引用本文：曾创展，蔡晨贾农，姚耀，等. 基于区块链的白盒化光网络故障检测[J]. 光通信技术，2024，48(2)：72-77.

基于区块链的白盒化光网络故障检测

曾创展[1]，蔡晨贾农[1]，姚 耀[1]，刘 伟[2]，杨 辉[2*]

（1.61932部队，北京 100000；2.北京邮电大学 信息光子学与光通信国家重点实验室，北京 100876）

摘要：针对白盒化通信网设备异常检测问题，提出了基于区块链的白盒化光网络故障检测方案。首先，设计了面向数据收集的区块链架构和基于区块链的数据收集与存储方案；然后，又设计了基于自编码器的数据异常检测算法；最后，将自编码器模型部署至区块链共识设备中，形成完整的基于区块链的异常检测机制。实验结果表明：在不同设备的采集数据下，所提方案可以快速实现数据上链存储，且能准确、快速地识别设备的异常数据，实现网络设备的实时异常检测。

关键词：白盒化光网络；异常检测；区块链；自编码器

中图分类号：TN256　**文献标志码**：A　**文章编号**：1002-5561(2024)02-0072-06

DOI：10.13921/j.cnki.issn1002-5561.2024.02.013　**开放科学(资源服务)标识码(OSID)**：

White box optical network fault detection based on blockchain

ZENG Chuangzhan[1], CAICHEN Jianong[1], YAO Yao[1], LIU Wei[2], YANG Hui[2*]

(1. Unit 61932, Beijing 100000, China; 2. State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Aiming at the problem of anomaly detection of white-box communication network equipment, a white-box optical network fault detection scheme based on blockchain is proposed. Firstly, a data collection oriented blockchain architecture and a data collection and storage scheme based on blockchain are designed. Then, a data anomaly detection algorithm based on autoencoder is designed. Finally, the autoencoder model is deployed to the blockchain consensus device to form a complete blockchain-based anomaly detection mechanism. The experimental results show that the proposed scheme can quickly realize the on-link storage of data collected by different devices, accurately and quickly identify the abnormal data of devices, and realize the real-time anomaly detection of network devices.

Key words: white box optical network, anomaly detection, blockchain, autoencoder

0 引言

随着软件定义网络（SDN）和第五代移动通信(5G)技术的广泛应用，运营商在部署5G基站设备时，面临的资金压力越来越大[1]。目前，设备供应商可以提供成套绑定的硬件服务器和专用软件，但运营商只能从单一设备制造商那里购买并部署硬件设备，这给运营商造成了巨大的成本压力，同时也阻碍了设备的更新升级。为了打破被设备供应商束缚的封闭局面，国内外通信运营商与IT公司纷纷开始研究网络白盒化[2]。然而，网络白盒化也面临一些问题。例如，一个基站可能包含来自多个制造商的设备，它们会具有不同的内部设计和技术规格[3]。相较于黑盒化网络，白盒化网络中的设备及其零部件是由多个厂商提供的，在异常检测中将面临责任推卸的问题，甚至存在篡改性能数据的风险。因此，如何检测和准确定位设备异常，以及保障设备数据的安全性与数据的可验证性是白盒化网络异常检测的难点。

区块链是一种分布式数据库，具有分布式去中心化、无需中央信任机构、数据一致、不可篡改、可加密、可验证、可追溯等优点[4]。它使用实用拜占庭容错(PBFT)算法来实现共识机制，并基于密码学来增强数据的安全性和可验证性。将区块链技术引入光网络和运营商

收稿日期：2023-05-01。

基金项目：国家自然科学基金面上项目(62271075)资助。

作者简介：曾创展(1990—)，男，博士，工程师，毕业于航天工程大学，主要研究方向为数据科学、人工智能。

*** 通信作者**：杨辉(1987—)，男，博士，教授，主要研究方向为意图驱动光网络、多波段光网络、区块链光网络。

设备的连接中,以构建可信分布式光网络,增强了白盒化光网络下不同厂商设备的信任度,让故障责任归属更明确[5]。因此,区块链技术无需任何第三方中心的介入就可以使参与者达成共识,以极低的成本解决了不同设备供应商之间信任与价值的可靠传递难题,同时区块链技术可以保障网络数据的有效性与准确性,提供准确的故障划分依据。

