面向业务流程动态调整的异常检测与修复方法

刘芙洁, 方贤文

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.241200037

面向业务流程动态调整的异常检测与修复方法

刘芙洁　方贤文
安徽理工大学数学与大数据学院　安徽 淮南 232001
(1980951805@qq.com)

摘　要　在数字化转型浪潮中,业务流程的异常检测与修复对保障企业运营效率和决策质量至关重要,同时也对其检测与修复技术提出了更高的要求。传统的异常检测方法已无法满足当前业务流程实时监控和适应性调整的需要,而现有方法多侧重于静态分析,没有充分考虑业务环境的复杂性与多变性,难以适应流程动态变化的需求。基于此,创新性地提出了一种自适应异常处理方法(Adaptive Anomaly Handling Method,AAHM)。该方法通过动态参数调整和实时数据反馈,提高异常检测的准确性和修复的有效性。为验证该方法的有效性,实验采用了4组真实事件日志进行仿真。结果表明,该方法通过特征向量补全和行为修复策略,能够有效识别并对异常行为进行修复,恢复业务流程的正常执行。此外,通过对实验结果进行事后检验分析,进一步验证了所提方法的有效性和合理性。

关键词:过程挖掘;异常检测;自适应方法;异常修复;事后检验
中图分类号　TP391

Anomaly Detection and Repair Methods for Dynamic Adjustment of Business Process

LIU Fujie and FANG Xianwen
College of Mathematics and Big Data,Anhui University of Science and Technology,Huainan,Anhui 232001,China

Abstract　In the wave of digital transformation,the anomaly detection and repair of business processes are crucial for ensuring the operational efficiency and decision-making quality of enterprises. Meanwhile,higher requirements are put forward for its detection and repair technologies. Traditional anomaly detection methods can no longer meet the needs of real-time monitoring and adaptive adjustment of current business processes. Most of the existing methods focus on static analysis and do not fully consider the complexity and variability of the business environment,so it is difficult to adapt to the needs of dynamic changes in processes. Based on this,this paper innovatively proposes the AAHM. This method improves the accuracy of anomaly detection and the effectiveness of repair through dynamic parameter adjustment and real-time data feedback. To verify the effectiveness of this method,four groups of real event logs are used for simulation in the experiment. The results show that this method can effectively identify and repair abnormal behaviors and restore the normal execution of business processes through feature vector completion and behavior repair strategies. In addition,through post hoc test analysis of the experimental results,the effectiveness and rationality of the proposed method are further verified.

Keywords　Process mining,Anomaly detection,Adaptive method,Abnormal repair,Post hoc test

1　引言

在当今数字化时代,随着业务流程的日趋复杂化,组织机构越来越依赖于先进的自动化技术,以提升业务流程的效率和透明度,同时确保决策的科学性和精准性。在这一背景下,过程挖掘(Process Mining,PM)研究从事件日志出发,辅助以智能分析技术,如数据分析、流程建模和流程分析等,提取事件日志中的知识,用于发现、监控和改进真实流程执行。通常,过程挖掘技术可以分为3类:1)过程发现,通过可视化过程模型呈现事件日志;2)一致性检查,比较日志和模型以检测偏差;3)改进,利用事件日志中的历史数据来完善流程模型。

然而,物理世界中的业务流程数据往往存在不确定性及异常特征。这些异常可能是由操作失误、系统故障、欺诈行为或其他非预期因素造成的。异常行为的及时检测与修复对于保障业务流程的顺畅执行和结果的准确性至关重要。例如,在金融服务行业中,异常交易行为的快速识别不仅能预防经济损失,还能避免监管风险;在医疗领域,对异常治疗流程的实时监控和调整则直接关系到患者的健康和安全。它们不仅

到稿日期:2024-12-04　返修日期:2025-03-18
基金项目:国家自然科学基金(61572035);安徽省重点研发计划(2022a05020005);安徽省自然科学基金(水科学联合基金,2308085US11)
This work was supported by the National Natural Science Foundation of China(61572035),Key Research and Development Program of Anhui Province(2022a05020005) and Natural Science Foundation of Anhui Province(Joint Fund for Water Science,2308085US11).
通信作者:方贤文(280060673@qq.com)

影响了业务流程的正常运行,更有可能误导基于数据分析的决策制定,从而导致严重的后果。尽管当前过程挖掘领域在事件日志的完整性修复和异常检测方面的研究热度持续上升,但当前相关研究大多侧重于静态分析,缺乏对业务流程动态变化的适应和响应能力,难以满足快速变化的业务需求。

为了解决这一问题,本文提出了一种自适应异常检测及修复方法,称为自适应异常处理方法(Adaptive Anomaly Handling Method,AAHM)。该方法的创新之处在于,其能够根据实时数据动态调整参数,以灵活应对数据中的异常行为。通过融合机器学习技术,算法能够学习和识别正常与异常行为模式,从而实现对潜在异常行为的精准预测和及时响应,增强对业务流程变化的敏感度;通过实时数据流式处理架构,持续更新模型参数,并融合不同场景下的知识图谱丰富特征信息,强化对复杂异常行为的判别与处置能力,使其能够适应不同领域和场景下的应用需求。

本文的创新性主要体现在以下两点:

1)融合了编码策略,并创新性地使用行为特征图构建与分析以及特征向量补全方法,能够适应不同模型架构,且有效捕捉、校正异常行为并保障数据完整性,为业务流程异常处理提供了拓展思路;

2)在模型性能优化方面,基于自适应共振理论对实时数据流中异常的高效检测和模型进行动态调整,结合人工注入异常增强泛化能力,有效应对业务流程动态变化,显著提升了异常检测与修复的效率和准确性。

本文其余部分组织如下:第 2 章讨论了相关工作;第 3 章介绍了预备知识;第 4 章通过动机案例给出 AAHM 方法的实施步骤;第 5 章采用真实的事件日志进行实验仿真,并对实验结果进行对比分析,以验证所提方法的有效性;最后总结全文并展望未来工作。

2 相关工作

在业务流程管理领域,异常检测是确保流程效率和质量的关键环节。随着技术的发展,多种方法被提出以识别和处理流程中的异常情况。本章将对现有文献进行分类整理,以展现当前研究的主要趋势和方法。

