设备多工况状态异常检测方法. 首先采用滑动时间窗口机制对告警日志进行划分，并根据每条日志的类型，将日志子集表征为日志特征向量；然后对健康状态下的样本点进行密度峰值聚类分析，以构建设备的基准健康状态矩阵；最后采用 W-ReLU 非线性映射模型量化评估设备的异常度，并据此进行异常检测. 结果表明，与现有的相似性度量方法相比，该方法具有更高的准确率和稳定性.

关键词：光通信设备；异常检测；特征权重；样本相似性
中图分类号：TN929；TP181　　**文献标志码**：A　　**文章编号**：1672-6510（2022）05-0063-08

Equipment Abnormal Detection Method for Dynamic Working Conditions Based on W-ReLU

ZHANG Yiying[1]，WANG Pengkai[1]，LIU Yiyang[1]，WU Yannian[2]，GUO Xiaoyan[3]
（1. College of Artificial Intelligence，Tianjin University of Science & Technology，Tianjin 300457，China；
2. China Gridcom Co.，Ltd.，Shenzhen 518109，China；
3. Information and Communication Company，State Grid Tianjin Electric Power Company，Tianjin 300000，China）

Abstract：An equipment abnormal detection method based on quantitative analysis of monitoring log data was studied to address the problem that it is difficult to find the abnormal status of optical communication equipment in time due to the uncertainty of the equipment health status under dynamic working conditions. To this end，in this article we propose an equipment abnormal detection method for dynamic working conditions based on weighted rectified linear function（W-ReLU）. In this method，the alarm logs were first divided by sliding time window mechanism，and log subsets were represented as log feature vectors based on the type of each log. Then，cluster analysis of density peaks was performed on the sample points under normal operation conditions to construct the baseline health state matrix. Finally，W-ReLU was used to quantitatively evaluate the anomaly degree of the equipment and perform anomaly detection accordingly. Experimental results show that the proposed method has higher accuracy and stability than the traditional similarity measurement methods.

Key words：optical communication equipment；abnormal detection；feature weight；sample similarity

光通信设备是电力通信网络以及各类主干网络的核心基础设备，保障该设备的稳定运行对于通信网络具有重要意义. 近年来，随着大数据和人工智能的发展，数据驱动的智能运维技术[1]成为相关领域的研究重点. 针对光通信设备的智能运维技术研究，国内外的研究人员从多个角度对设备告警日志的压缩以及序列模式挖掘[2-6]（如基于数据增强和深度学习模型的告警预测[7-8]、基于告警信息的故障预测与诊

收稿日期：2022-02-28；**修回日期**：2022-03-20
基金项目：国家自然科学基金青年基金项目（61807024）
作者简介：张翼英（1973—），男（满族），辽宁盘锦人，教授；**通信作者**：王鹏凯，硕士研究生，pk365726186@163.com

断[9-10]等方面）开展研究. 这些研究在一定程度上消除了光通信设备传统运维方式的弊端. 然而，在光通信设备的实际运行过程中，设备出现异常状态的概率比较小，正样本和异常样本存在极端不平衡的情况，从而导致基于分类模型的泛化性能较差.

异常检测作为一种单一分类技术（one-class classification，OCC），通过检测样本点是否符合已有正样本的数据分布，实现样本点的划分. 异常检测算法包括基于概率统计的方法、基于最近邻的方法以及基于聚类的方法[11]. 基于概率统计的方法对分布模型的依赖程度较高. 基于最近邻的方法和基于聚类的方法需要进行样本间的相似性计算，这种方式受到样本点多维数据分布差异的影响，采用不同的相似性度量方式在特定的 OCC 任务中往往有不同的结果[12]，从而影响算法的准确率. 此外，基于深度学习和构建样本超球体[13]的异常检测方法，利用多层网络结构以及核函数等非线性映射，将原始特征变换为抽象特征，并利用判别函数实现异常样本点的识别. 虽然这种方法在某些任务中取得了较好的效果，但是由于其“黑盒”特性，导致其具有决策风险，不适用于需要较高可靠度的通信及工业领域的异常检测.

本文的研究思路是通过比较实时告警日志样本点与设备健康状态下的代表性样本点之间的相似性是否超出异常判定的阈值，实现日志模式的异常检测. 这种方法的检测效果取决于特定工况条件下的代表性样本点的选取以及样本间的相似性度量方法. 通过对设备健康状态下的历史告警日志样本集进行密度峰值聚类分析[14]（clustering by fast search and find density peaks，DPC），取类别簇中心样本点作为健康状态下的代表性样本点，并考虑特征的分布以及每条告警的重要性定义特征的组合权重，实现特征加权处理；最后采用加权线性修正函数（W-ReLU）计算样本点与代表性样本点的异常度，消除部分特征变化对计算结果的影响. 以此提高日志异常检测的准确率，并降低误警率.