本文提出一种基于区块链的白盒化光网络故障检测方案,以优化在检测大型复杂厂商设备时的检测和存储机制,实现白盒化网络异常的自动检测和智能诊断。

1 基于区块链的白盒化光网络故障检测方案

本文首先设计基于区块链的数据收集与存储方案;然后,针对数据异常识别问题,又设计了基于自编码器的异常检测算法;最后,将自编码器与区块链结合形成完整的基于区块链的异常检测机制。

1.1 基于区块链的数据收集与存储方案

基于区块链的白盒化光网络故障检测的网络架构如图 1 所示。图中左边是由不同厂商的设备组成的白盒化网络场景。该场景下,运营商会从许多通信公司购买 5G 统一标准的设备与元器件来构建 5G 网络,不同厂家的设备可以通过开放的接口进行互联。它们之间通过信息交互进行写作,因而会产生大量的数据流。为了让所有设备更好地协同工作,本文使用区块链进行数据收集与异常检测。SDN 控制器通过基于 YANG/NETCONF 的南行接口来处理节点和设备代理。管理平面嵌入区块链分类帐,向控制器暴露接口,并且配备有能够从选定的数据平面接收消息的本地 NETCONF 客户端。同时,将自编码器模型嵌入到区块链的智能合约中,转换成可以自动执行的代码。智能合约会不断重复执行,检测白盒化光网络的设备数据信息,标注异常设备数据。

但是,由于设备产生的数据量太大,该网络架构无法在单一链上实现共识和存储,所以本文采用了链上共识、链下存储的方式实时检测和存储设备信息。首先,需要选出合适的设备组成一个区块链共识组。为了保障共识消息的传播效率,本文设计了一种基于时间戳的共识组选举算法。选举算法流程如下:所有白盒化设备向其它设备发送一个选举请求,请求包含时间戳 t_1 和自身信息,时间戳 t_1 是各自设备发送请求的时间。当设备在接收到选举请求时,定会将收到所有请求的时间转换为时间戳 t_2,并计算出接受信息与发送信息的时间差 $t_3=t_2-t_1$,然后将请求设备信息与 t_3 记录在表单中。在收到所有选举设备的请求后,每个设备会广播各自记录的表单。最后,根据表单中对应设备的时间差 t_3 的总和,选择时间差最小的 n 个设备组件成为区块链共识组。

区块链共识组通过完整的网络结构实时监控运营商设备网络中具有动态活动特征的节点信息,并为数据平面提供数据支撑。在数据平面上,由于区块大小和费用的限制,不能直接在以太坊、联盟链等公共区块链上存储解释[6]。而星际文件系统(IPFS)中的链外分布式点对点存储是一种可行的解决方案[7]。因此,本文将区块链与 IPFS 相结合,设计了基于区块链的 IPFS(BlockIPFS),为区块链提供数据支撑。

BlockIPFS 是一种可信的、易于使用的、高性能的分布式文件系统。在BlockIPFS 中,区块链负责管理文件系统的元数据(元数据是原始数据的存储信息),区块链账本负责存储与事务相关的元数据。由于元数据