2.1 异常检测方法

异常检测方法的研究主要集中在如何有效地从事件日志中识别出不符合预期模式的行为。文献[1]提出了一种结合流程挖掘、模糊多属性决策和模糊关联规则学习的方法,以检测业务流程中的异常情况。该方法首先使用流程挖掘技术检查事件日志和流程模型之间的一致性,然后应用模糊多属性决策来计算异常率,最后通过模糊关联规则学习生成关联规则,用于异常迹的检测。文献[2]介绍了一种基于事件流的在线流程挖掘器,其核心在于使用概率和非确定性自动机的集合,这些自动机会随着事件流的发展动态更新,从而实现实时的异常检测。文献[3]关注于错误事件的分类和数据质量的监控,通过持续训练新传入的分类模型,以支持在线流程挖掘中的异常检测。文献[4]提出了一种基于多图的异常检测框架 GAMA,该框架利用全局图将迹转换为多图,并使用图神经网络(GNN)学习多图嵌入,通过注意力机制聚合多个图形来捕获不同属性之间的内在关系。文献[5]提出了 BINet,这是一种实时多视角业务流程异常检测的神经网络架构,通过预测迹中的下一个事件并将这些事件与事件日志中的实际事件进行比较来识别异常。文献[6]出了一种用于分类数据的在线异常检测算法,该算法能够实时处理数据流并检测异常。文献[7]提出一种扩展似然图以包含事件属性的方法,该方法结合上下文,并考虑到并非每个意外事件都会被立即检测为异常,而是基于一定的发生可能性,从而减少误报数量。文献[8-11]使用自编码器检测和分析业务流程执行中发生的异常。文献[12]将编码的表征能力与元学习策略相结合,以增强对事件日志中异常迹的检测能力。文献[13]提出了异常检测启发式方法,该方法将从事件日志中挖掘的关联规则用于区分正常和异常迹。文献[14]利用特定上下文中活动发生的频率,检测给定特定事件数据中的异常行为。文献[15]利用序列挖掘技术在流程挖掘领域进行异常值检测。文献[16]提出了一种名为 LOGLG 的新型弱监督日志异常检测框架,以探索序列中关键字之间的语义联系。文献[17]通过引入一种自动化方法,在业务流程执行中进行多视角一致性检查和异常检测,从事件日志中提取预期和意外行为,并识别偏差模式。文献[18]提出了一种利用机器学习模型预测下一阶段活动的概率,并将不可预测的事件视为异常,进行在线事件异常检测的方法。

2.2 数据修复技术

数据修复技术的研究关注于如何纠正和优化事件日志中的错误或异常数据。文献[19]提出了一种基于数据视角修复过程模型的技术,该技术能同时保持控制流结构的完整性,这是底层数据 Petri 网的无限状态空间的有限符号抽象。文献[14]提出一种事件数据修复方法,该方法基于迹中的频繁上下文检测异常值,并随后修复此行为,而不是删除异常值的迹。文献[20]提出了一种基于日志自动机恢复错误时间戳的方法,该方法基于正确的频繁行为修复迹中事件的总顺序。概念漂移处理的研究关注于如何在业务流程随时间变化时,持续有效地进行异常检测。文献[21]将概念漂移检测、定位和过程模型修复集成在一起,提出了一种在概念漂移下自动修复过程模型的方法(AIMED)。

随着技术的不断进步,研究者提出了多种创新的方法来提高异常检测的准确性和效率。这些方法不仅包括传统的流程挖掘技术,还涵盖了机器学习、深度学习、自编码器和多图嵌入等先进技术。通过这些研究,可以看到业务流程管理领域正朝着更加智能化和自动化的方向发展。

传统的结合流程挖掘、模糊多属性决策和模糊关联规则学习的方法在执行原理上需依次执行多个相对独立的步骤,包括流程挖掘与模型一致性检查、模糊多属性决策计算异常率,以及模糊关联规则学习检测异常迹,各环节之间缺乏强关联性联系,使得整体流程繁琐复杂。而本文方法将事件日志转换为行为特征图,借助编码策略输入 ART 模型训练学习,可以深度融合数据特征与模型训练流程,从而利用全局数据模式精准识别异常,其技术原理更为简洁高效。此外,传统方法因多步骤且各步骤需分别处理复杂信息,在面对大规模业务流程数据时,计算量随数据量呈指数级增长,而

AAHM 方法依靠其编码方式和 ART 模型的快速学习及反馈调整机制，时间复杂度增长相对缓慢，在处理海量业务流程数据以及应对复杂动态变化时，在算法复杂度方面展现出显著优势，能更高效地利用计算资源，实现准确的异常检测与修复。

3 预备知识

定义 1(事件，迹，事件日志[22]) 业务系统中任一事件 e 是其对应活动 $a\in A$ 在某时刻的执行步骤，e 的执行携带时间戳、资源、活动名称等相关信息，形式化记为 $e=\{c,a,r,t,attr_{\{1,2,\cdots,n\}}\}$。其中 c 为事件所属流程实例 ID；a 为事件对应的执行活动标签；r 是执行活动时涉及的资源；t 是事件的时间戳；$attr_{\{1,2,\cdots,n\}}$ 是日志中所具有的附加属性值，对于任意的 $1\leqslant i\leqslant n$，$attr_i$ 均属于属性域 $attr$。一条迹 σ 是一个非空的事件序列，即 $\sigma=\langle e_1,e_2,\cdots,e_n\rangle$。事件日志 $EL=\langle\sigma_1,\sigma_2,\cdots,\sigma_n\rangle$ 是一个非空的迹序列，$\xi=c\times a\times r\times t\times attr_{\{1,2,\cdots,n\}}$ 称为事件域，事件日志 $EL\in\xi$，其中每个事件只能出现一次，即可以通过唯一标识符识别事件。

表 1 列出了实例日志 EL 的一部分片段，包含案例 ID、事件 ID，以及时间戳、活动和资源等属性值。

表 1 事件日志片段

Table 1 Event log fragment

案例 ID	事件 ID	属性				
		时间戳	活动	资源	成本	…
1	001	2023-05-01 09:00	Submit Order	Supplier A	0	…
	002	2023-05-01 09:30	Process Order	Supplier A	50	…
	003	2023-05-01 10:00	Ship Goods	Transporter X	100	…
2	001	2023-05-02 08:30	Submit Order	Supplier B	0	…
	002	2023-05-02 09:00	Process Order	Supplier B	50	…
	003	2023-05-02 10:00	Ship Goods	Transporter Y	80	…
3	001	2023-05-03 10:30	Submit Order	Supplier C	0	…
	002	2023-05-03 11:00	Process Order	Supplier C	50	…
	003	2023-05-03 12:00	Ship Goods	Transporter Z	120	…
…	…	…	…	…	…	…

定义 2(标签 Petri 网[22]) $N=(PL,TR,F,AC,lf)$ 是一个元组，PL 是一组库所的有限集合，TR 是一组变迁的有限集合，其中 $PL\cap TR=\varnothing$；$F\subseteq(PL\times TR)\cup(TR\times PL)$ 是一组有向边的集合，称为流关系；AC 是一组活动标签，$lf\in TR\to AC$ 是标签函数。

