第XX卷 第XX期 XXXX年XX月XX日
DOI：10.7500/AEPS20240523001
电力系统自动化
Automation of Electric Power Systems
Vol. XX No. XX XX XX, XXXX

基于大语言模型的配电主站日志异常检测

王 申[1]，魏兴慎[2]，朱卫平[3]，朱道华[4]，关志涛[1]
（1. 华北电力大学控制与计算机工程学院，北京市 102206；2. 南京南瑞信息通信科技有限公司，江苏省南京市 211106；
3. 国网江苏省电力有限公司，江苏省南京市 210024；4. 国网江苏省电力有限公司电力科学研究院，江苏省南京市 211103）

摘要： 日志异常检测是监控配电主站系统运行并识别异常行为的关键技术之一。已有的基于深度学习的日志异常检测方法依赖于大量的带标注的训练数据，而在配电主站系统中缺少带标注训练数据，这会导致日志异常检测性能显著下降。文中基于大语言模型的上下文推理特性，提出了一种无需训练的配电主站日志异常检测方案LogAdapt。所设计的上下文学习示例筛选算法针对不同的在线日志，从少量带标注的本地日志中动态筛选出若干高质量的上下文学习示例；结合任务描述和人类经验知识，自动构建出文本提示，以指导大语言模型完成配电主站日志异常检测任务。实验结果表明，所提方案相比现有方案性能更优。

关键词： 日志异常检测；配电主站；大语言模型；提示工程；上下文学习；深度学习

0 引言

配电主站系统是一种用于集中监控和控制配电系统的智能化系统，在配电网络中起着至关重要的作用，是整个配电网的监视、控制和管理中心，能够在配电系统发生故障时，自动判别、隔离故障区段，并恢复非故障区段供电，对提高新型电力系统的可靠性和运行效率具备重要意义[1-3]。配电主站系统接入的网络形式多样、终端数量巨大、跨越多个分区，是电力监控系统网络安全的薄弱点；一旦其遭受网络攻击，可能造成如乌克兰大停电事件的严重事故，对公众安全构成威胁，故保证配电主站系统的安全运行至关重要[4,5]。

日志异常检测是配电主站系统安全防护关键技术之一，可用于监控系统活动和异常行为，有效识别网络攻击并预警[6]。日志中通常记录了系统运行过程中的各种重要信息，包括数据变动、错误和告警信息等，为故障处理和安全分析提供了重要依据。配电主站系统的异常通常由终端或线路故障、网络攻击或软件配置错误造成，通常表现为网络通信、软件运行或文件存储的异常，并在日志中有所体现。然而，随着用电规模的增大，配电业务也日益复杂，其在运行过程中会生成大量的日志，这使得手动分析日志变得不切实际。近些年已经出现了很多基于日志的异常检测方法[7-11]，包括基于机器学习的方法和基于深度学习的方法。在这些方法中，如决策树、主成分分析、不变量挖掘等机器学习方法存在效率低、适应性弱等缺陷[12]。相较之下，基于深度学习的方法，如SwissLog[7]、LogEncoder[8]，展现出了更为优越的性能，能够捕捉更为复杂的日志表征信息。

然而，基于深度学习的日志异常检测方法通常需要大量的域内训练数据，训练数据不足会使其性能显著下降。相比于其他领域，配电主站日志的标注需要投入大量人工成本。配电主站系统涉及SCADA系统、通信协议、智能终端等多种技术和设备的集成，增大了日志复杂性；其特定的业务逻辑和运行规则使得工作人员需要结合电力市场规则甚至物理规律等方面对日志进行标注；同时，与其他领域类似，主站系统的业务需求是不断变化的[13]，新业务出现或旧业务升级时，日志形式也通常会发生变化，进一步增加了数据标注的难度。文献[14]中也指出了日志的不稳定性，提及某软件在多次版本升级后，未发生变化的日志模板仅占30%左右，这为现有日志异常检测方法带来了很大挑战。

近期，有研究开始尝试基于大语言模型（large language model，LLM）的电力系统人工智能应用[15-17]，为解决上述挑战提供了新思路。目前，基于LLM的日志异常检测相关研究较为有限。LogPrompt[18]和LogGPT[19]两项相关研究仅采用了相对初级的提示设计，并未能充分释放LLM强大的潜能，这也导致其在日志异常检测应用上的性能表

收稿日期：2024-05-23；修回日期：2024-10-25。
上网日期：XXXX-XX-XX。
国家电网有限公司科技项目（5400-202340217A-1-1-ZN）。