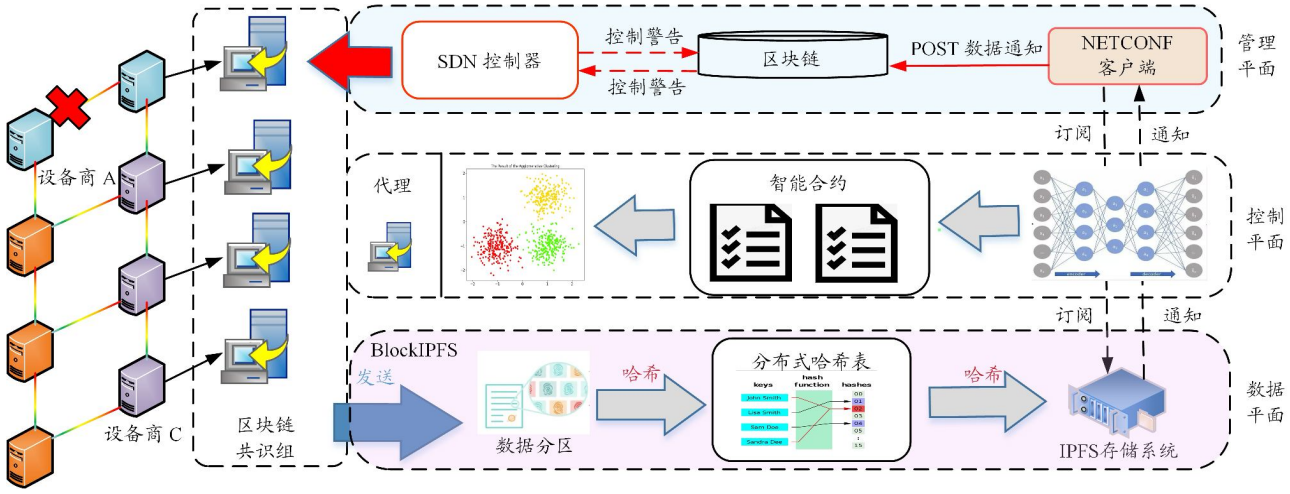

图 1 基于区块链的白盒化光网络故障检测的网络架构

的数量比原始数据小很多（比如原始数据为视频、图片、文本等,而元数据仅为原始数据的信息与索引),故不会产生高的存储量或网络开销。原始数据被存储在 IPFS 中，区块链仅保存数据的哈希值与相关信息。一旦文件成功添加到 IFPS,其多哈希与文件信息经过共识被添加到区块链中,区块链中的所有设备都能访问该哈希。为了解决存储中的哈希冲突,区块链还提供了一种额外的方法来共享内容的多哈希值,即将某一层冲突的哈希值交由下一层进行填充。IPFS 操作和它们在区块链中的相关元数据存储是同步的。为了确保操作的连续性,当区块链存储操作失败时,设备会收到失败的警告,并有选择是否进行区块链操作或中止整个进程的权限。

本文设计的基于区块链的数据收集与存储方案的流程如下:1)设备对网络数据进行收集,将数据信息发送至区块链共识组进行存储前的共识;2)经过区块链共识组公示后，设备将数据发送至 IPFS 存储系统;3)IPFS 对数据进行存储,并生成数据的哈希值;4)IPFS 将储存结果(包括数据信息、数据哈希等)发送给区块链共识组进行共识,若共识成功,则将 IPFS 存储结果发布至区块链分类账。区块链的所有设备均可通过访问区块链分类账与 IPFS 系统获取原始数据,且设备每次访问原始数据时,区块链分类账都会对访问进行记录。

1.2 基于自编码器的数据异常检测算法

由于白盒化光网络的部署规模越来越大,导致数据快速增长,使得传统有监督的深度学习模型难以应对。而自编码器是一种无监督学习的方法,不需要标注训练数据就可以自动从数据中学习到有用的特征表示,并从未标记的数据中发现异常。此外,白盒化光网络产生的数据具有复杂多维的特征,不同指标的特征权重可能有较大差异,传统人工设定权重的方式已不在适用。而自编码器通过学习数据的压缩表示来提取数据重要特征,并通过最小化重构误差来生成输入数据的重建。在二维图像中,这种重建过程可以清晰地展示出异常数据。同时,自编码器能够适应不同类型和分布的数据,并具有一定的鲁棒性。即使在一定程度的噪声或异常数据的情况下,自编码器也能够通过学习数据的分布特征来恢复原始数据,从而提高异常检测的鲁棒性。

自编码器的目标是学习数据的低维表示,这意味着它可能无法完全捕捉数据中的所有复杂性和细节。虽然这有助于降低数据的维度,但也可能导致信息损失。白盒化光网络中的数据特征较为冗余,并且数据一致重复性较高,这使得自编码器能够很好地适应白盒化光网络的场景。但是,自编码器的训练过程可能相对复杂,因此需要选取容量合适的样本进行训练。

