doi:10.3969/j.issn.1001-4616.2025.05.012

基于经验回放的日志异常检测模型的更新

卜意磊[1],李嘉硕[2],赵　斌[2],庞文迪[1]

(1.江苏省市场监管局数据中心,江苏 南京 210036)
(2.南京师范大学计算机与电子信息学院,江苏 南京 210023)

[**摘要**]　基于日志的异常检测是异常检测问题的重要分支,得到越来越多的关注. 然而,现有研究往往忽略了在长时间检测的场景下,数据分布及模式变化对日志异常检测产生的影响. 为了实现持续性有效检测的目标,本文提出了基于增量学习的日志异常检测模型的更新方法,使用黑暗经验回放策略在原有的先进方法 MLog 的基础上进行改进. 在利用原有数据充分训练模型的基础上,更新算法使用聚类得到的范例样本和收集的新样本增量更新模型,其中对范例样本应用蒸馏损失,保留更多知识从而减少遗忘. 进一步,为了保留更多已学习到的特征信息,本文方法提取 MLog 的特征融合层的中间层特征,使用范例样本的类别原始分数和中间特征共同约束模型的更新,实现完整经验重放. 在真实数据集上的实验结果表明,在持续检测需求场景下,本文方法能够有效提高检测模型训练的时间效率,并且获得了与全量训练相近的检测效果.

[**关键词**]　异常检测,日志分析,增量学习,深度学习
[**中图分类号**]TP301　[**文献标志码**]A　[**文章编号**]1001-4616(2025)05-0104-10

Update of A Log Anomaly Detection Model Based on Experience Replay

Bu Yilei[1],Li Jiashuo[2],Zhao Bin[2],Pang Wendi[1]

(1.Jiangsu Provincial Market Management Bureau Data Center,Nanjing 210036,China)
(2.School of Computer and Electronic Information,Nanjing Normal University,Nanjing 210023,China)

Abstract:Anomaly detection based on logs is an important branch of the anomaly detection problem,gaining increasing attention. However,existing studies often overlook the impact of changes in data distribution and patterns in long-term detection scenarios on log anomaly detection. To achieve the goal of sustainable and effective detection,this paper proposes an update method for a log anomaly detection model based on incremental learning,which improves upon the existing advanced method MLog using a Dark Experience Replay(DER) strategy. Building on the existing data to thoroughly train the model,the updating algorithm incrementally updates the model using exemplary samples obtained from clustering and newly collected samples. In this approach,a distillation loss is applied to the exemplary samples to retain more knowledge and reduce forgetting. To retain more learned feature information,this method called Full Experience Replay(FER) extracts the intermediate features from the feature fusion layer of MLog,to utilizes the original class scores of the sample data and the intermediate features to jointly constrain the model's update. Experiments on real dataset show that in scenarios with continuous detection requirements,the method proposed in this paper can effectively improve the training time efficiency of the detection model while achieving detection performance comparable to that of full training.

Key words:anomaly detection,log analysis,incremental learning,deep learning

现代软件系统的规模不断增大,结构也越发复杂,软件系统容易出现更多的系统问题和潜在风险,高可用性和可靠性对于大型软件系统至关重要[1]. 为此,学术界开始关注基于日志的异常检测研究. 日志异常检测在维护系统安全、提高系统可靠性和稳定性等方面发挥着重要作用[2]. 目前,日志异常检测已经在超级计算机[3]、分布式数据库[4]等多种信息系统中应用.

收稿日期:2024-11-03.
基金项目:江苏省市场监管局科技计划项目(KJ2025043).
通讯作者:赵斌,博士,副教授,研究方向:大数据分析与挖掘. E-mail:zhaobin@njnu.edu.cn

近年来,机器学习和深度学习技术的发展为日志异常检测问题提供了更为有效的解决方案. 许多研究[5-6]致力于解决由日志演变和噪声导致的日志数据不稳定问题. 然而,日志数据的分布或模式随时间发生变化也是导致日志不稳定性的原因之一,进而影响日志异常检测的效果. 当前已有研究人员注意到该问题并展开了相关研究[7-11]. 经过调研发现,虽然现有解决方案试图使用误判信息或在线学习的方式增量更新模型,但更新时存在数据利用不充分或知识遗忘等问题,缺乏持续学习的能力. 为了提升检测模型的增量式学习能力,本文在日志异常检测研究中引入增量学习(incremental learning)的经验回放(experience replay,ER)[12]技术. 经验回放利用额外存储的历史数据样本或已学习的特征帮助模型"回忆"知识,将过去的经验重新引入到模型的训练过程中,从而实现模型的持续学习,有效应对数据分布及模式变化造成的不利影响.

在日志异常检测研究中引入增量学习技术将面临新旧知识的平衡问题. 在实际的日志检测系统中,日志检测模型将面对持续到来的日志数据,检测模型在学习新数据时可能会遗忘之前学到的知识,尤其在训练数据分布发生变化时,容易出现"灾难性遗忘"现象. 增量学习方法需要采用合适的策略保留已学习到的关键知识,并在更新阶段约束模型训练,平衡新旧知识的权重,以保持模型的整体表现.

针对上述问题,本文设计并实现了基于经验回放的日志异常检测模型的更新方法,基本思路是在先进的日志异常检测方法 MLog 上进行改进,增强模型的持续学习能力. 当新产生小批量数据时,使用新数据和从旧数据中提取的范例样本对模型进行更新. 更新方法基于黑暗经验回放(DER)[13]策略,并保留重要的中间特征进行完整经验回放(full experience replay,FER)[14]. 与训练时不同之处在于更新时范例样本使用蒸馏损失,存储范例样本的原始类别分数和中间特征并在模型更新时进行回放,从而有效保留先前学到的知识.