1　基于日志特征的设备状态分析

光通信设备由各种类型的单板以及电源、风扇等辅助设备组成，是一种典型的具有复杂结构的技术密集型系统设备. 该设备在运行时，其健康状态受到设备自身和外界环境因素的影响较大. 运维人员主要通过分析网管系统的各种监控日志对其健康状态进行判断. 然而，由于日志关联性[15]以及告警门限的人为设置等因素，日志中的大多数信息对于判断设备的实时健康状态没有意义，因此采用人工分析的方式难以确定设备是否健康.

对监控数据的变化情况进行分析. 在光通信设备无出厂缺陷、安装配置符合部署规范以及运行寿命期处于稳定的条件下，当日志特征发生变化时，其可能的原因包括两种：一种是设备的运行工况特性发生变化，即网络承载的业务量和负载等产生变化；另一种是设备在某一工况特性下，出现了性能劣化和故障等异常状态. 设备的运行状态与监控数据的变化关系如图 1 所示.

因此，在对设备的运行状态进行异常检测时，首先要排除运行工况发生变化而导致的告警模式异常的情况，即确定设备的实时工况特性，然后再分析告警模式与当前工况条件下的健康状态告警模式是否相同. 当两者差异较大时，表明设备的运行状态发生了异常.

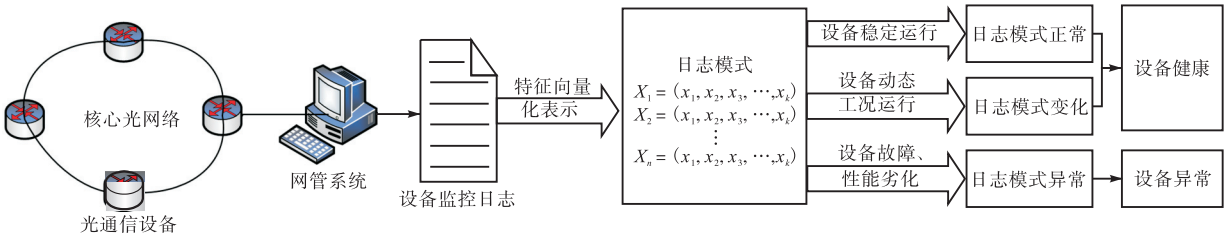


图 1　设备运行状态与监控日志异常分析

Fig. 1　Abnormal analysis of equipment operation status and monitoring logs

2　日志模式异常检测

2.1　总体架构

本文提出的告警日志模式异常检测方法包括离线模型构建和实时异常检测. 在离线模型构建阶段，对健康状态下的样本点进行 DPC 聚类分析以及统计分析，得到模型参数；在实时异常检测阶段，基于 DPC 聚类原则确定设备的当前工况，然后计算当前工况条件下的样本点异常度，判断设备是否异常，从

而实现多工况条件下的告警模式异常检测，总体架构如图 2 所示.

2.1.1　离线模型构建

(1) 对健康状态下的历史告警日志按照固定大小的滑动时间窗口进行划分，统计划分后的日志子集中各类告警的数量，将日志子集表征为特征向量，得到健康状态下的历史样本集合.

(2) 对历史样本集合中的样本点进行 DPC 聚类，取聚类簇中心构建基准健康状态矩阵，并将每个样本点归类.

(3) 基于组合权重法确定每个工况条件下的特征权重，并基于所有样本点异常度的均值统计值确定判定阈值.

2.1.2　实时异常检测

(1) 计算待测样本点与基准健康状态矩阵中每个向量的距离，取距离最小值所对应的簇中心作为待测样本点的健康基线，实现工况识别.

(2) 计算待测样本点的加权线性修正函数值，得到样本点异常度，若其大于阈值，则判定为异常，否则判定为正常.