定义 3(直接跟随图(见图 1)，行为特征矩阵) 业务流程图中的直接跟随图(Directly Follows Graph，DFG)定义为一个四元组 $DFG=(E,A,\boldsymbol{F},\boldsymbol{T})$，其中 $E=\{e_1,e_2,\cdots,e_{|E|}\}$ 是事件的有序序列，表示一组按时间顺序记录的事件；$A=\{a_1,a_2,\cdots,a_{|A|}\}$ 是活动的集合，表示在事件集合 E 中出现的所有不同活动；$\boldsymbol{F}$ 是一个 $|A|*|A|$ 的矩阵，其中 $F(i,j)$ 表示活动 a_i 直接跟随在活动 a_j 之后的频率，即 $F(i,j)$ 表示在事件集合 E 中，活动 a_i 在活动 a_j 之后出现的次数；$\boldsymbol{T}$ 也是一个 $|A|*|A|$ 的矩阵，其中 $T(i,j)$ 表示从活动 a_i 到活动 a_j 的平均时间间隔，即 $T(i,j)$ 表示在事件集合 E 中，从活动 a_i 到活动 a_j 的平均时间间隔；$\boldsymbol{FT}$ 表示直接跟随图的行为特征矩阵。

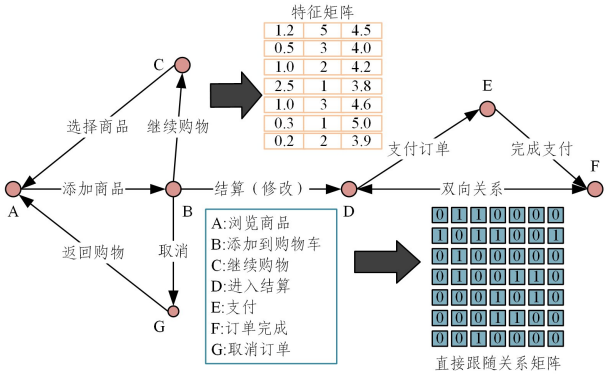


图 1 直接跟随图

Fig. 1 Directly follows graph

定义 4(直接跟随邻接矩阵) 直接跟随邻接矩阵 DDFAM(Directed Directly Follows Adjacency Matrix)是一个二维矩阵，用于表示行为图中活动节点之间的直接跟随关系。假设 $\boldsymbol{ADFG}=(A,D)$ 是一个直接跟随行为图，其中 A 是活动集合，D 是直接跟随关系的集合。DDFAM 的具体表达如下：

$$\boldsymbol{DDFAM}=[ddfam_{ij}]_{m\times m} \tag{1}$$

其中，$ddfamij$ 是矩阵 $\boldsymbol{DDFAM}$ 中的元素，对应于活动节点 i 和活动节点 j 之间的直接跟随关系。如果存在一条有向边从活动节点 i 指向活动节点 j，即活动 i 在某个事件日志中直接跟随活动 j，那么 $ddfam_{ij}=1$；如果不存在这样的有向边，即活动 i 从未在任何事件日志中直接跟随活动 j，那么 $ddfam_{ij}=0$。对于所有 $i,j\in[1,m]$，矩阵 $\boldsymbol{DDFAM}$ 满足以下条件：

1)如果 a_i 是事件日志中的活动，a_j 是活动 a_i 之后直接发生的活动，那么 $ddfam_{ij}=1$；

2)如果 a_i 之后没有直接发生活动 a_j，或者 a_j 不是 a_i 的直接后继，那么 $ddfam_{ij}=0$。

定义 5(行为特征图(Activity Feature Graph)) 行为特征图 $AFG=(DFG,\boldsymbol{FT})$ 由一条迹的直接跟随图和该迹的活动平均行为特征矩阵构成。其中 DFG 代表了一条迹中事件的行为走向，即控制流视角；$\boldsymbol{FT}$ 代表了一条迹中事件的数据变化，即数据流视角。

4 AAHM 方法

图 2 给出了自适应异常处理方法(Adaptive Anomaly Handling Method，AAHM)的流程框架，该框架通过动态调整和反馈机制优化异常检测与修复流程。首先通过特征提取及编码阶段对事件日志进行处理，将活动记录转换为可直接输入模型的格式。该方法提供直接输入编码、水平堆叠编码

和一维向量编码 3 种编码策略，以适应不同的学习模型架构。

将处理好的数据输入异常检测模型，利用编码后的数据训练分类器，以区分正常与异常行为。检测结果通过正向和负向反馈机制进行评估：正向反馈确认正常行为，促进模型稳定性；而负向反馈则针对异常行为，自适应地动态调整和进行异常修复。本章将对方法的各个步骤进行详细阐述。

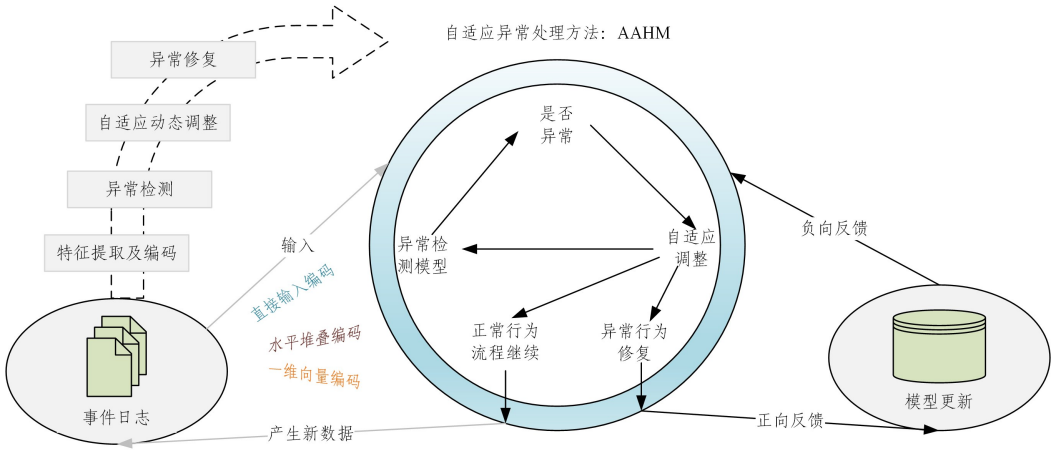


图 2 整体框架图

Fig. 2 Overall framework graph

4.1 特征提取及编码

图 3 给出了 3 种编码策略：1）直接将行为特征矩阵 ***FT*** 与直接跟随邻接矩阵 ***DDFAM*** 作为图神经网络的输入，以便模型捕捉活动节点间的复杂时序依赖性；2）通过水平堆叠 ***FT*** 与***DDFAM***形成二维图像矩阵，为卷积神经网络（CNN）提供必要的空间结构特征，使得网络能够有效识别局部模式；3）将二维矩阵展平为一维向量，适用于不依赖空间结构信息的机器学习或深度学习模型。这些编码策略基本满足了不同学习模型的输入需求，能够有效识别异常模式。