在基于机器学习或数据挖掘的异常检测中,本文获取在训练期间能捕获正常行为的模型,然后检查测试数据是否能与训练模型拟合[8]。如果测试数据与训练的模型不一致,则将其视为异常。降维异常检测是基于检测数据具有相互关联的特征,降维后的特征可以嵌入到一个低维的子空间中。在这个子空间中,正态样本和异常样本呈现出明显的区别。在训练阶段,有正规数据作为训练集$\{x(1),x(2),\cdots,x(m)\}$。假设每个数据样本 $x(i)\in R^D$, R^D 由 D 个不同变量的向量表示。将数据压缩到低维重构子空间,并重现输出$\{\hat{x}(1),\hat{x}(2),\cdots,\hat{x}(m)\}$,使重构误差变小,$m$ 为输入向量的个数。重构误差 $E(i)$可表示为

$$E(i)=\sqrt{\sum_{j=1}^{R}\left(x_j(i)-\hat{x}_j(i)\right)} \tag{1}$$

其中,R 是降维后的指标个数,$x_j(i)$表示原始数据第 i 条数据的 j 个指标参数,$\hat{x}_j(i)$表示解码输出数据第 i 条数据的 j 个指标参数。

确定子空间后，在测试阶段将测试数据投影到子空间中,重构原始数据。本文使用式(1)中的重构误差作为判断数据异常的分数。当测试样本为满足测试阶段学习到的正态相关的正常实例时，其重构误差值较低;当样本异常时,其重构误差值较大。由于网络数据维度较多，所以本文使用自动编码器作为实现这一过程的技术手段[9-10]。图 2 为自动编码器结构示意图。

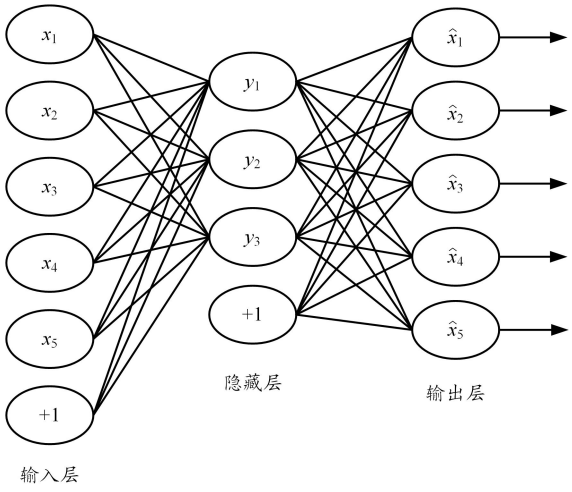

图 2 自动编码器结构示意图

首先,自动编码器是一个无监督神经网络,其目标是学习复制输入向量$\{x(1),x(2),\cdots,x(m)\}$,然后输出$\{\hat{x}(1),\hat{x}(2),\cdots,\hat{x}(m)\}$。在隐藏层中,输入被压缩成少量的神经元。$l$层单元$i$的激活如式(2)所示。

$$a_i^l=f\left(\sum_{j=1}^{n} W_{ij}^{l-1} a_j^{l-1}+b_i^l\right) \tag{2}$$

其中,a_i^l表示l层单元i的输出向量,W和b分别表示权重和偏置参数,f表示每一层对应的激活函数。输入层中,$a^{(1)}=x$;输出层中,$a^{(3)}=\hat{x}$。本文在隐藏层中使用了sigmoid函数,由于没有将每个输入示例预先缩放到一个特定的区间(如$[-1,1]$),因此在输出层中使用了线性函数。在训练期间,最小化式(3)中的关于W和b的目标函数。目标函数包含正则化项,参数λ决定正则化的强度。

$$J(W,b)=\frac{1}{m}\sum_{j=1}^{m}\left(\frac{1}{2}\left|\left|x(i)-\hat{x}(i)\right|\right|^2\right)+\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}\left(W_{ji}^l\right)^2 \tag{3}$$

其中,n_l为网络中的层数,s_l为层中单元数。为了避免高压缩编码中通常存在的高纠缠度,本文使用一小部分神经元对输入进行编码,并通过增加隐藏单元数量和向输入中添加噪声来实现。为了训练更高精度的编码器,本文通过选择输入的固定数量的分量为0,来破坏输入,以训练更高精度的编码器。