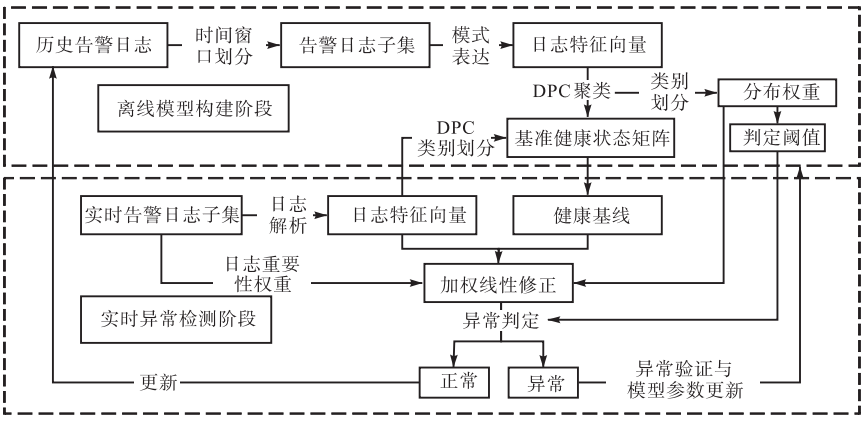


图 2　设备实时异常检测总体架构

Fig. 2　Framework of real-time abnormal detection for equipment

2.2　日志子集划分与向量表示

告警日志是一种结构化的时间序列数据，包含告警名称、告警等级以及产生时间等字段. 为了满足异常检测实时性的要求，需要对一段时间内的告警模型进行分析. 滑动窗口机制是一种针对序列数据的处理方法，能够根据序列的产生时间或者元素个数来对全量序列数据进行划分[16]. 因此，本文设计了滑动时间窗口对历史告警日志数据进行划分，图 3 所示为产生 3 个告警日志子集的实例.

滑动时间窗口的属性包括窗口的长度和步长，窗口长度表示每次划分的时间跨度，步长表示每次滑动的时间间隔. 一般设置步长小于窗口长度，以获取较多的日志子集，并保证每条告警都被划分到相应的日志子集中.

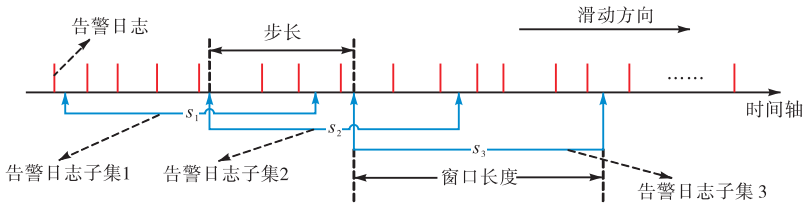


图 3　滑动时间窗口对历史告警日志的序列划分

Fig. 3　Sequence division of historical alarm logs by sliding time window

在完成告警日志子集划分之后，对其进行向量化表示. 首先对已知的告警类型按照每种类型所反映的不同信息维度，将告警日志划分为连接域、配置域、主设备域、性能域和辅助设备域 5 种不同的日志类型，然后统计日志子集中不同日志类型的数量. 将不同类型的告警数量作为特征值进行向量化表示，向量中每个维度的变量表示及其含义见表 1.

2.3　基准健康状态矩阵构建

对光通信设备在不同工况条件下的告警特征变化情况进行分析. 假设在某一工况条件下，设备正常运行时，日志的特征向量存在一个最优值，此时设备状态是最健康的，那么在多工况条件下，设备的健康状态则对应一个最优值特征向量集合. 本文基于健康基线[17]的概念，定义基准健康状态矩阵，对设备在

多工况条件下的健康状态进行模式表达.

表 1　光通信设备监控日志类型划分

Tab. 1　Classification of monitoring logs for optical communication equipment

日志特征	变量	含义
连接域	x_1	信号消失、帧丢失及自动倒换的相关告警数量
配置域	x_2	业务配置、单板配置、网元状态以及对端告警指示的相关告警数量
主设备域	x_3	激光器光功率异常、单板状态性能的相关告警数量
性能域	x_4	指针调整、指针丢失、误码越限、时钟源丢失的相关告警数量
辅助设备域	x_5	风扇、电源、继电器以及环境检测类的相关告警数量

假设光通信设备在运行时存在 m 种工况特性，在某一工况特性 c 的条件下，设备的健康基线为 $X^c=\left(x_1^c,x_2^c,\cdots,x_k^c\right),c=1,2,\cdots,m$，其中 x_i^c 表示在工况特性 c 下第 i 个特征的最优值. 基准健康状态矩阵 $\boldsymbol{D}_{m\times k}$ 可表示为

$$\boldsymbol{D}_{m\times k}=\begin{bmatrix}X^1\\X^2\\\vdots\\X^m\end{bmatrix}=\begin{bmatrix}x_1^1 & x_2^1 & \cdots & x_k^1\\x_1^2 & x_2^2 & \cdots & x_k^2\\\vdots & \vdots & & \vdots\\x_1^m & x_2^m & \cdots & x_k^m\end{bmatrix}$$

该矩阵存储了设备在不同工况条件下的健康基线向量集合，因此基准健康状态矩阵实际上是代表性日志特征向量的集合. 本文基于 DPC 算法对健康状态下的样本点进行聚类分析，并取各个类别的聚类中心构建基准健康状态矩阵.