1 相关工作

1.1 研究现状

1.1.1 基于日志的异常检测

日志中记录了系统运行的详细信息,能够帮助管理人员识别和解决系统中的问题. 基于日志的异常检测方法一直以来都得到相关专家和学者的广泛关注. 机器学习和深度学习技术已经在日志异常检测领域得到广泛应用,并取得了非常好的实验效果. 例如,针对日志语句演变和噪声问题,HilBERT[5]实现了基于层次化 Transformer 的专用日志表示模型,显著提升了检测方法的鲁棒性. Logs2Graph[15]将日志转换为包含语义信息的有向加权图,并利用图神经网络检测图级别的异常.

然而,减少日志不稳定性对异常检测结果的影响仍然是一项具有挑战性的任务. 现有方法主要关注由于系统更新或解析错误导致的日志语句演变、处理噪声或新的执行路径导致的少量新增序列,针对日志数据的分布及模式变化这一情况的研究较为有限. 例如,LogAnomaly[7]采用周期离线训练的模型更新方式,对于两次离线训练之间出现的新日志则采用模板近似与现有模板匹配. 研究[8-9]将检测错误的日志序列用于后续模型的更新任务,现实场景中依赖人工反馈才能实现. 此外,在线学习(online learning)也被用于解决日志不稳定的问题. 考虑到训练与测试数据的特性不同,ROEAD[10]使用在线镜像下降算法(online mirror descent,OMD)最小化累计损失,结合人工反馈动态更新模型参数. LogOnline[11]利用日志头部信息实现了额外的"正常"检测模型以识别高置信度的正常序列,用于增量更新异常检测模型. 现有方法仅使用部分新数据更新模型,更新时会遗忘部分旧知识,缺乏持续学习的能力.

1.1.2 深度增量学习

传统的深度学习假设数据服从独立同分布的特性,在训练阶段一次性处理所有数据. 然而,现实场景中数据分布可能会发生变化,深度学习方法更新时会面临计算成本高、难以适应新数据等问题. 而增量学习能够在动态环境中自适应学习,减少遗忘并提升模型的持续学习能力. 深度增量学习方法的研究已经取得有效进展,当前的深度增量学习方法大致可以划分为基于正则化的方法、基于回放的方法和基于动态网络的方法. 基于正则化的方法核心在于利用正则损失约束模型,放弃其遗忘关键知识,典型方法包括 oEWC[16]、SI[17]. 基于回放的方法利用额外存储的旧样本信息帮助模型巩固"知识",典型方法包括 DER[14]、SNCL[15]等. 基于动态网络的方法如 PNN[18]等,通过参数分配或网络扩张为后续更新提供冗余

空间.

增量学习的的任务场景主要包括任务增量学习、类增量学习和域增量学习. 表 1 展示了具有代表性的三类方法在多个数据集上进行增量任务的准确率.

表 1 增量更新方法在分类任务中的准确率比较

Table 1 The accuracy of incremental update methods in classification tasks

方法/数据集	S-CIFAR-10		S-Tiny-ImageNet		P-MNIST Domain-IL	P-MNIST Domain-IL
	Class-IL	Task-IL	Class-IL	Task-IL		
oEWC	19.49±0.12	68.29±3.92	7.58±0.10	19.20±0.31	75.79±2.25	77.35±5.77
SI	19.48±0.17	68.05±5.91	6.58±0.31	36.32±0.13	65.86±1.57	71.91±5.83
DER	70.51±1.67	93.40±0.39	17.75±1.14	51.78±0.88	87.29±0.46	92.24±1.12
SNCL	76.35±1.21	94.02±0.43	20.27±0.76	52.58±0.67	88.53±0.41	93.05±1.02
PNN	—	95.13±0.72	—	67.84±0.29	—	—

‘—’表示实验因超时或不适用任务而无法运行.

研究显示,基于回放的方法在多种增量学习任务上具有更好的综合表现[12]. 经验回放是增量学习中的常用策略,黑暗经验回放(DER)及完整经验回放(FER)是这类方法的典型代表,其核心思想是利用额外存储的范例样本和已学习的特征约束模型更新,已经在图像分类、图像恢复、目标检测等研究任务中得到应用. 例如,针对图像的多种恶劣天气去除任务,Cheng 等[19]提出了一种具有统一网络结构和有效知识回放的持续学习策略,将旧模型的预测结果和学习到的中间特征的主成分作为监督,对新模型进行蒸馏训练. Mo 等[20]将增量学习技术引入目标检测领域,专门针对单阶段检测器提出了一种多级前景提示增量学习算法,提供了图像级、特征级和知识级的前景提示.

1.2 日志异常检测方法 MLog

本文方法是在先进的异常检测方法 MLog[21]上进行改进的. MLog 方法遵循了当前日志异常检测领域的通用方法框架,首先将日志事件模板转换为包含语义信息和频率信息的向量表示. 为了捕获日志序列中的全局依赖和局部依赖,事件模板序列被输入到由 Mogrifier LSTM 和 CNN 构建的检测网络模型中,学习日志序列和标签的映射关系. 新产生的日志序列被输入到训练好的检测模型中,得到异常检测结果.

考虑到系统演变会导致日志语句发生变化,以及日志语句本身具有相似性等数据特点,MLog 提取事件模板的语义信息,使用语言模型 Bert 学习初始模板语义向量,然后将每个日志事件模板的发生频率与其初始语义向量集成来构建最终的模板语义表达. 模板语义表达被按顺序输入到异常检测模型中,首先通过基于 Mogrifier LSTM 的模型建模日志序列,然后通过基于 CNN 的特征融合层提取更为丰富和有意义的特征,提高了模型对重要特征的关注.