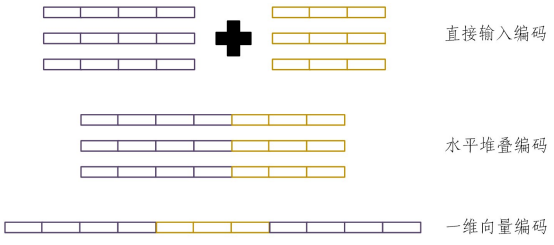


图 3 特征编码方式

Fig. 3 Feature coding method

4.2 异常检测

将每一条事件日志转换为行为特征图，这一过程不仅捕捉了单个事件的属性，而且整合了事件之间的时序和关联信息。随后构建异常检测流程，如图 4 所示，事件日志被分为正常和异常两大类，并分别进行了标签化处理。对于那些仅包含正常行为的日志，采用人工注入异常，以模拟真实世界中可能出现的各类异常情况，从而增强模型的泛化能力。在这一过程中，每一条注入异常的迹都被重新标记，以确保模型能够学习到异常行为的特征。

分类器的训练是通过行为特征图作为输入，以及对应的异常和正常标签作为监督信号来完成的。异常检测模型实质上是一个图的二分类任务，其中每一条迹都被视为一个包含控制流和数据流信息的图结构。对这些图结构进行细致的分析和分类，能够有效识别和定位异常行为，从而为业务流程的监控和优化提供有力的数据支持。

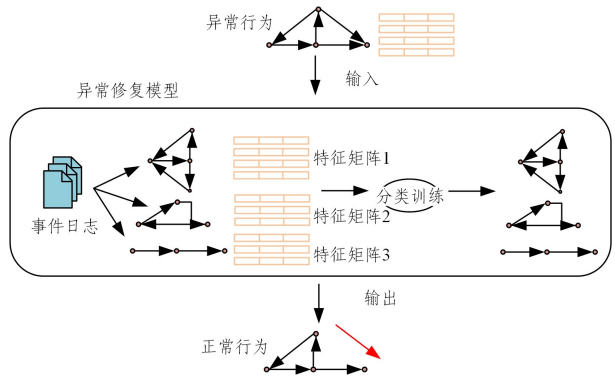


图 4 异常检测流程图

Fig. 4 Flow graph of anomaly detection

4.3 自适应动态调整

自适应共振理论（Adaptive Resonance Theory，ART）作为一种动态的神经网络模型[23]，适合处理实时数据流中的异常检测任务。

针对 ART 网络参数的初始化，节点数量的确定基于对历史业务流程数据规模和复杂度的分析。对过往数据中不同活动数量、资源种类以及事件关联的复杂程度进行统计评估，学习率的初始化参考同类业务流程中数据变化的平均速率以及期望的模型收敛速度。若数据变化相对缓慢，且希望模型能较快稳定，学习率可设置在较低水平，如 α_1；反之，若数据更新频繁，则可适当提高学习率至 α_2。警觉性参数的设定依据业务流程中正常行为模式的稳定性和可变性程度。对于具有严格规范和较少异常波动的流程，将警觉性参数设为较高值，如 β_1，以减少误判；对于相对灵活多变的业务流程，降低警觉性参数至 β_2，确保能够及时捕捉到潜在异常。以此展示本文方法的灵活普适性。

ART 自适应模型整体流程如图 5 所示，系统还包括一个紧密结合了性能评估指标的反馈调整环节。每经过一个固定的时间周期（如具体时间周期 T）或处理一定数量的事件日志

（如具体数量 N）后，模型会根据关键性能指标的数值预设目标值的偏差情况，对模型参数进行微调。同时，在反馈调整过程中，会同步记录参数调整的历史信息，形成一个参数调整轨迹日志。通过对该日志的分析，可以进一步了解模型在不同阶段的性能变化趋势以及参数调整的有效性。整个流程通过一个直观的流程图来表示，确保了异常检测过程的透明性和可追踪性。

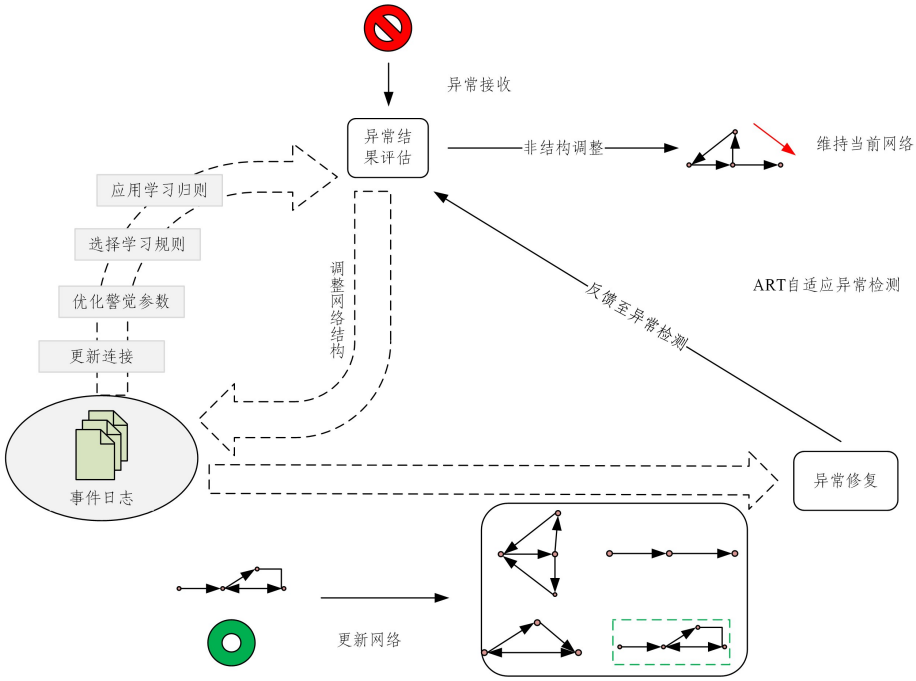


图 5 ART 自适应模型

Fig. 5 ART adaptive model

自适应算法的具体流程和解释如算法 1 所示。

算法 1 Adaptive Dynamic Adjustment Algorithm

INPUT：事件日志集 L，初始参数 θ

OUTPUT：ART 自适应模型 M

```
1. θ_ART ← init(L);
2. F ← extract_features(L);
3. foreach f ∈ F do
4.     θ_ART ← fast_learn(f,θ_ART);
5.     θ_ART ← update_weights(θ_ART,f);
6. foreach f′ ∈ new_logs do
7.     match_result← match(f′,θ_ART);
8.     if match_result=no_match then
9.         label(f′) ← anomaly;
10.    if label(f′)=anomaly then
11.        θ_ART ← adjust_threshold(θ_ART,f′);
12.        acc,rec← calculate_metrics(θ_ART,F);
13.    if acc <threshold_acc or rec < threshold_rec then
14.        θ_ART ← fine_tune(θ_ART,F);
15.        M← retrain(L,θ_ART);
16.        M← θ_ART
```