DPC 算法的伪代码（密度峰值聚类算法）：

输入：训练数据集 $D=(X_1,X_2,\cdots,X_n)$；截断距离 r_c.

输出：每个样本点 X_i 的局部密度 ρ_i 和聚类中心距离 δ_i；每个点的类别.

1. for i：$1\to n-1$.

2. for j：$i+1\to n$.

3. 计算 X_i 与 X_j 的距离.

4. for i：$1\to n$.

5. 计算 X_i 的局部密度 ρ_i.

6. 计算 X_j 的聚类中心距离 δ_i.

7. 以 ρ_i 为横坐标、δ_i 为纵坐标画出决策图，选择具有较大 ρ_i 和 δ_i 的数据点作为聚类簇中心.

8. 将其余点归类为距离最近且局部密度比自身大的数据点所在的簇.

DPC 聚类首先需要计算样本点的局部密度，对于连续型的特征变量，可采用高斯核函数计算样本点的局部密度 ρ_i，公式为

$$\rho_i=\sum_{j\in I_S,j\neq i}\exp\left(-\frac{r_{ij}^2}{r_c^2}\right)\tag{1}$$

式中：I_S 表示特征集，r_{ij} 表示样本点 i 和样本点 j 的欧氏距离，r_c 为截断距离. 基于 DPC 聚类得到的簇中心具有局部密度最大和类间距离最大的特点，因此 DPC 聚类可以从正样本中学习健康状态的代表性日志特征向量.

2.4　设备异常度的量化评估

设备的异常度可以通过计算样本点与健康基线之间的相似性得到. 然而，传统的相似性度量方法大多基于向量之间的各种距离度量，这种方法没有考虑不同特征的权重因子，也没有考虑特征的不同变化对异常度量化值的影响，因此并不能准确地反映设备当前状态与健康状态相比的异常程度. 针对上述问题，本文提出一种 W-ReLU 非线性映射模型，对设备的异常度进行量化评估.

2.4.1　组合权重计算

在日志的多维特征中，不同特征的变化对设备健康状态具有不同的影响，因此需要对原始样本点进行加权处理，即对设备健康状态影响较大的日志特征赋予更大的权重. 本文所设计的组合权重 w_i 的计算公式为

$$\begin{cases}RV_i=\dfrac{\overline{x}_i}{\sigma_i}\\[2ex]AW_i=\dfrac{N_c\cdot\sum\limits_{a_v\in S_i'}l(a_v)}{x_i'\cdot\sum\limits_{a_u\in D_i^c}l(a_u)}\\[2ex]w_i=\dfrac{\theta\cdot RV_i}{\sum\limits_{i=1}^{k}RV_i}+\dfrac{(1-\theta)AW_i}{\sum\limits_{i=1}^{k}AW_i}\end{cases}\tag{2}$$

式中：σ_i 为特征 x_i 的标准差；$\overline{x}_i$ 为特征 x_i 的平均值；k 为特征维度；N_c 为工况特性 c 下正样本点的个数；x_i' 为待测样本点的特征值；$l(a)$ 为告警 a 反映设备状态的重要性，设定 $l(a)=\{1,2,3,4\}$ 分别对应 4 种具有不同等级的告警重要性；S_i' 和 D_i^c 分别为第 i 个维度的待测样本点日志子集和设备当前工况 c 条件下的全体日志集合.

特征权重由两部分组成，包括特征的分布权重和重要性权重. 分布权重由特征值的反变异系数决定，这里定义反变异系数为变异系数[18]的倒数. 在历史数据中，若某一维度的特征值分布较为集中，则在异

常检测中具有较高的权重. 特征的重要性权重由待测样本点的告警重要性的均值决定. 在实际应用中，通过比例参数 θ 调整两者的比例.

2.4.2　W-ReLU 非线性映射模型

衡量两个样本点之间的差异可以使用各种相似性度量函数，然而大多数相似性度量方法考虑了所有特征值的绝对值，样本间的相似性并不具备可解释性. 对于告警日志特征而言，特征值越小越好，因此待测样本点的某些特征值小于健康基线值时，表明该样本点在这些特征上没有出现异常，样本点的异常度映射需要对这些特征值的差异进行屏蔽.

线性修正函数 ReLU 是一种非线性映射函数[19]，该函数能够较好地模拟生物神经元的特性，并且计算简单，因此被广泛用于神经网络的激活函数. 本文利用该函数对样本点的异常度进行映射，对样本点中特征值小于健康基线值的特征进行屏蔽，并且通过加权处理提高重要特征对异常度的影响.