MLog 的优势在于其使用包含语义信息和频率信息的模板向量化方法,减小日志语句演变和类不平衡问题的负面影响. 此外,其通过结合 Mogrifier LSTM 和卷积神经网络(CNN),提高异常检测的准确性和鲁棒性. 这使得 MLog 在复杂系统的维护和故障诊断中具有重要的实际应用意义.

2 本文方法

2.1 问题定义

本文将日志异常检测问题建模为二分类问题. 假定日志序列形式化表示为 $S=e_1,\cdots,e_j,\cdots,e_s$,其中 e_j 表示第 j 个日志事件,s 是序列长度. 日志异常检测问题的输入是序列集合 $\boldsymbol{L}=\{S_1,S_2,\cdots,S_n\}$,$S_i$ 是 $\boldsymbol{L}$ 的第 i 个日志序列,$Y=\{y_1,y_2,\cdots,y_n\}$ 是与 $\boldsymbol{L}$ 对应的标签集合,其中 $y_i\in\{$正常,异常$\}$. 日志异常检测学习一个分类器 $f_\theta:\boldsymbol{L}\rightarrow Y$,$\theta$ 是模型参数. 对于给定的新实例,分类器能够输出预测标签.

增量更新是在已训练模型的基础上更新模型. 给定当前训练得到的模型表示 f_{θ_T}和旧数据集合 $\boldsymbol{L}_T$,T 代表历史时间区间. 假定经过新时间区间 ΔT 后收集的新数据集合为 $\boldsymbol{L}_{\Delta T}$,增量学习任务就是在已有模型的基础上更新得到新的分类器 $f_{\theta_{T+\Delta T}}:\boldsymbol{L}'\rightarrow Y'$,其中 $\boldsymbol{L}'=\text{sample}(\boldsymbol{L}_T)\cup\boldsymbol{L}_{\Delta T}$,$Y'$是与之对应的标签集合,更新后检测模型对于新旧数据实例都保持较好的检测效果.

2.2 异常检测模型更新总体框架

异常检测模型更新的总体框架主要分为两个模块:日志异常检测模块和增量更新模块. 在日志异常检测部分,本文采用了基于日志模板语义信息和混合神经网络的新型日志异常检测方法 MLog,其中包含 3 个步骤:日志解析、日志模板向量化和异常检测模型训练和预测. 增量更新模块实现基于黑暗经验回放(DER)策略的增量更新方法. 如图 1 所示,使用一段时间内收集到的新数据及代表性旧数据(即范例样本)更新模型参数. 具体来说,在已经使用原有数据集充分训练模型的基础上,当收集足够时间跨度的新样本之后,从旧数据中抽取代表性样本生成范例样本集,共同应用范例样本与小批量新样本更新模型,其中分别对范例样本和新样本应用蒸馏损失和交叉式损失. 为了减少训练偏差较大的样本被采样重播的影响,增量更新模块也对部分范例样本使用真实标签以增强新旧模型分类的一致性. 本文方法针对 MLog 模型结构,进一步使用完整经验回放(FER)[14]加强特征融合层的监督,从而保留更多已学习到的关键特征.

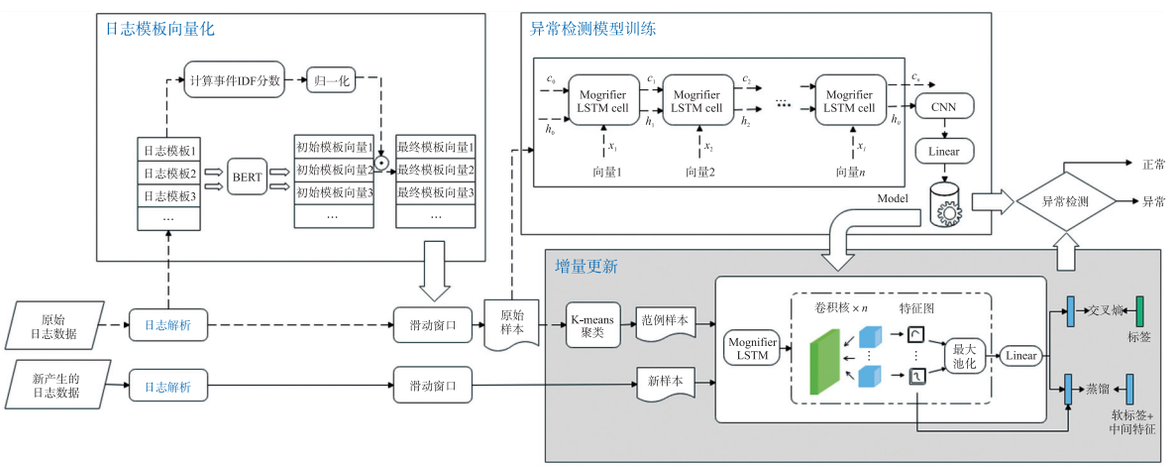

图 1 异常检测模型更新总体框架

Fig. 1 Overview of anomaly detection model updating

2.3 基于黑暗经验回放的日志异常检测模型的更新

在实际应用中,系统日志数据通常是持续生成的,系统的正常行为和异常行为可能会随着时间而变化. MLog 方法利用了日志中包含的语义和时序信息,并使用 Mogrifier LSTM 和 CNN 构建异常检测模型,能够有效应对由于日志演变和噪声导致的不稳定性问题. 然而,MLog 缺乏增量学习的能力,不能很好适应数据分布随时间的变化,这可能会导致对于新的正常或异常模式的识别准确度下降. 因此,本文在原有方法的基础之上,增加了增量更新模块以实现持续性有效检测的目标,通过在训练过程中回放存储在缓冲区中的范例样本来减轻灾难性遗忘. 需要注意的是,黑暗经验回放(DER)策略中范例样本的损失函数改为衡量预测值与软标签(旧模型网络输出)之间的差距,这样做避免了在概率空间中由于压缩函数(如 softmax)导致的信息损失. 由于可能会采样到在先前任务中存在高度偏差的样本,更新阶段也对部分范例样本使用真实标签,提高模型应对局部扰动的能力. 此外,本文方法保留 CNN 特征融合层的中间层特征,全面重放过去的经验并使用完整经验重放(FER)进行监督,帮助模型在学习新任务时,保持对旧任务的知识,进一步减少遗忘.