算法 1 通过初始化步骤设置 ART 网络基础参数 θ，利用特征提取步骤从事件日志集 L 中生成特征向量集 $\boldsymbol{F}$，为模型提供输入数据。第 5－7 行为模型训练阶段，算法遍历特征向量集 $\boldsymbol{F}$，对每个向量 $\boldsymbol{f}$ 应用快速学习机制，并更新网络权重，以逐渐构建对正常行为模式的认知。基于自适应模型的异常检测对新的事件日志特征向量 $\boldsymbol{f}'$ 进行模式匹配，以判断其是否与已知的正常行为模式相匹配。若无匹配，则将 $\boldsymbol{f}'$ 标记为异常。第 12－13 行展示了模型检测到异常时的动态性，其调整 ART 网络的警觉性阈值，以提高后续异常检测的敏感性。性能评估阶段，采用计算准确率和召回率来衡量模型的检测效能。若模型性能未达到预设阈值，则对参数进行微调，并使用整个日志集 L 对模型进行重训练，以提升其性能。算法输出结果为经过自适应动态调整后的 ART 自适应模型 M。

4.4 异常修复

修复机制基于对行为特征图（AFG）的深入分析，结合直接跟随图（DFG）和行为特征矩阵（$\boldsymbol{FT}$）来综合考量事件的时间序列和数据属性。

异常行为的修复过程首先聚焦于对正常行为轨迹的分类学习，以此训练模型对正常模式的识别能力。随后，该模型被用于对异常行为进行分类预测，通过匹配最相似的正常行为图谱，实现对异常行为的校正。在此过程中，将行为修复任务定义为多分类问题，利用正常行为图作为输入，通过模型预测其对应的正常类别标签。针对修复过程中可能出现的属性缺失问题，进一步提出了特征向量补全的方法。该方法将行为特征图中的节点特征向量串联成单一迹向量，并识别出其中的缺失值。此外，通过构建待修复矩阵，结合完整迹向量与缺失迹向量，采用了包括 GAIN[24] 在内的一系列数据插补技术，对缺失特征进行估计和填充。这些方法的优势在于它们能够直接处理带有缺失值的数据集，无需进行复杂的预处理或训练步骤，从而能够快速有效地恢复数据的完整性。

基于特征图的异常修复算法的具体流程和解释如图 6 和算法 2 所示。

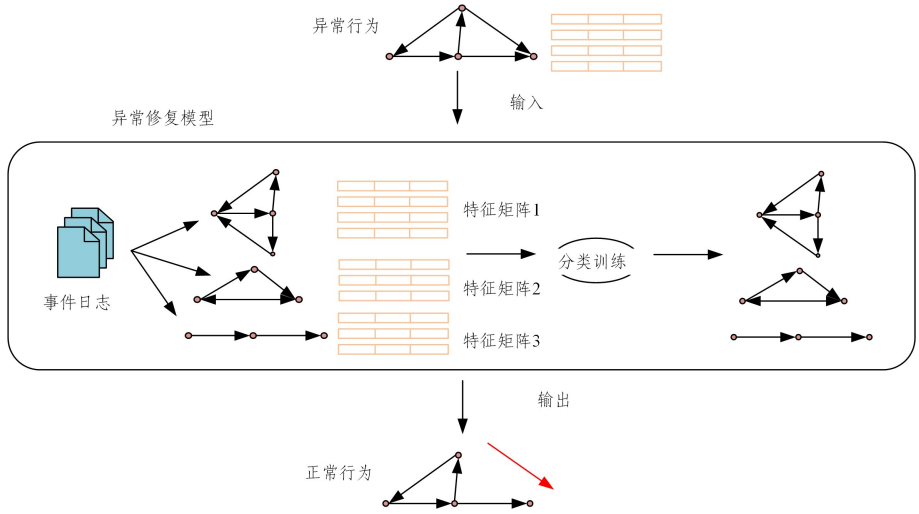


图 6 异常修复流程图

Fig. 6 Flow graph of abnormal repair

算法 2 Behavioral Anomaly Rectification through Feature Graph Enhancement

INPUT: AB, NB, **FT**, DFG, 修复参数 φ

OUTPUT: 修复后的行为集 R

```
1. φ ← init_repair(AB, NB, FT, DFG);
2. foreach a∈ AB do
3.     AFG_a← construct_AFG(a, FT, DFG);
4.     C← search_NB(AFG_a, NB);
5.     Va← extract_features(AFG_a);
6.     foreach n∈ C do
7.         AFG_n← construct_AFG(n, FT, DFG);
8.         Vn← extract_features(AFG_n);
9.         d← calc_distance(Va, Vn);
10.        n_best∈ C;
11.        a ← update_behavior(a, n_best, φ);
12. foreach a∈ AB ∪ NB do
13.    v← create_trace_vector(a);
14.    M← identify_missing_values(v);
15.    foreach m∈ M do
16.        e← gain(v, m);
17.        v← fill_missing_value(v, m, e);
18.        R← create_repaired_set(AB, φ);
19. return R
```

算法 2 通过异常行为集 AB、正常行为集 NB 以及 $\boldsymbol{FT}$ 和 DFG 进行传参，以初始化修复参数 φ。第 3－13 行构建异常行为 a 的行为特征图 AFG_a，在正常行为集中搜索与 AFG_a 最相似的正常行为集合，从而提取 AFG_a 的特征向量。与其对比的操作是构建正常行为 n 的行为特征图 AFG_n 特征向量提取 AFG_n 的特征向量，以便计算异常行为与候选正常行为之间的距离。选择与异常行为最相似的正常行为，最终根据 n_best 更新异常行为 a 的特征，进行行为修复。算法后段（第 14－21 行）的主要功能是将行为特征图中的节点特征向量串联成单一迹向量，从而识别迹向量中的缺失值，使用 GAIN 算法对缺失值进行估计，同时填充迹向量的缺失值，根据修复参数 φ 创建包含所有修复后行为的新集合 R 并返回。