1.3 基于区块链与自编码器的异常检测策略

在机器学习完成部署前,区块链共识组负责收集机器学习模型训练的数据。当区块链收集到足够多的数据后,区块链中的网络主控制器(或者选择网络中算力较强的设备)会通过区块链获取这些数据。主控制器通过区块链分类账得到所有数据的存储信息,再基于这些信息从IPFS存储系统中下载全部数据,并采用基于自编码器的数据异常检测算法完成机器学习模型的训练。最后,将训练好的模型部署到区块链共识组设备中。

在光学异常检测阶段,本文使用正常数据进行训练重构。由于测试数据集中的正常数据与训练阶段构建的正常轮廓一致,重构误差较小,而异常数据的重构误差相对较高。因此,可以通过计算重构误差来检测异常数据。本文使用主成分分析(PCA)对数据进行降维至2维,使结果以集群的方式显现。由于重构空间更能抵抗数据噪声,因此重构空间的距离提供了足够的细节来检测并区分原始空间中的异常数据。

白盒化光网络数据异常检测流程如图3所示。设备会将产生的所有数据存储到链下IPFS中,然后IPFS会将哈希值和存储地址以区块形式上传到区块链,由区块链共识组进行共识验证。同时,异常检测合约根据地址获取所有数据,并计算重构误差。当大型设备中的某个设备出现问题时,共识组会通过自编码器检验设备数据以识别异常。共识组将问题数据保存在区块链中,方便故障定位和后期维护。然后,共识组的代理会创建一个通知代理程序,该程序在检测到指定的异常后上报相应的异常告警。所有订阅实体都会发送通知。最后,控制器对异常数据进行计算处理后,根据分析结果定位具体设备,并对设备进行维护。通过区块链与自编码器的有机结合,实现了异常检测从数据收集到异常识别全流程的安全可控。

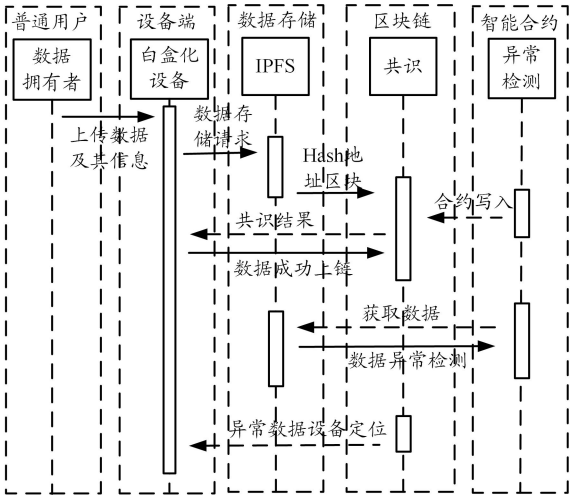

图3 白盒化光网络数据异常检测流程

2 仿真结果

本文对白盒化光网络中基于区块链的异常检测的性能进行验证。仿真平台采用多核服务器,配备16个物理2.90GHz CPU内核和80GB RAM,运行Linux 4.15操作系统。

首先,本文基于区块链搭建数据收集平台,其网络拓扑图如图4所示。此拓扑按照COST239设置,较为典型与通用。为了模拟机器学习的环境,本文在拓扑中设置了3个算力较强的节点,其它节点按照普通网络节点配置。基于Fabric搭建了区块链网络平台(BlockFD),负责数据收集与网络共识。在数据收集模拟中,使用中兴试验网的数据来模拟真实网络,数据共有29 556条,其中正常数据为23 509条,异常数据