设 x_i 、x_i^c 分别表示待测样本和健康基线的第 i 个特征值，w_i 表示第 i 个特征的组合权重，则待测样本点 n 在工况特性 c 条件下的异常度 δ_n^c 的计算公式为

$$\delta_n^c=\sum_{i=1}^{k} w_i \cdot \mathrm{ReLU}\left(x_i'-x_i^c\right) \tag{3}$$

式(3)中 δ_n^c 的实际意义是表示设备在当前工况 c 条件下，告警日志样本点 n 中大于健康基线的部分特征值超出健康基线值的加权百分比之和，其值越大，表明设备健康状态越差.

2.4.3　异常判定阈值

假设在某一工况条件下，所有健康状态下的样本点的异常度服从某一概率分布，当观测数据足够多时，用样本点异常度的均值代替期望. 当待测样本点的异常度超过设备处于健康状态下样本点的异常度期望值时，表明待测样本点可能为异常样本点. 因此，异常判定阈值可以通过计算已有健康状态下的样本点异常度均值并加入经验参数进行确定，公式为

$$\delta_t^c=\frac{1}{N_c}\sum_{i=1}^{N_c}\delta_i^c-\gamma, i=1,2,\cdots,N_c \tag{4}$$

式中：δ_t^c 为异常判定的阈值，N_c 为在工况特性 c 下的正常样本点个数，δ_i^c 为在工况 c 条件下第 i 个健康状态样本点的异常度，γ 为经验参数. 当实时观测日志样本点的异常度超过阈值时，判定设备状态异常，此时通过人工对异常状态进行验证，并根据验证结果调整当前工况下的经验参数，并对阈值进行更新.

2.5　时间复杂度分析

对本文方法的时间复杂度进行分析. 设历史告警个数为 n，分析离线模型构建阶段，统计历史告警中每个类型的告警数量的时间复杂度 $O(n)$. 对样本点进行 DPC 聚类的时间复杂度为 $O(n^2)$，计算特征权重的时间复杂度 $O(n)$，判定阈值的确定需要计算每个样本点与类别中心的异常度，因此该阶段的时间复杂度为 $O(n^2)$.

分析实时异常检测阶段. 在日志子集划分阶段，设滑动时间窗口内的告警日志数量为 n_s. 统计当前时间窗口内每个类型告警数量的时间复杂度 $O(n_s)$. 确定设备当前运行工况，需要计算样本点与多个簇中心的距离并排序，这部分的时间复杂度与工况类别个数有关，在给定基准健康状态矩阵的情况下，可以认为其时间复杂度为 $O(1)$. 计算待测样本点的异常度，并判断当前样本点的异常度是否超过判定阈值. 该阶段的时间复杂度 $O(n_s)$ 与滑动时间窗口参数设置有关.

3　实验分析

为了验证本文方法的有效性，利用北京某电力通信公司网管系统采集的 SDH 设备近 1 个月的设备监控告警日志数据，构建离线模型，用已有的异常状态告警记录作为算法的测试集，验证本文方法对同一设备在不同运行状态下异常检测的有效性. 表 2 所示为部分原始告警日志数据的示例.

表 2　设备监控告警日志数据示例

Tab. 2　Example of equipment monitoring alarm log data

级别	名称	发生时刻	清除时刻
次要	LP_REI_VC12	19:21:20	19:21:23
提示	BEFFEC_EXC	19:25:30	19:33:00
紧急	CONNECT_FAIL	19:31:15	19:31:25
紧急	NE_NOT_LOGIN	19:31:15	19:31:29
紧急	COMMU_BREAK	19:31:15	19:31:25
重要	COMMU_SWITCH	19:31:23	19:31:54

实验验证了本文所设计的组合权重方法以及 W-ReLU 模型计算设备异常度的有效性，并对比了本文算法和其他异常检测算法的准确率和误警率. 准确率是算法识别正确的异常样本点占全部异常样本点的比例，误警率是将正常样本点误识别为异常样本点的数量占全部正常样本点的比例.

3.1　样本点 DPC 聚类分析

对健康状态下的告警日志按照固定时间长度的

窗口进行划分. 这里为了模拟多工况环境，每个样本点都从设备闲时和忙时两种工况条件下采集，设定时间窗口的大小为 2 h，经过数据预处理，共得到 144 个闲时工况样本点和 96 个忙时工况样本点.

按照式(5)对每个样本点的各个维度进行归一化处理，以消除不同量纲对聚类结果的影响.

$$x_i' = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}, i = 1,2,\cdots,k \tag{5}$$

对归一化后的样本点进行 DPC 聚类分析，结果如图 4 所示. 本实验采用高斯核函数定义某点的局部密度，并设定截断距离，使每个样本点的平均邻居个数为全体数据点总数的 2%.