5 实验分析

5.1 实验设置

表 2 列出了 4 组真实的事件日志数据集，即 BPIC_2014、BPIC_2015、败血症日志（Sepsis Cases）和道路交通罚款日志（Road Traffic Fines）。BPIC_2014 是荷兰合作银行集团 ICT 从“HP 服务管理器”的 ITIL 服务管理工具中提取的详细信息；BPIC_2015 由荷兰 5 个城市提供，日志中的案例包含有关主要的申请信息以及各个阶段的异议程序；败血症日志（SC）是一份医院事件日志，记录了疑似患有危及生命的脓毒症的患者从医院急诊室到出院的轨迹，在共 1 050 个轨迹中，有 846 个独特的轨迹变体；道路交通罚款日志（RTF）是意大利当地警方处理道路交通过程中获得的事件日志，在约 10 000 个轨迹的样本中，只有 69 个不同的轨迹变体。在实验阶段，将所提及的日志集视为无误的标准事件日志。随后，为了模拟异常行为并进行相应的修复分析，对这些日志中的流程轨迹实施 3 种随机操作：1）在某些随机选定的事件之后插入相同事件的副本；2）随机移除一些事件；3）随机交换某些事件的顺序。通过这些操作构建异常行为的实例（已 5%，10%，15%和 20%比例注入），并进一步进行异常检测和修复的深入分析。

表 2 事件日志属性

Table 2 Event log attribute

事件日志	可用性	活动数量	事件数量	迹数量	资源数量
BPIC_2014	公开	39	466 155	46 507	242
BPIC_2015	公开	356	262 628	5 649	72
SC	公开	16	15 012	1 050	39
RTF	公开	30	145 800	10 005	62

为证明本文方法的可行性，进行了两组实验，分别是异常行为检测的准确率和异常行为修复的准确率。实验选择 CNN，DBN，LOF，AE，LSTM 和 XGBoost 这 6 种分类器方法与本文方法进行对比实验。CNN，DBN 和 LOF 分类器采用本文所使用的方法。AE，LSTM 和 XGBoost 分类器在自然语言处理任务中采用一种前缀迹编码的方法，在这种方法中，

整个轨迹被视为一个前缀迹，其中每个事件及其属性被当作一个单词特征。编码过程中，选取轨迹中最长的长度作为前缀迹的长度，并且预测的标签保持一致。对于非数值属性，这两组实验技术都采用 one-hot 编码策略。

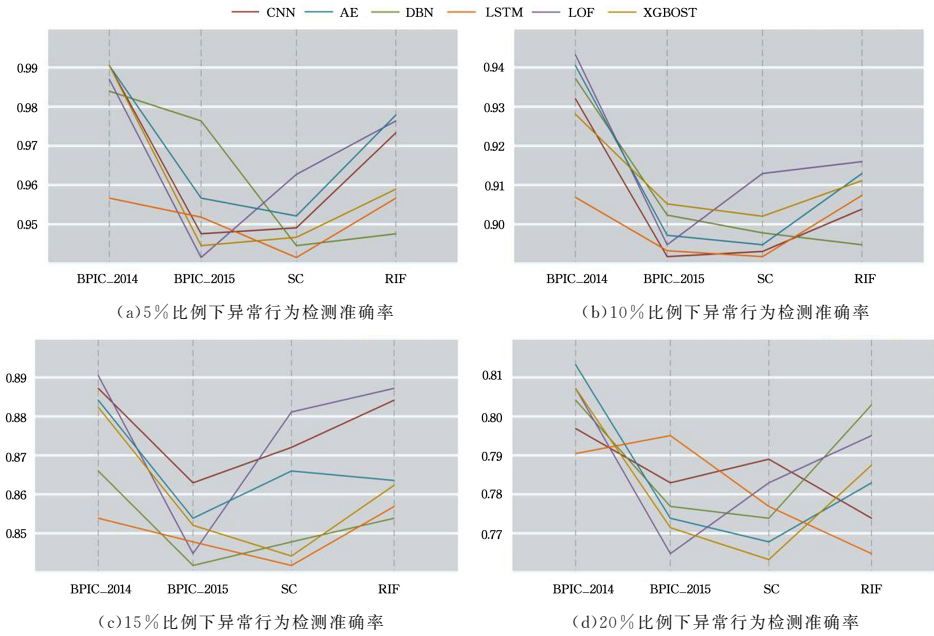
考虑到实验中涉及的分类器方法在处理数据时可能存在本质上的差异，选择了 Kruskal-Wallis 检验作为验证方法。Kruskal-Wallis 检验是一种非参数方法，它不依赖于数据的分布形态，允许比较多个独立样本组的中位数，特别适用于样本数据不满足正态分布或方差齐性假设的情况。通过 Kruskal-Wallis 检验，能够识别出在不同异常注入比例下哪些分类器展现出显著的性能差异。

5.2 异常检测

上述 6 类分类器在传统异常检测领域的侧重点不同，例如：CNN 在处理具有空间结构特征的数据方面表现出色，而业务流程事件日志中的活动序列和关联信息在一定程度上可被视为具有空间结构特征；DBN 能够学习复杂的非线性关系，对于业务流程中复杂的因果关系和行为模式可能具有较好的建模能力；LOF 是一种基于密度的异常检测方法，其优势在于能够有效识别局部数据密度差异较大的异常点，在业务流程中对偏离正常行为模式的孤立异常事件可能具有较高的检测敏感性；AE 在特征学习和数据降维方面有独特的优势，能够提取数据的关键特征，有助于在异常检测和修复中发现隐藏在复杂数据中的异常信息；LSTM 常被用于处理时间序列数据，考虑到业务流程事件具有时间顺序性，LSTM 可以利用其记忆单元捕捉长短期的时间依赖关系，从而识别异常行为的时间模式；XGBoost 是一种高效的梯度提升决策树算法，在分类和回归任务中表现优异，能够通过集成多个弱学习器来提高预测准确性，对于业务流程异常检测和修复中的分类问题可以提供有效的解决方案。