2.3.1 黑暗经验回放策略

日志异常检测的增量更新任务是不断地利用新的时间区间内收集的数据更新模型,可以被看作多个有顺序的分类任务,以 θ 为参数的分类器 f 按照时间顺序进行优化. T 代表历史时间区间,用 $z_\theta(S)$ 表示模型输出的原始类别分数,$f_\theta(S) \triangleq \mathrm{softmax}(z_\theta(S))$ 表示类别的概率分布. 学习目标是在给定当前更新的时间区间 ΔT 时能够对 $t \in \{1,\cdots,T+\Delta T\}$ 的日志序列样本正确分类:

$$\underset{\theta}{\operatorname{argmin}} \sum_{t=1}^{T+\Delta T} \mathcal{L}_t, \quad \text{where } \mathcal{L}_t \triangleq \mathbb{E}_{(x,y)\sim L_t}\left[\mathcal{L}(y, f_\theta(S))\right]. \tag{1}$$

式中,$\mathcal{L}(y, f_\theta(S))$ 是真实标签 y 与模型预测标签的交叉熵损失. 为了尽可能保留模型已经学到的知识,模型需要模仿其对旧样本的原始响应,最小化如下目标:

$$\mathcal{L}_{\Delta T}+\eta\mathbb{E}_{(S)\sim \boldsymbol{L}_T}[D_{KL}(f_{\hat{\theta}}(S)\parallel f_{\theta}(S))]. \tag{2}$$

式中,$\hat{\theta}$ 是 T 时刻更新完成之后的模型参数,η 是平衡损失项的超参数,$(S)\sim \boldsymbol{L}_T$ 代表 S 从旧数据集合 $\boldsymbol{L}_T$ 采样,KL 散度用于衡量原始响应与当前模型预测的差异. 优化上述目标需要已知先前的数据分布 $\boldsymbol{L}_T$. 然而考虑储存空间资源及计算资源的限制及成本,来自以往时间的日志数据不是完全可用的. 因此,本文方法使用 Kmeans 聚类方法提取范例样本作为后续可学习的经验知识,聚类时使用轮廓系数(silhouette score)来确定最佳聚类数目并多次随机初始化聚类中心. 在范例样本的训练集 $\mathcal{M}_T$ 中,DER 保留了样本对应的原始类别分数 $z_{\hat{\theta}}(S)$,这些分数提供了关于数据点的更丰富的信息描述.

$$\mathcal{L}_{\Delta T}+\eta\mathbb{E}_{(S,z)\sim \mathcal{M}_T}[D_{KL}(\mathrm{softmax}(z_{\hat{\theta}}(S))\parallel f_{\theta}(S))]. \tag{3}$$

在温和的假设下,等式(3)中 KL 散度的优化等价于最小化相应的原始类别分数和预测类别分数之间的欧几里德距离. 模型更新时的优化目标最终转化为:

$$\mathcal{L}_{\Delta T}+\eta\mathbb{E}_{(S,z)\sim \mathcal{M}_T}[z_{\theta}(S)-z_{\hat{\theta}}(S)_2^2]. \tag{4}$$

式中,$\hat{\theta}$ 是旧模型参数,θ 是需要更新训练的模型参数. 需要注意的是,在输入流中可能会发生突然的数据分布变化,这会导致范例样本进行重播时受到旧任务中存在高度偏差的样本的影响. DER++通过在目标函数中增加一个额外的项,提升缓冲区数据点相对于其真实标签的条件似然性,从而更好地应对这种分布变化.

$$\mathcal{L}_{\Delta T}+\eta\mathbb{E}_{(S',y',z')\sim \mathcal{M}_T}[\parallel z'_{\theta}(s)-z_{\hat{\theta}}(S')\parallel_2^2]+\zeta\mathbb{E}_{(S'',y'',z'')\sim \mathcal{M}_T}[\ell(y'',f_{\theta}(S''))]. \tag{5}$$

DER++引入了一个额外的系数 ζ,用于平衡目标函数中的最后一项. 当 $\zeta=0$ 时,DER++与 DER 等价. DER++通过优化目标函数,促进模型达到更平坦的最小值. 这种平坦的最小值有助于模型在参数空间中探索邻近区域,提高了模型对局部扰动的容忍度,减少了对知识的遗忘.

2.3.2　完整经验回放

范例数据重播可以帮助模型保持在旧数据上的性能,但 DER 仅仅保留范例样本及对应的原始类别分数,只计算网络最终输出的损失,缺少对中间层特征的监督. 因此,本文在 MLog 的 CNN 特征融合层上进行完整经验回放(full experience replay,FER). CNN 特征融合层能够同时捕获日志序列中的全局依赖性和局部依赖性,提高了对重要特征的关注,这些特征包含了更多的关键知识,能够帮助模型更好地保持对过去任务的记忆. 完整经验回放(FER)通过保留过去样本的完整信息(包括输入、原始类别分数和中间层特征),使得特征融合层在不同层上保持知识的稳定,进一步有效减少知识的遗忘. FER 的完整结构图如图 2 所示.