现并不理想。本文提出了一种基于自适应提示学习的配电主站日志异常检测方案LogAdapt，充分利用了LLM强大的上下文理解与推理能力，相比已有方法检测性能更优。

LogAdapt主要包含两个模块：日志预处理和提示构建。在日志预处理模块中，通过日志解析、日志分组等步骤，将半结构化的日志转化为更易处理的结构化日志；在提示构建模块中，本文设计了一套的上下文学习（in-context learning，ICL）示例动态筛选策略，以提供高质量的ICL示例，然后结合任务描述和人类经验知识，自动为不同日志消息定制文本提示，以更好地指导LLM完成日志异常检测任务。

1）无需训练，针对配电主站系统中缺少带标注训练数据问题，提出了一种适用于训练数据稀缺场景的配电主站日志异常检测方案。

2）通过分析LLM的上下文推理特性，设计了一套自适应提示学习方法，为LLM动态筛选高质量、有针对性的ICL示例，以优化其性能表现。

3）实验结果表明，尽管没有训练过程，LogAdapt的检测性能仍优于现有方法，对提升配电主站安全具有实用价值。

1 配电主站日志异常检测概述

1.1 配电主站日志异常检测流程

配电主站系统的日志异常检测包含日志解析、特征表示、异常检测3个步骤，如图1所示。

1.1.1 日志解析

配电主站等系统产生的日志是一种半结构化数据，包含了多种类型的信息如日期、来源、级别等，为了更好地分析日志，通常需要先对日志进行解析，提取出日志模板等信息并生成结构化日志。例如：从日志消息“Oct 1 10：47：44 tsrtdb2 kernel：[85317.768290]transport_model[1463]：segfault at 7f4bf44f72cc ip 00007f4bf44f72cc sp 00007ffc5e023d98 error 15 in libclntsh. so [7f4bf414d000+3ab000]”提取出日志模板“transport_model[＜*＞]：segfault at ＜*＞ ip ＜*＞ sp ＜*＞ error ＜*＞ in libclntsh.so[＜*＞]”，其中‘＜*＞’代表日志中的变量。日志模板作为开发者手工设计的系统状态语义描述，包含了诸多关键信息，通常会作为日志异常检测的重要分析对象。更多的日志解析细节见附录A图A1。

常见的日志解析方法可以分为频繁模式挖掘、聚类、启发式等几种类型。频繁模式挖掘方法通过识别日志数据中的频繁词串，进而提取出潜在的模板；聚类方法通过将相似的日志分到相同的类别来提取日志模板；而启发式方法则利用专家经验，根据日志的独特结构和词汇特点设计出更为匹配的解析算法，例如Drain[20]使用固定深度的树结构，通过逐层划分日志条目中的字段，进而快速准确地提取日志模板。其中，启发式方法由于充分利用了日志数据的结构相对稳定性、词汇有限性等特点，往往具备更优的性能[21]。

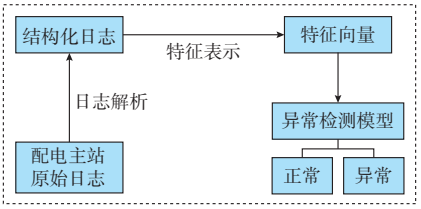


图1 配电主站日志异常检测过程

Fig. 1 Process of log abnormity detection in distribution master station logs

1.1.2 特征表示

日志模板需要转化为模型能够识别的特征向量才能够实现异常模式的判别。在进行特征表示之前，需要首先对日志进行分组，将日志消息切片形成具备时间维度的日志序列，再将日志序列转化为特征向量。常见的特征表示形式有顺序向量、计数向量和语义向量，分别采用日志的顺序、频数和语义信息来表征日志。其中顺序向量可以反映出上下文信息，计数向量便于展示日志分布，语义向量更能凸显日志的语义特性。

1.1.3 异常检测

该步骤主要目的是利用提取出来的特征表示训练得到一个日志异常检测模型，已经有很多深度学习模型如循环神经网络（recurrent neural network，RNN）、卷积神经网络（convolutional neural networks，CNN）、Transformer[22]得到了应用。例如，Deeplog[9]和LogAnomaly[10]等使用RNN的变体长短期记忆（long short-term memory，LSTM）神经网络来预测下一个日志事件；文献[23]使用CNN模型学习系统日志中的事件关系进而检测异常；AllInfoLog[24]和NeuralLog[11]等基于Transformer的编码器来进行日志异常判别。