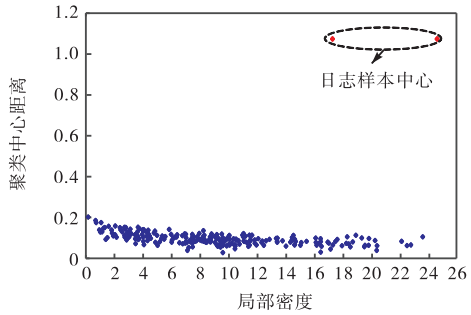


图 4　基于 DPC 算法的日志特征聚类分析

Fig. 4　Cluster analysis of log features based on DPC algorithm

DPC 聚类能够准确地将所有样本点划分为两类，右上角的两个点具有较高的局部密度和较大的聚类中心距离，因此代表两种工况条件下告警日志特征的健康基线. 设定比例参数 $\theta = 0.4$，分别计算两种工况条件下的组合权重(表 3).

表 3　两种工况条件下的健康基线及组合权重

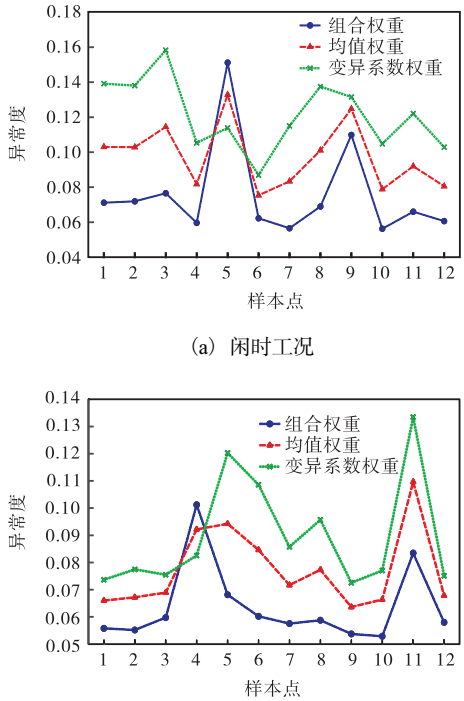
Tab. 3　Health baseline and feature weights under two working conditions

工况	特征	x_1	x_2	x_3	x_4	x_5
闲时	健康基线	0.21	0.18	0.24	0.23	0.19
	变异系数	0.11	0.18	0.06	0.09	0.12
	分布权重	0.18	0.11	0.33	0.22	0.16
忙时	健康基线	0.69	0.51	0.59	0.72	0.72
	变异系数	0.06	0.08	0.05	0.08	0.06
	分布权重	0.25	0.18	0.31	0.12	0.14

3.2　模型有效性验证

采集该设备在闲时和忙时两种工况条件下各 12 个样本点，每种工况条件下都包含两种异常情况，分别采用均值权重法、变异系数权重法、组合权重法进行异常度的计算，实验结果如图 5 所示.

由图 5 可知：在两种工况条件下，本文所提出的组合权重法能够较为明显地检测出异常样本点，即闲时工况条件下的样本点 5、样本点 9 和忙时工况条件下的样本点 4、样本点 11，表现出较为明显的异常度增大. 然而，采用变异系数权重法和均值权重法出现了未检出和误识别的情况，这是由于变异系数权重法对正样本中离散程度较大的特征赋予更大的权重，导致健康状态下某些样本点的异常度偏大. 在设定各个特征权重相同的情况下，设备异常度不能很好地反映重要日志特征的变化. 因此，本文采用的组合权重法能够削弱不重要特征的变化，提高重要特征对异常度的影响.