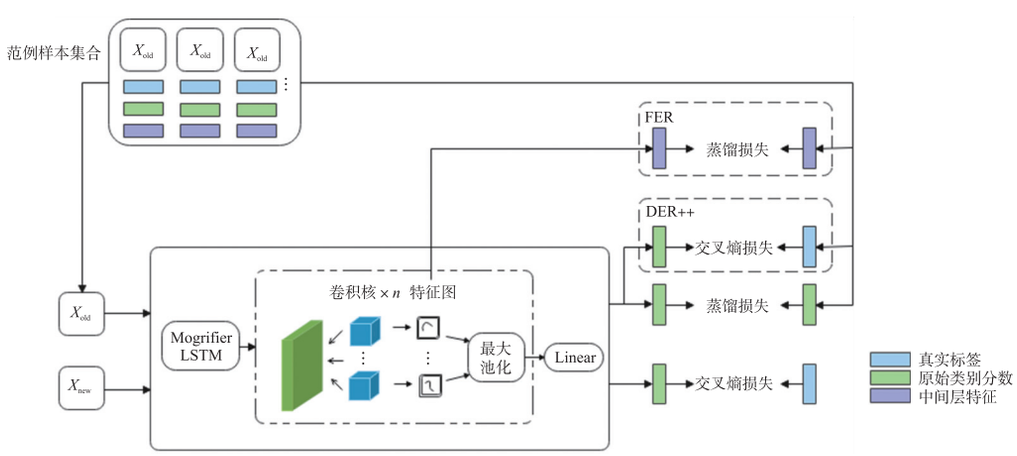

图 2　FER 结构示意图

Fig. 2　Structure of FER

假定重放缓冲区表示为 $\mathcal{M}=\{m[(S_i,y_i)]\}_{i=1}^{M}$,$m[(S_i,y_i)]=\{S_i,y_i,z_{\hat{\theta}}(S_i),\{h_{\hat{\theta},l,c}(S_i)\}\}$,其中 $\hat{\theta}$ 是旧模型参数,$z_{\hat{\theta}}(S_i)$ 和 $h_{\hat{\theta},l,c}(S_i)$ 分别是范例样本的原始类别分数和中间层特征. $\mathcal{L}_{\mathrm{FER}-h}$ 用于计算卷积层中间特征的损失:

$$\mathcal{L}_{\mathrm{FER}-h}=\mathbb{E}_{(S,\{h_{\hat{\theta},l,c}(S)\})\sim \mathcal{M}}\sum_{l}^{L}\sum_{c}^{C_l}\parallel h_{\theta,l,c}(S)-h_{\hat{\theta},l,c}(S)\parallel_2^2. \tag{6}$$

$\mathcal{L}_{\text{FER}-z}$使用当前模型参数 θ 和旧模型参数 $\hat{\theta}$ 对范例样本的响应,计算范例样本的原始类别分数的损失:

$$\mathcal{L}_{\text{FER}-z}=\mathbb{E}_{(S,z_{\hat{\theta}}(S))\sim\mathcal{M}}\parallel z_{\theta}(S)-z_{\hat{\theta}}(S)\parallel_2^2. \tag{7}$$

$\mathcal{L}_{CE-M}$表示缓冲区中的范例样本的标签损失,本文方法使用交叉熵损失衡量 0/1 标签损失:

$$\mathcal{L}_{CE-M}=\mathbb{E}_{(S,y)\sim\mathcal{M}}\ell(f_{\theta}(S),y). \tag{8}$$

使用完整经验回放时,用于每轮更新的范例样本的完整损失函数表示为:

$$\mathcal{L}_{\text{FER}}=\alpha\mathcal{L}_{CE-M}+\beta\mathcal{L}_{\text{FER}-z}+\gamma\mathcal{L}_{\text{FER}-h}. \tag{9}$$

通过以上方式,FER 提供了一种有效的机制来增强模型的学习能力,特别是在面对多轮更新任务时,能够显著提高模型的性能和稳定性.

本文方法通过引入增量学习方法,回放范例样本的原始类别分数和中间层特征,减少了信息在概率空间中的损失,提升了模型的学习效率,增强了模型的适应能力,使其能够在实际应用中保持高效的检测性能,为长期监控和检测提供支持.

算法 1　基于经验回放的日志异常检测模型的更新方法

输入:训练集 $\boldsymbol{L}_T$ 和更新集 $\boldsymbol{L}_{\Delta T}$,已训练模型 f_{θ} 及参数 θ,$\bar{f}_{\theta}$ 代表评估模式,超参数 α,β,γ,学习率 λ

#生成范例样本集 $\mathcal{M}_T$

$\mathcal{M}_T\leftarrow$Kmeans($\boldsymbol{L}_T$)

for(S',y') in $\mathcal{M}_T$:

　　$z',h'\leftarrow\bar{f}_{\theta}(S')$　#获取范例样本的中间特征

　　$\mathcal{M}_T\leftarrow(S',y',z',h')$

#增量更新

for (S,y) in $\boldsymbol{L}_{\Delta T}$ and(S',y',z',h') *in* $\mathcal{M}_T$:

　　#新数据更新

　　$y_{\theta}\leftarrow f_{\theta}(S)$

　　$reg\leftarrow\ell(y_{\theta},y)$　#计算损失

　　$\theta\leftarrow\theta+\lambda*\nabla_{\theta}reg$　#更新模型参数

　　#范例数据回放

　　$y'_{\theta}\leftarrow f_{\theta}(S')$

　　$reg'\leftarrow\alpha\ell(y'_{\theta},y')+\beta\parallel z_{\theta}(S')-z'\parallel_2^2+\gamma\sum\sum\parallel h_{\theta,l,c}(S')-h'\parallel_2^2$　#参照式(9)计算损失

　　$\theta\leftarrow\theta+\lambda*\nabla_{\theta}reg'$

#更新训练集

$\boldsymbol{L}_{T+1}\leftarrow(\mathcal{M}_T,(S',y'))\cup(\boldsymbol{L}_{\Delta T},(S,y))$

end for

3　实验与结果分析

本文利用真实的数据集对增量更新的日志异常检测模型的更新方法进行评估,以验证本文方法的有效性. 首先详细描述实验设置情况,包括数据集、基线方法、实验准备和评估指标等. 其次,实验验证每轮更新对检测效果产生的影响,并设计实验比较不同更新策略及策略组合的效果. 本研究将增量更新的日志异常检测方法与先进的基线方法进行比较,分析增量更新带来的优势与劣势,本文还探究存储空间大小对实验效果的影响.