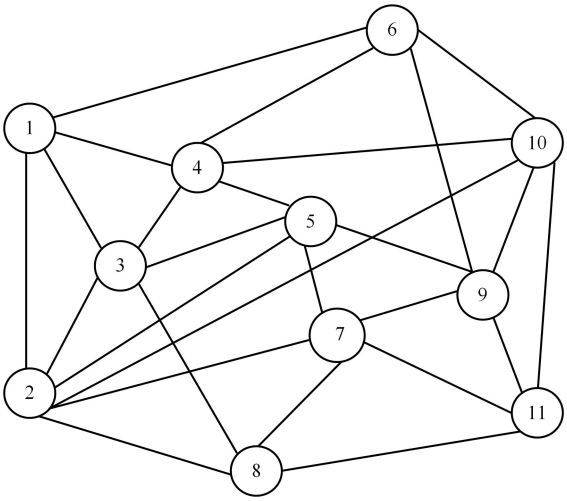

图 4 网络拓扑图

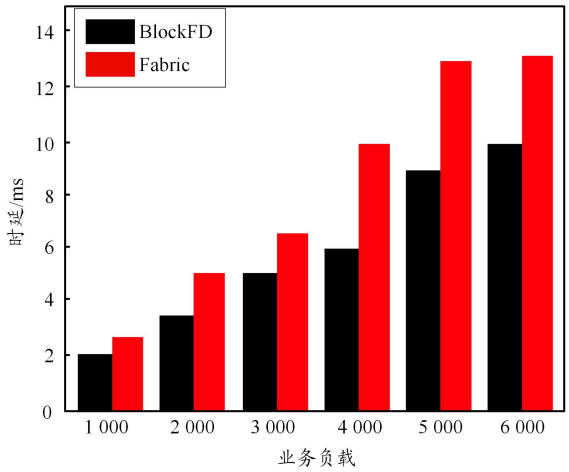

图 6 时延性能对比

为 6 047 条。数据以天为单位,主要由 16 个特征组成,包括不可用时间、环境温度、激光器偏置电流、激光器温度偏移、输入/输出光功率以及最大值/最小值平均值等。此外,自编码器本质上是一种对数据压缩降维的方法,对数据特征有较强的包容性,并在数据泛化方面表现相对灵活。每个网络节点负责上传试验网单个设备的数据,该数据集包括时延、抖动、丢包等网络性能参数。

当业务负载为 1 000~6 000 时,BlockFD 和传统 Fabric 的吞吐量和时延性能对比情况分别如图 5 和图 6 所示。由图 5 可知,当流量负载为 1 000 时,BlockFD的吞吐量比 Fabric 的吞吐量高 80 TPS。随着流量负载的增加,BlockFD 的吞吐量也逐渐增加。当流量负载为 6 000 时,BlockFD 的吞吐量达到 253.6 TPS,因此 BlockFD 比传统 Fabric 区块链的吞吐量更高。由图 6 可知,BlockFD 的时延均优于传统 Fabric,特别在流量负载量较大时,

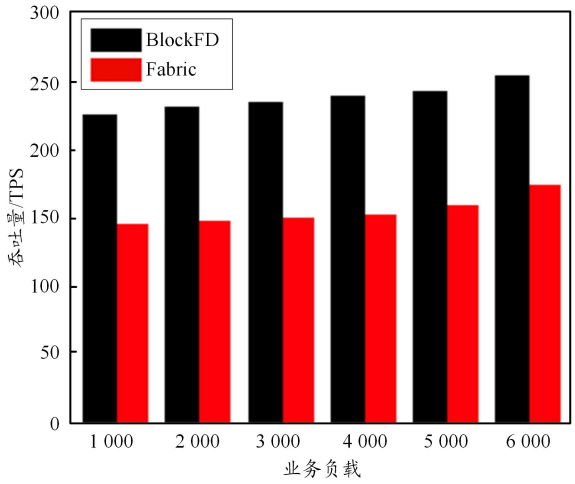

图 5 吞吐量性能对比

时延性能会较为明显。这是因为 BlockFD 通过选举算力强和相对时延较低的节点组成共识组,从而加速了查询和认证的过程。但两者的时延目前仍然较高,未来还需进一步优化改进。整体仿真结果表明,BlockFD 在吞吐量和时延方面具有较好的性能。

本文基于区块链收集的数据对所提检测方案进行了神经网络的训练,并测试其性能,如图 7 所示。可以看出,自编码器的神经网络在训练轮次为 20 时,其识别准确度已经达到了 70%以上;而且在训练轮次为 30 时,其神经网络识别准确度已经突破了 95%,并且其精度已经趋于稳定。由此可见,采用基于自编码器的数据异常检测算法的模型收敛速度快、收敛程度好,可以实现较高精度的异常检测。基于自编码器的数据异常检测的可视化结果如图 8 所示。图中数据是多维数据经过一系列矩阵分析后提取的主要的 2 个特征成分。可以看出,正常数据聚合程度较高,异常数据离群明显。因此,本文提出的检测方案可以有效地检测离群数据点。