1.2 配电主站系统日志异常检测

针对训练数据稀缺场景下配电主站日志异常检测面临的挑战，LLM展现出了良好的应用潜力。LLM是指以Transformer为基础架构的大规模预训练语言模型，其参数量高达数十亿乃至数万亿，具备强大的自然语言理解和生成能力。LLM的发展历程可以追溯到Transformer的提出，引起了BERT[25]等预训练模型的蓬勃发展，人们无须从头训练模型，

只需在较小的下游数据集上微调，便可以得到符合预期的模型。而随着GPT-3[26]的提出，人们发现无须训练，只需要特定的文本提示，LLM便可以充分利用预训练过程中学到的知识完成相应下游任务。

LLM涌现出的一个重要的能力便是ICL，无需更新参数，只须在上下文中给出一些解决特定任务的示例，LLM便可以从示例中类比学习到特定任务的解决方法，而其无须训练的特性也为解决训练数据稀缺场景下的配电主站系统日志异常行为检测提供了可能。然而，多项研究表明LLM在ICL上的性能很大程度上依赖于ICL示例的选择[27,28]，不合适的示例甚至会降低LLM的表现。本文通过设计一套自适应ICL示例筛选方法，解决了这一问题，实现了LLM在日志异常检测任务中的应用。

2　基于LLM的日志异常检测方案

本文提出了一种基于ICL范式的LLM自适应提示学习方法，并基于此形成了一套配电主站日志异常检测方案LogAdapt。首先，使用日志解析器将来自配电主站系统的原始日志解析为结构化日志，然后通过聚类的方法从本地日志采样出一个多样化的日志候选集。当给定一个在线日志消息作为查询的时候，LogAdapt会从日志候选集合中筛选出N个相似度最高的样本并升序排列，然后将这些样本按照固定的格式构建形成ICL示例，最后将任务描述、经验知识、ICL示例与在线日志按照顺序排列形成最终的提示并送入到LLM中，完成在线日志异常模式的判别。LogAdapt的基本工作流程如图2所示。

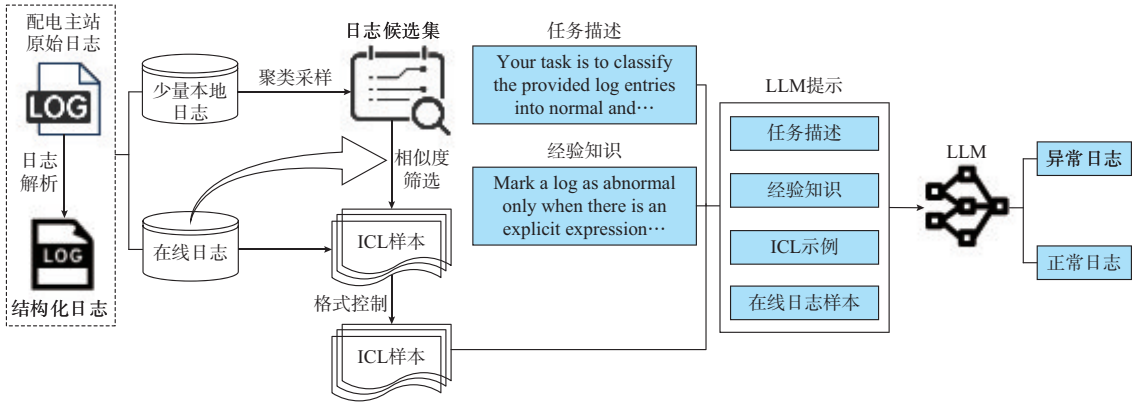


图2　LogAdapt基本工作流程
Fig. 2　Basic workflow of LogAdapt

2.1　日志预处理

日志预处理的主要目的是将半结构化的原始日志转化为结构化数据。首先，使用启发式的日志解析方法Drain[20]将来自配电主站系统的原始日志解析为包含日志模板、日志变量、时间戳等部分的结构化日志。然后，对日志消息按照时间进行排序，使用固定窗口的分组方法将连续的日志消息切片为具备时间维度信息的日志序列。

2.2　提示构建

日志异常检测文本提示主要包含任务描述、经验知识、ICL示例、在线日志4个部分，如图3所示。本文根据经验手工设定了任务描述和经验知识两部分的内容，并在ChatGPT的帮助下进行了进一步优化；在ICL示例部分设计了一套筛选算法，根据当前正在查询的在线日志动态的选取最为合适的ICL示例。

2.2.1　任务描述和经验知识的注入

LLM在预训练阶段获取了强大的自然语言理解和生成能力，通过提供明确的任务描述，可以为LLM提供准确的指导，使其更有效地理解和执行特定任务。此外，注入人类在配电主站系统中的异常判别知识，为LLM提供了实质性的经验支持，进一步提高了其在特定任务上的执行成功率。具体来说，经验知识来源于工作人员积累起来的一些具备较高通用性的异常判别经验，一般为“在某情况下判定某日志为异常/正常”的形式，比如日志解析通常