3.1　实验设置

本文实验使用 LogHub[22] 提供的开源真实世界数据集 BGL,是从位于加利福尼亚州利弗莫尔的劳伦斯利弗莫尔国家实验室的 BlueGene/L 超级计算机系统中收集的开放日志数据集,包含 4 747 963 个日志条目. BGL 为每个日志条目设置了警报标签,在日志的第一列中,“-”表示非警报消息,而其他符号则表示警报消息. 与另一常用的日志数据集 HDFS 相比,BGL 数据集具有更长的时间跨度,且经过对 BGL 数据集进行统计与分析,发现随着时间推移,日志序列数据的分布发生了变化,这与本研究的动机相符合. 序列

生成采用滑动窗口的机制,共生成 16 432 条有效序列. 对于生成序列集合,划分为训练集、验证集和测试集,然后按照时间跨度进一步划分训练集为原始训练集和更新集.

实验采用了 4 个基线方法:PCA、DeepLog[9]、ROEAD[10]和 LogOnline[11]. PCA 通过在事件计数向量中查找模式来生成正常和异常子空间,然后通过计算模板计数向量对异常空间的投影长度来确定日志序列是否异常. DeepLog 利用长短期记忆网络学习日志序列模式,为当前给定序列预测下一个可能的日志事件,当日志模式偏离模型预测时则报告为异常. ROEAD 通过鲁棒的特征提取去除噪声影响,并结合在线演化异常检测(OEAD)动态更新模型参数,以提高异常检测的准确性和效率. LogOnline 通过在线学习机制持续学习新出现的正常模式,其使用日志头信息(时间戳、组件名称和日志级别)训练一个自动编码器,用于筛选可靠的正常序列模式更新检测模型.

基于日志的异常检测的研究普遍采用分类任务中的常用评价指标,即精确率(precision)、召回率(recall)和 F1-score. 计算公式如下:

$$\text{precision}=\frac{TP}{TP+FP}, \tag{15}$$

$$\text{recall}=\frac{TP}{TP+FN}, \tag{16}$$

$$\text{F1score}=2\times\frac{\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}. \tag{17}$$

3.2　实验结果及分析

为验证增量学习方法的有效性,本文从检测性能、时间效率两个方面设计实验,并将本文方法与其他先进的检测方法进行比较. 首先,展示基础方法 MLog 使用增量更新方法学习的检测性能变化. 具体来说,本文从 BGL 数据集中提取 65%作为训练集,5%作为验证集,30%作为测试集,然后将 BGL 训练数据集和验证集按时间进行划分. BGL 数据集的时间跨度为 6 个月,前两个月的数据作为初始训练集,后续每 15 d 产生的数据作为更新训练集,实现多步增量更新实验. 具体来说,首先使用 7—8 月的初始训练集训练得到初始模型,然后从 9 月 16 号开始,模型每半个月更新一次. 超参数取值 $\alpha=1$,$\beta=1$,$\gamma=0.5$,范例样本集大小 $m=600$. 本文将基于经验重放的增量更新方法和全量更新方法学习后的检测效果进行对比,如图 3 分别展示两种更新方法的准确率、召回率和 F1-分数随更新轮次的变化情况,实验结果取 3 次运行的均值.

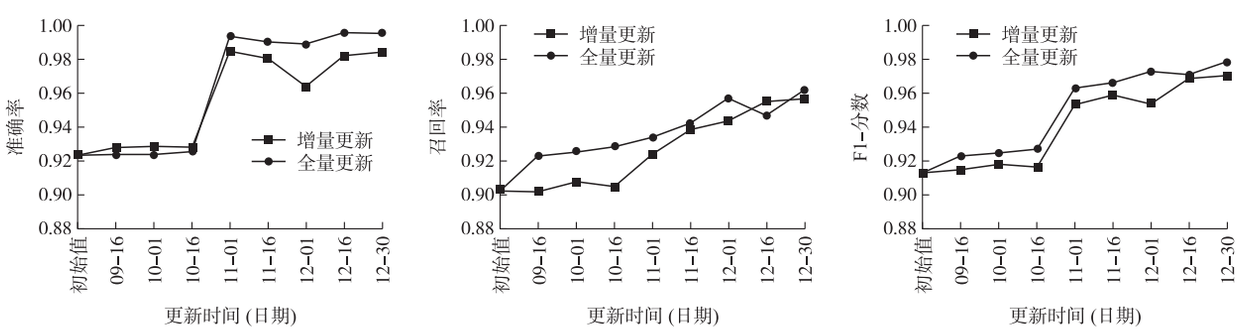

图 3　评价指标随更新轮次的变化

Fig. 3　Variation of evaluation metrics across update rounds

从上述实验结果可以看出,增量更新与全量更新方法的变化趋势总体一致,且指标差距较小,说明仅使用少量数据的增量更新方法能够在保留旧知识的情况下学习到新样本中所包含的信息,达到与全量更新相近的效果. 总体来说,增量更新方法在检测效果上略逊于全量更新方法,因为全量更新能够确保数据完整性和一致性,实验结果符合预期. 然而,全量更新需要耗费更多的时间资源和计算资源. 全量训练和增量训练均采用早停策略,根据损失变化设定训练终止条件.