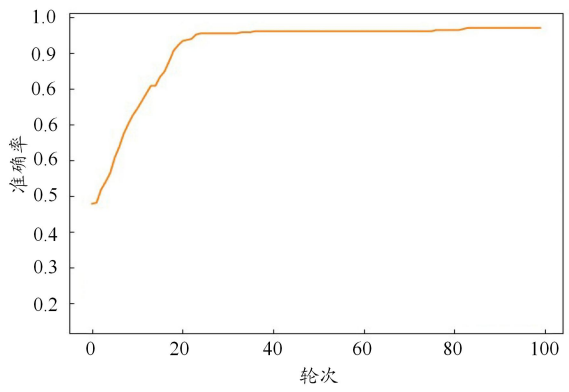

图 7 自编码器训练准确率

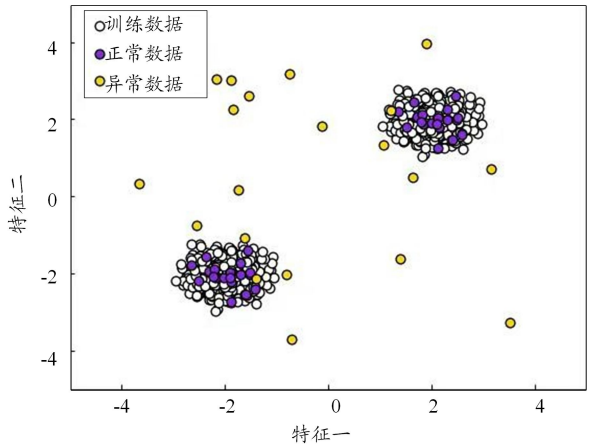

图 8 基于自编码器的数据异常检测的可视化结果

3 结束语

针对白盒化网络的设备异常检测问题,本文提出了一种基于区块链的白盒化光网络故障检测方案。该方案利用区块链技术实现设备数据的收集,并基于收集到的数据训练自编码器用于设备异常检测,将训练好的自编码器模型部署到区块链共识组,区块链共识组使用自编码器模型与共识机制,实现对设备数据的实时分析和异常检测。仿真结果表明:在不同网络设备的检测数据下,所提出的检测方案均可以实现快速与准确的数据收集,并能精准地识别网络中的异常数据,实现对白盒化网络的实时异常检测。在未来的工作中,还需要再结合现阶段流行的聚类算法对网络数据进行更深入的异常分析。

参考文献:

[1] 张梅琼. 基于云网络的 SDN 运维节能系统 [J]. 软件,2021,42(12):174–176.
[2] 荆继武, 李畅. 密码技术的现状与白盒化发展趋势 [J]. 中国信息安全,2021(8):49–53.
[3] 王文兵,张志民. O-RAN 5G 白盒化小基站硬件架构的研究[J]. 通信技术,2020,53(11):2847–2854.
[4] 邵奇峰,金澈清,张召,等. 区块链技术:架构及进展[J]. 计算机学报,2018,41(5):969–988.
[5] 梁永燊. 基于区块链的可信光网络高效控制技术研究[D]. 北京:北京邮电大学,2020.
[6] 游聪,王利朋,潘常春,等. 基于区块链的数据安全管理系统[J]. 计算机时代,2022,358(4):34–37,42.
[7] 高洁. IPFS+ 区块链技术在科学数据共享中的运用分析[J]. 图书情报导刊,2023,8(1):35–39.
[8] 蒋建春,马恒太,任党恩,等. 网络安全入侵检测:研究综述[J]. 软件学报,2000(11):1460–1466.
[9] 张国梁,郭晓军. 基于自编码器的网络异常检测研究综述[J]. 信息安全学报,2023,8(2):81–94.
[10] 蔚森. 基于自编码器的异常检测算法研究[D]. 南京:南京邮电大学,2022.