(a)　闲时工况

(b)　忙时工况

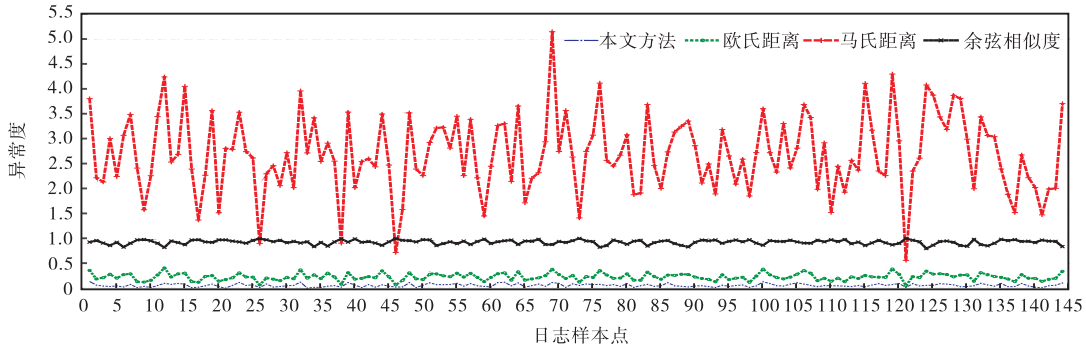
图 5　两种工况条件下不同特征权重对比

Fig. 5　Comparison of different feature weights under two working conditions

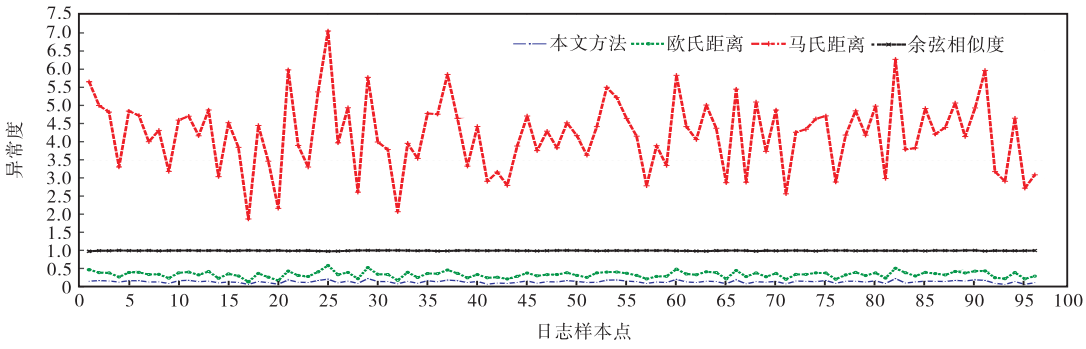
健康状态下的样本点应该具有较低的异常度和较为稳定的数据分布. 为了衡量不同相似性度量方法的稳定性，对两种工况条件下所有健康状态的样本点分别计算每个样本点与对应工况条件下健康基线的不同距离度量，实验结果如图 6 所示.

在两种工况条件下，对于健康状态下的样本点，马氏距离具有较大的波动性. 这是由于马氏距离虽然考虑了特征间的相互关系，但对于部分没有出现异常的特征，马氏距离受到这些特征的变化影响较大，因此这种相似性度量方式不稳定. 虽然余弦相似度具有更高的稳定性，但在忙时工况下，这种相似性度量

方式出现了失效的情况，表现出对日志特征的变化不敏感. 这是由于余弦相似度更加关注特征维度之间的差异，而并非数值的差异. 欧氏距离和本文方法虽然都更加关注特征值的差异，但欧氏距离的分布显然具有更大的方差. 这是由于欧氏距离同样也受到没有出现异常的特征变化的影响，若某一样本点的特征值都低于健康基线，很明显，该样本点是正常的，而该样本点与健康基线的欧氏距离可能较大，显然不符合实际的设备健康状态. 采用本文的相似性度量方式得到的异常度最小，符合正常样本点的类别特性，并且具有较高的稳定性，能够较好地衡量样本点所对应设备的当前状态.



(a) 闲时工况



(b) 忙时工况

图 6　两种工况条件下的不同距离度量方法对比

Fig. 6　Comparison of different distance measurement methods under two working conditions

3.3　与其他异常检测算法的对比

通过对比本文方法和支持向量数据描述(SVDD)[20]、局部离群因子[21]两种异常检测算法的准确率和误警率，评估本文所提方法的异常检测性能(表 4).

表 4　不同算法的异常检测性能对比

Tab. 4　Comparison of abnormal detection performance of different algorithms

异常检测方法	准确率/%	误警率/%
支持向量数据描述	68.2	1.2
局部离群因子	81.9	7.9
本文方法	95.4	2.1

两种对比算法的参数通过不断调优，以保证对异常数据点具有较高的准确率. 本文方法和支持向量数据描述算法用健康状态下的样本作为模型训练集，用包含故障注入实验检测数据在内的样本点作为测试集；而局部离群因子算法直接对全体样本进行异常检测.

与其他两种异常检测算法相比，本文对告警日志的异常检测具有最高的准确率和较低的误警率. 支持向量数据描述算法将不同工况条件下的所有样本首先进行核函数映射，在高维空间中划分出一个超球体来描述正常样本空间，这种方法一方面没有考虑告警重要性的影响，另一方面由于设备运行工况的动态变化，导致模型所划分出的超球体空间偏大. 局部离群因子算法对每个样本点计算离群因子，这种方法考虑数据的局部密度信息，在实际的设备运行过程中，其告警日志特征会随着设备实时运行工况而产生变化，当样本点的某些维度特征值变化较大时，也会导致异常检测的误警率上升.