在每轮更新中,全量方法与增量方法所耗费的时间对比如表 2 所示,全量更新的训练时长随着训练样本增加主要呈现增长趋势,增量更新只使用该时间段内新产生的样本和范例样本进行训练,所用时长约为全量训练的 30%~50%,时间指标提升效果显著.

本文将更新完毕后 BGL 数据集上的最终检测效果与基础方法 MLog 和其他非增量基线、增量基线方法进行对比. 非增量基线在完整的数据集上训练,增量基线 ROEAD 和 LogOnline 按照时间顺序使用训练集前两个月作为初始训练集,剩下的数据进行单步增量更新. 本文方法在进行多轮增量更新后表现略低于全量基线 MLog,但高于全量训练的 PCA、DeepLog 和增量训练的 ROEAD 和 LogOnline,如表 3 所示.

表 2　更新方法用时统计情况

Table 2　Update time comparison

	09/16	10/01	10/16	11/01	11/16	12/01	12/16	12/30
全量更新用时/s	1 077	1 135	1 221	1 135	1 335	1 284	1 465	1 473
增量更新用时/s	581	563	422	486	285	392	331	335

表 3　本文方法与基线方法的效果比较

Table 3　Comparison with baselines

	方法名称	准确率	召回率	F1-分数		方法名称	准确率	召回率	F1-分数
非增量方法	PCA	46.98%	59.75%	52.60%	增量方法	LogOnline	78.14%	80.46%	79.28%
	DeepLog	87.55%	98.91%	88.72%		ROEAD	90.26%	85.11%	86.79%
	MLog	99.66%	96.20%	97.90%		本文方法	98.50%	95.70%	97.07%

本文探究了范例样本集 $\mathcal{M}_T$ 的大小对实验效果的影响. 表 4 展示了在采用不同大小的范例样本集的情况下进行多步增量更新时,F1-分数的变化趋势,实验结果取 3 次运行的均值. 结果显示当 m 较小时,检测结果不理想,随着 m 增大,检测效果越来越好,m 取 600 或 800 时表现相近,最终更新轮次在 $m=600$ 时表现最佳.

表 4　不同大小范例样本集的 F1-分数变化

Table 4　Variation of F1-score with different sizes of example sets

范例样本集大小	09/16	10/01	10/16	11/01	11/16	12/01	12/16	12/30
$m=200$	89.62%	91.44%	90.11%	89.46%	91.36%	92.22%	95.46%	95.14%
$m=400$	90.05%	89.37%	91.74%	92.79%	90.94%	93.22%	96.38%	96.22%
$m=600$	**91.53%**	91.81%	91.65%	**95.39%**	**95.95%**	95.41%	**96.90%**	**97.07%**
$m=800$	91.09%	**92.17%**	**92.73%**	94.93%	94.47%	**96.18%**	96.18%	96.35%

* 字体加粗表示该更新轮次指标的最高值.

此外,本文还实现式(9)中损失项系数 β 和 γ 的敏感性实验,实验结果如表 5 和表 6 所示. 当 $\beta=1.0$,$\gamma=0.5$ 时表现最佳,其他取值存在某些更新轮次表现不佳的情况,如 $\beta=0.5$,$\gamma=0.5$ 时在 10/01 更新时 F1-分数仅有 88.21%. 通过对 BGL 数据集进行分析,其 9 月份仅新增 3 个未见过的日志事件,而 11 月份新增了 68 个未见日志事件,说明 BGL 数据分布的变化程度是不一致的,在选取损失项系数时要注意平衡其在不同更新轮次下的表现.

表 5　F1-分数随 β 取值的变化

Table 5　Variation of F1-score with β

超参数 β	09/16	10/01	10/16	11/01	11/16	12/01	12/16	12/30
$\beta=0.5$	90.45%	88.21%	**91.73%**	**95.43%**	94.76%	95.25%	96.36%	96.03%
$\beta=1.0$	**91.53%**	91.81%	91.65%	95.39%	**95.95%**	**95.41%**	**96.90%**	**97.07%**
$\beta=1.5$	88.01%	**92.42%**	91.53%	95.33%	94.74%	93.88%	96.01%	95.90%

* 字体加粗表示该更新轮次指标的最高值.

表 6　F1-分数随 γ 取值的变化

Table 6　Variation of F1-score with γ

超参数 γ	09/16	10/01	10/16	11/01	11/16	12/01	12/16	12/30
$\gamma=0.2$	90.63%	91.61%	91.18%	94.52%	95.65%	94.22%	96.62%	96.71%
$\gamma=0.5$	**91.53%**	**91.81%**	91.65%	95.39%	**95.95%**	95.41%	**96.90%**	**97.07%**
$\gamma=1.0$	90.08%	87.30%	**91.85%**	**95.46%**	94.93%	**95.55%**	96.11%	96.40%

* 字体加粗表示该更新轮次指标的最高值.

本文还设计了消融实验,比较了不同更新策略在 BGL 数据集上的检测效果,超参数取值 $\eta=\zeta=1$. 图 4 展示了使用 DER、DER++、FER 3 个增量更新方法进行多轮更新实验的 F1-分数,实验结果表明,添加中间层特征约束的 FER 表现优于只使用类别分数约束模型更新的 DER 和 DER++,证明本文方法在特征融合层使用完整经验约束的有效性.