注入比例选择 5%，10%，15% 和 20% 这 4 个阶段性指数。在实际业务流程中，异常情况通常以一定的低概率出现，但在某些复杂或不稳定的环境下，异常比例可能会有所上升。基线结果表明，5% 的比例代表了相对比较稳定的业务流程中可能出现的轻微异常情况，可用于测试分类器在常规环境下的检测和修复。随着比例逐渐增加到 10%，15% 和 20%，通常情况下，20% 比例下的干扰模型与原模型的相似度差异低于 60%，不再具备实验参考性，旨在模拟业务流程受到不同程度干扰或处于较为复杂多变的场景，考查分类器在应对逐渐增多的异常数据时的性能变化趋势，从而全面评估分类器的鲁棒性和适应性。

图 7 展示了 6 种分类器在不同事件日志和不同异常注入比例下的异常检测准确率，揭示了本文方法所采用的基于自适应方法与传统的前缀迹编码方式相比，在异常检测任务中并不逊色。特别是在高异常注入比例的情况下，本文方法展现出了更加卓越的性能，这表明它能够有效地捕捉到异常模式，即使在数据受到严重干扰时也能保持较高的准确率。为了进一步论证本文方法的有效性和鲁棒性，进一步进行了 Kruskal-Wallis 检验实验。



(a)5%比例下异常行为检测准确率　(b)10%比例下异常行为检测准确率

(c)15%比例下异常行为检测准确率　(d)20%比例下异常行为检测准确率

图 7　不同异常注入率下的检测准确率

Fig. 7　Detection accuracy under different abnormal injection rates

表 3 列出了基于 Kruskal-Wallis 检验的分类器性能分析，针对在不同异常注入比例下的表现进行的秩次分配示例，在这里仅以 BPIC_2014 数据集作为示例。

累积秩次的计算提供了一个量化不同分类器整体性能的检测方法，它通过累加每个分类器在所有测试条件下的秩次来实现。

累积秩次数值如表 4 所列。分析结果可知，LOF 和 CNN 分类器在累积秩次上表现最佳，分别以 13 和 14 的累积秩次数取得显著优势。这两种分类器在不同异常注入比例下均能保持较高准确率，反映出其自适应性能和对异常数据的鲁棒性。特别是 LOF，作为一种基于局部密度的方法，它在检测异常行为方面展现出了较高的敏感性和准确性。而 AE，

LSTM 和 XGBOOST 的累积秩次数较高，分别为 27，24 和 23，这可能表明这些方法在处理特定类型的异常数据或在特定的异常注入比例下存在一定的局限性，与自适应方法相比结果不佳。

表 3 基于 Kruskal-Wallis 检验的 BPIC_2014 示例

Table 3 BPIC_2014 example based on Kruskal-Wallis test

(%)

分类器	BPIC_2014			
	5	10	15	20
CNN	3.5	3	3	3
DBN	2.5	2.5	2	2
LOF	3.5	2.5	1	1
AE	4	4	4	4
LSTM	5	5	5	5
XGBOOST	6	6	6	6

表 4 基于异常检测的不同分类器累积秩次数值

Table 4 Cumulative rank values of different classifiers based on anomaly detection

分类器	累积秩次
CNN	14
DBN	20
LOF	13
AE	27
LSTM	24
XGBOOST	23

在相同的异常注入比例设置下，以往相关研究中的异常检测和修复方法，如前文提到的结合流程挖掘与模糊关联规则学习的方法，在低比例异常时表现尚可，但在高比例异常(如 15%和 20%)时，其检测准确率和修复成功率明显低于 AAHM 方法。这主要是因为该方法的模型结构相对简单，在处理复杂业务流程数据时，其基于模糊关联规则学习的方式对异常的识别能力有限，且缺乏如 AAHM 方法中的自适应调整机制，难以根据数据的动态变化，及时优化模型，从而在面对高比例异常时无法有效应对。再比如基于事件流的在线流程挖掘器方法，其虽然在处理具有一定实时性的业务流程数据方面有一定优势，但在整体的不同异常注入比例测试中，AAHM 方法在平均准确率方面更优。

从另一层面来看，相较于传统前馈神经网络、循环神经网络等，本文方法提出的快速学习规则使得 ART 模型在面对实时数据流时，能迅速捕捉并更新正常行为模式转换为内部表示，有效避免了数据快速更新导致的模型滞后。例如在供应链流程中，新数据流入时，本文模型相比传统模型，可依据已有记忆和学习机制快速判断并调整权重。同时，警觉性参数设置提升了异常判别的精准度。因业务流程有规律，正常模式集中，异常偏离常规，故合理设置该参数可减少误判和漏判，保持较高检测性能。这是传统神经网络模型较难实现的自适应性。

5.3 异常修复

表 5 列出了不同分类器在多个事件日志分类和不同异常注入比例下的异常修复成功率对比。通过这些实验发现，CNN，DBN 和 LOF，即采用了本文所提出的自适应方法的分类器，在多数情况下的修复成功率均高于采用了传统前缀迹编码的 AE，LSTM 和 XGBOOST 分类器。特别是在异常注入比例较高时，本文方法展现出了更为卓越的性能。这一点在 BPIC_2014 和 BPIC_2015 数据集上尤为明显。值得注意的是，本文方法在高异常注入的情况下依然能够达到 50%以上的修复成功率。

表 5 不同分类器异常修复成功率对比

Table 5 Comparison of success rate of anomaly repair in different classifiers

(%)

事件日志分类	异常比例注入	CNN	DBN	LOF	AE	LSTM	XGBOOST
BPIC_2014	5	82.16	81.56	82.57	68.89	65.86	63.25
	10	81.18	79.14	79.42	61.17	58.98	60.58
	15	73.31	74.25	78.96	53.22	48.74	58.55
	20	69.92	74.13	78.75	51.47	43.52	51.71
BPIC_2015	5	81.58	77.85	74.78	58.12	57.52	52.55
	10	74.42	69.75	71.52	52.63	48.85	40.26
	15	63.58	64.25	58.41	43.75	41.93	38.74
	20	60.97	54.49	52.63	39.18	31.14	29.89
SC	5	54.39	47.58	41.33	28.46	39.42	32.92
	10	51.23	44.25	35.62	22.83	28.66	20.86
	15	43.29	38.74	37.91	13.59	21.75	15.28
	20	40.57	30.62	34.53	7.63	13.95	9.77
RTF	5	67.23	59.41	69.88	49.25	35.74	32.96
	10	59.43	55.74	61.47	32.13	28.12	20.17
	15	53.26	54.23	52.82	23.78	15.83	18.18
	20	50.92	41.69	40.75	19.15	13.18	16.23

为了进一步验证这些观察结果的统计显著性，同样进行了 Kruskal-Wallis 检验，结果如表 6 所列。实验结果表明，不同分类器的性能存在显著差异，本文方法在异常修复上累积秩次数依旧占优。