4　结　语

本文针对光通信设备的智能运维问题，提出了基

于 W-ReLU 的设备多工况状态异常检测方法. 与已有的异常检测算法相比,该方法对基于健康状态下的监控告警日志数据进行离线模型构建. 在实时异常检测中,仅需计算生产环境下待测样本点的异常度是否超过阈值进行异常判定,避免了对全量样本点进行距离计算导致效率低下的问题,实现了在类别不平衡条件下的高效率异常检测. 并且随着告警数量的增加,该方法具有较好的迭代更新能力.

对基于监控日志的设备异常检测问题,本文根据滑动时间窗口内各类型的告警数量进行向量化表示. 时间窗口的大小以及不同的日志子集划分方法对异常检测模型的影响,以及告警特征与设备异常度的非线性映射关系,仍需要进一步研究.

参考文献:

[1] 蒲天骄,乔骥,韩笑,等. 人工智能技术在电力设备运维检修中的研究及应用[J]. 高电压技术,2020,46(2):369–383.

[2] 刘芳. 基于协同过滤的舰船通信网络告警数据多维度分析方法[J]. 舰船科学技术,2020,42(4):121–123.

[3] 贺艳芳,石坚. SDH 告警显示预处理和告警关联分析[J]. 科学技术与工程,2006(4):487–491.

[4] 齐小刚,胡秋秋,姚旭清,等. 一种有效的通信网络告警分析方法[J]. 西安电子科技大学学报,2019,46(4):1–8.

[5] 张光兰,杨秋辉,程雪梅,等. 序列模式挖掘在通信网络告警预测中的应用[J]. 计算机科学,2018,45(S2):535–538.

[6] 杨剑,蓝明超. 基于业务时间窗选取的告警聚类及关联方法[J]. 光通信研究,2019(1):33–36.

[7] 刘军,王颖,张冰,等. 电力骨干通信网中告警预测方法研究[J]. 光通信研究,2019(5):9–13.

[8] 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]//IEEE. 2020 International Conference on Computing,Networking and Communications (ICNC). New York:IEEE,2020:294–298.

[9] 杨济海,刘洋,刘杰,等. 基于并行的 F-LSTM 模型及其在电力通信设备故障预测中的应用[J]. 武汉大学学报(理学版),2019,65(3):263–268.

[10] 王峰,李兴华,李晓龙,等. 基于健康画像的光通信设备故障预测算法[J]. 光通信技术,2021,45(9):31–35.

[11] 张颖君,刘尚奇,杨牧,等. 基于日志的异常检测技术综述[J]. 网络与信息安全学报,2020,6(6):1–12.

[12] 邵俊健,王士同. 高维数据的增量式聚类算法的距离度量选择研究[J]. 计算机工程与科学,2019,41(2):214–223.

[13] JOVE E,CASTELEIRO-ROCA J,QUINTIAN H,et al. A new method for anomaly detection based on non-convex boundaries with random two-dimensional projections[J]. Information fusion,2021,65:50–57.

[14] 陈叶旺,申莲莲,钟才明,等. 密度峰值聚类算法综述[J]. 计算机研究与发展,2020,57(2):378–394.

[15] 徐冰珂,周宇喆,杨茂林,等. 面向电信行业网络告警系统的告警过滤算法[J]. 计算机应用,2018,38(10):2881–2885.

[16] 冯士龙,台宪青,马治杰. 改进的基于日志聚类的异常检测方法[J]. 计算机工程与设计,2020,41(4):1087–1092.

[17] 高鑫磊. 民用飞机发动机起动系统健康监测与故障诊断方法研究[D]. 南京:南京航空航天大学,2019.

[18] 白丽丽,白尚旺,党伟超,等. 基于离差最大化组合赋权的煤矿安全评价研究[J]. 计算机应用与软件,2021,38(4):82–87.

[19] 曲之琳,胡晓飞. 基于改进激活函数的卷积神经网络研究[J]. 计算机技术与发展,2017,27(12):77–80.

[20] 高罗莹,田增山,李玲霞,等. 一种基于 SVDD 的 WLAN 室内被动入侵检测方法[J]. 重庆邮电大学学报(自然科学版),2020,32(2):200–209.

[21] 秦辉东,杨加,李笑难,等. 基于多特征的 DNS 异常检测技术研究[J]. 深圳大学学报(理工版),2020,37(S1):36–43.

责任编辑: 郎婧