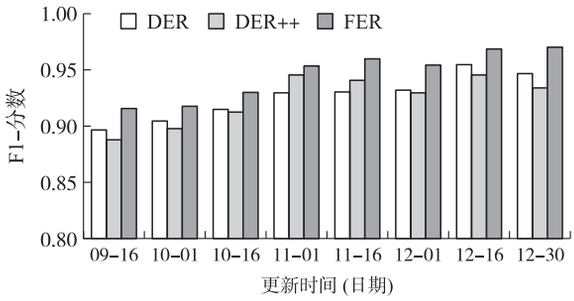

图 4　不同更新策略的比较

Fig. 4　Comparison of different update strategies

4　总结

日志异常检测对于提高复杂软件的可用性和安全性至关重要. 在现有的日志异常检测研究中,模型往往一经确定就不会再进行更新,少数考虑模型更新的方法也仅仅采用判断错误的样本进行更新训练,或采用周期性的更新方式重新训练模型. 然而,在现实应用场景中,日志数据分布可能在较长一段时间内发生变化,现有研究不能很好地适应这些变化,及时做出反馈. 因此,本文提出了基于黑暗经验回放(DER)的增量式日志异常检测模型的更新方法,旨在通过增量学习策略,持续有效地对新数据进行检测,提升检测模型对新出现模式的适应能力. 在新数据到来时,该方法通过使用新数据和旧数据中的提取样本进行模型更新,并在更新过程中运用蒸馏损失以保留已有知识. 进一步,本文方法提取特征融合层的关键特征实现完整经验回放(FER),提高模型的整体学习能力和性能,从而有效应对数据分布的变化和灾难性遗忘现象. 在真实数据及 BGL 上的实验结果表明,本文方法能够在保留原有知识的同时有效捕获新知识,检测效果略低于全量更新基线 MLog,但高于其他基线方法,并大幅减少了训练时间.

[参考文献]

[1] LE V H, ZHANG H. Log-based anomaly detection with deep learning: How far are we? [C]//Proceedings of the 44th International Conference on Software Engineering. Pittsburgh, PA, USA: ICSE, 2022: 1356-1367.

[2] 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.

[3] XIAO T, QUAN Z, WANG Z J, et al. Loader: A log anomaly detector based on transformer[J]. IEEE transactions on services computing, 2023, 16(5): 3479-3492.

[4] ZHANG L, JIA T, JIA M, et al. Multivariate log-based anomaly detection for distributed database[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona, Spain: KDD, 2024: 4256-4267.

[5] HUANG S, LIU Y, FUNG C, et al. Improving log-based anomaly detection by pre-training hierarchical transformers[J]. IEEE transactions on computers, 2023, 72(9): 2656-2667.

[6] QI J, LUAN Z, HUANG S, et al. Logencoder: Log-based contrastive representation learning for anomaly detection[J]. IEEE transactions on network and service management, 2023, 20(2): 1378-1391.

[7] MENG W, LIU Y, ZHU Y, et al. Loganomaly: Unsupervised detection of sequential and quantitative anomalies in unstructured logs[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China: 2019, 19(7): 4739-4745.

[8] 刘春波,梁孟孟,侯晶雯,等. 面向不稳定日志的一致性异常检测方法[J]. 湖南大学学报(自然科学版),2022,49(4):89-99.

[9] DU M, LI F, ZHENG G, et al. Deeplog: Anomaly detection and diagnosis from system logs through deep learning[C]//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. USA: CCS, 2017: 1285-1298.

[10] HAN S, WU Q, ZHANG H, et al. Log-based anomaly detection with robust feature extraction and online learning[J]. IEEE transactions on information forensics and security, 2021, 16: 2300-2311.

[11] WANG X, SONG J, ZHANG X, et al. LogOnline: A semi-supervised log-based anomaly detector aided with online learning mechanism[C]//2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE). Echternach, Luxembourg: IEEE, 2023: 141-152.

[12] 孙文举,李清勇,张靖,等. 基于深度神经网络的增量学习研究综述[J]. 数据分析与知识发现,2025(1):1-30.

[13] BUZZEGA P,BOSCHINI M,PORRELLO A,et al. Dark experience for general continual learning:a strong,simple baseline [J]. Advances in neural information processing systems,2020,33:15920-15930.

[14] YAN Q,GONG D,LIU Y,et al. Learning bayesian sparse networks with full experience replay for continual learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. USA:CVPR,2022:109-118.

[15] LI Z,SHI J,VAN LEEUWEN M. Graph neural networks based log anomaly detection and explanation[C]//Proceedings of the 2024 IEEE/ACM 46th International Conference on Software Engineering:Companion Proceedings. Lisbon,Portugal:ICSE Companion,2024:306-307.

[16] SCHWARZ J,CZARNECKI W,LUKETINA J,et al. Progress & compress:A scalable framework for continual learning[C]// Proceedings of the 35th International Conference on Machine Learning. Stockholm,Sweden:PMLR,2018:4528-4537.

[17] ZENKE F, POOLE B, GANGULI S. Continual learning through synaptic intelligence [C]//Proceedings of the 34th International Conference on Machine Learning. Sydney,Australia:PMLR,2017:3987-3995.

[18] RUSU A A,RABINOWITZ N C,DESJARDINS G,et al. Progressive neural networks[J/OL]. arxiv preprint arxiv:1606.04671,2016.

[19] CHENG D,JI Y,GONG D,et al. Continual all-in-one adverse weather removal with knowledge replay on a unified network structure[J]. IEEE transactions on multimedia,2024(26):8184-8196.

[20] MO J,ZOU R,HUA Y. Multi-level foreground prompt for incremental object detection[J]. IEEE access,2024(113):4048-4066.

[21] FU Y,LIANG K,XU J. MLog:Mogrifier LSTM-based log anomaly detection approach using semantic representation[J]. IEEE transactions on services computing,2023,16(5):3537-3549.

[22] ZHU J,HE S,HE P,et al. Loghub:A large collection of system log datasets for ai-driven log analytics[C]//2023 IEEE 34th International Symposium on Software Reliability Engineering(ISSRE). Florence,Italy:ISSRE,2023:355-366.

[责任编辑:黄　敏]