表 6 基于异常修复的不同分类器累积秩次数值

Table 6 Cumulative rank values of different classifiers based on exception repair

分类器	累积秩次
CNN	17
DBN	26
LOF	15
AE	21
LSTM	35
XGBOOST	22

结束语 本文指出业务流程动态变化过程中，传统异常检测方法无法完成快速响应，因而基于自适应理念，提出了自适应异常处理方法 AAHM，并详细介绍了其中的关键步骤。首先，将事件日志中的迹转换为行为特征图，并通过多元化的编码方式，在 ART 模型中进行训练学习，从而实现不同场景下的自适应调整。其次，对比所提出的基于行为特征图的检测和修复方法与常规编码方法的实验结果发现，在融入自适应模型之后，本文方法的准确率和修复率均不弱于常规的编码方式。

尽管 AAHM 方法在当前研究中展现出一定优势，但仍存在一些亟待突破的局限性。例如在应对超高复杂度与噪声干扰严重的业务流程数据时，异常检测的准确性与稳定性面临挑战，复杂的数据波动易引发误判或漏判；对于部分业务规则模糊不清、充满不确定性的场景，模型的理解与适应能力尚显不足，难以精准捕捉异常特征。同时，在整合多源异构数

据方面，如融合企业内部不同系统数据与外部市场数据时，现有方法在数据结构统一、语义理解与特征融合上存在缺陷，限制了异常检测与修复的全面性与深度。

后续研究方向将聚焦于高效的智能数据预处理技术，结合深度学习的自动编码器与小波变换等方法，有效降低数据噪声、填补缺失值，增强数据质量，提升本文方法在复杂数据环境下的可靠性，进而提升在业务流程异常处理中的效能与普适性，推动其在更多领域的深度应用与发展。

参 考 文 献

[1] SARNO R，SINAGA F，SUNGKONO K R. Anomaly detection in business processes using process mining and fuzzy association rule learning[J]. Journal of Big Data，2020，7(1)：5.

[2] VAN ZELST S J，SANI M F，OSTOVAR A，et al. Detection and removal of infrequent behavior from event streams of business processes[J]. Information Systems，2020，90：101451.

[3] KRAJSIC P，FRANCZYK B. Catch me if you can：online classification for near real-time anomaly detection in business process event streams[J]. Procedia Computer Science，2022，207：235-244.

[4] GUAN W，CAO J，GU Y，et al. GAMA：A multi-graph-based anomaly detection framework for business processes via graph neural networks[J]. Information Systems，2024，124：102405.

[5] NOLLE T，LUETTGEN S，SEELIGER A，et al. Binet：Multi-perspective business process anomaly classification[J]. Information Systems，2022，103：101458.

[6] KO J，COMUZZI M. Keeping our rivers clean：Information-theoretic online anomaly detection for streaming business process events[J]. Information Systems，2022，104：101894.

[7] BÖHMER K，RINDERLE-MA S. Multi-perspective anomaly detection in business process execution events[C]// On the Move to Meaningful Internet Systems：OTM 2016 Conferences. Springer，2016：80-98.

[8] NOLLE T，LUETTGEN S，SEELIGER A，et al. Analyzing business process anomalies using autoencoders[J]. Machine Learning，2018，107：1875-1893.

[9] KRAJSIC P，FRANCZYK B. Variational Autoencoder for Anomaly Detection in Event Data in Online Process Mining[C]// ICEIS. 2021：567-574.

[10] KRAJSIC P，FRANCZYK B. Semi-supervised anomaly detection in business process event data using self-attention based classification[J]. Procedia Computer Science，2021，192：39-48.

[11] NGUYEN H T C，LEE S，KIM J，et al. Autoencoders for improving quality of process event logs[J]. Expert Systems with Applications，2019，131：132-147.

[12] TAVARES G M，JUNIOR S B. Process miningencoding via meta-learning for an enhanced anomaly detection[C]// Proceedings of European Conference on Advances in Databases and Information Systems. Cham：Springer，2021：157-168.

[13] BÖHMER K，RINDERLE-MA S. Mining association rules for anomaly detection indynamic process runtime behavior and explaining the root cause to users[J]. Information Systems，2020，90：101438.

[14] FANI SANI M，VAN ZELST S J，VAN DER AALST W M P. Repairing outlier behaviour in event logs[C]// Business Information Systems：21st International Conference，BIS 2018. Cham：Springer，2018：115-131.

[15] FANI SANI M，VAN ZELST S J，VAN DER AALST W M P. Applying sequence mining for outlier detection in process mining[C]// On the Move to Meaningful Internet Systems. OTM 2018 Conferences. Cham：Springer，2018：98-116.

[16] GUO H C，GUO Y H，YANG J，et al. Loglg：Weakly supervised log anomaly detection via log-event graph construction[C]// Proceedings of International Conference on Database Systems for Advanced Applications. Cham：Springer，2023：490-501.

[17] MEHR A S M，DE CARVALHO R M，VAN DONGEN B. Explainable conformance checking：understanding patterns of anomalous behavior[J]. Engineering Applications of Artificial Intelligence，2023，126：106827.

[18] LEE S，LU X，REIJERS H A. The analysis of online event streams：Predicting the next activity for anomaly detection[C]// Proceedings of International Conference on Research Challenges in Information Science. Cham：Springer，2022：248-264.

[19] ZAVATTERI M，BRESOLIN D，DE LEONI M. Repair of unsound data-aware process models[C]// Proceedings of International Conference on Business Process Management. Cham：Springer，2023：383-395.

[20] CONFORTI R，LA ROSA M，TER HOFSTEDE A H M，et al. Automatic repair of same-timestamp errors in business process event logs[C]// Business Process Management：18th International Conference，BPM 2020. Cham：Springer，2020：327-345.

[21] GUAN W，CAO J，GU Y，et al. AIMED：An automatic and incremental approach for business process model repair under concept drift[J]. Information Systems，2023，119：102285.

[22] FANG H，LIU W C G，WANG W S，et al. Discovery of process variants based on trace context tree[J]. Connection Science，2023，35(1)：2190499.

[23] MASUYAMA N，AMAKO N，YAMADA Y，et al. Adaptive resonance theory-based topological clustering with a divisive hierarchical structure capable of continual learning[J]. IEEE Access，2022，10：68042-68056.

[24] LEI S. A feature selection method based on information gain and genetic algorithm[C]// International Conference on Computer Science and Electronics Engineering. IEEE，2012：355-358.

LIU Fujie，born in 2000，postgraduate. Her main research interests include Petri nets and process mining.

FANG Xianwen，born in 1975，Ph. D，professor. His main research interests include Petri nets and trusted computing.

（责任编辑：何杨）