提示的组成：<任务描述> <经验知识> <ICL示例> <在线日志>

<Your task is to classify the provided log entries into normal and abnormal categories. Please follow these steps for classif ication:>< (a) Mark a log as normal when it contains invalid values, such as memory addresses, floating numbers, or register values. (b) Mark a log as normal when it lacks essential information. (c) Do not consider placeholders denoted as or missing values as abnormal patterns. (d) Mark a log as abnormal only when there is an explicit expression of an alert in the textual content, such as keywords like “error” or “interrupt” .> < Here are some examples: log entries: {ICL example}. Response: {'abnormal' / 'normal'} …> <log entries: {logs}>

图3　用于日志异常检测的文本提示
Fig. 3　Text prompts for log abnormity detection

会将日志中的变量替换为占位符‘<*>’，这是一种正常现象，因而不可以将其判定为异常模式，又比如日志中存在的较长的内存地址字符串、浮点数也是正常现象等等。基于此，本文手工设定了多套文本提示，并在一个小的验证集上进行了测试，保留性能最优的任务描述、经验知识，作为不同在线日志定制化提示中的固定成分，如图3中的橙色和蓝色部分文本字符所示，并与动态构建的ICL示例、在线日志共同组成LLM的输入文本提示。

2.2.2 ICL示例的构建

ICL示例的构建过程主要分为候选集采样、示例样本筛选、格式控制3个步骤。这3个步骤分别旨在确保示例候选样本的多样性、与查询样本在语义上的相似性和最终任务答案的可提取性，具体过程如下：

首先进行候选集的采样，即：从本地日志中采样出更具有代表性的样本，以达到去除干扰、平衡数据分布的作用，从而得到一个多样化的ICL示例的候选集合，为待查询的在线日志提供更为均衡的判别信息，进而大幅减少大语言模型在ICL推理中出现归纳偏差的风险。采样细节详见附录B表B1。

具体而言，首先使用基于Transformer的预训练语言模型BERT（bidirectional encoder representations from Transformers）将本地日志X向量化，取BERT模型最后一层隐藏层的输出作为其特征表示V。然后使用基于聚类的采样算法（如K-means算法），根据本地日志的特征表示将其划分为K个簇。接下来，从每个簇中随机选取一个样本代表该簇，从而形成一个包含K个样本的候选集C。通过这种采样方法得到的候选集的每个样本都属于不同的簇，从而具备了较好的多样性。

然后进行示例样本的筛选，即：从候选集中筛选出与待查询样本在语义上的相似性最高的n个样本，筛选细节详见附录B表B2。

同样的，首先使用BERT提取待查询日志x_q和候选集C的特征表示，然后计算x_q和候选集C中每一条日志的余弦相似度，选择相似度最高的n个样本作为示例样本。受到文献[27]的启发，本文将n个示例样本按照相似度进行升序排列，把相似度较高的样例放在更靠近查询样例的位置，这样的排列有助于LLM更有效地捕捉到相似上下文，进而提升其在上下文推理方面的性能。余弦相似度D的计算方法如式(1)所示。

$$D=\frac{\boldsymbol{v}_i\boldsymbol{v}_j}{\|\boldsymbol{v}_i\|_2\|\boldsymbol{v}_j\|_2} \tag{1}$$

式中：v_i和v_j分别代表样本i和样本j对应的特征向量。通过这种示例筛选算法，可以筛选到与查询样本最相似的数个高质量示例样本，让LLM更好地把握查询样本的语义、语境和语言特征，进而生成更准确的答案。

最后进行格式控制，将筛选得到的示例样本构建为特定的格式。注意到LLM经常会输出一些重复内容或不相关信息，对LLM的预测准确率造成了极大影响。LogAdapt通过将ICL示例构建为特定的格式，进而使得LLM从示例中类比学习到相关信息，并按照特定的格式生成文本内容。这样的格式控制方法也可以更容易地从LLM的输出中提取异常判别的答案，例如，对一个标签为异常的示例样本example，其格式化后的样式为：“log entries:{example}, Response:{abnormal}”。这样通过简单的正则表达式便可以在LLM的输出结果中抽取出查询日志的预测结果，同时单个token的输出也极大地提升了LLM的预测效率，从而显著提升了本方案在配电主站日志异常检测领域的可用性。

LogAdapt方案中LLM处理提示的过程如图4所示。

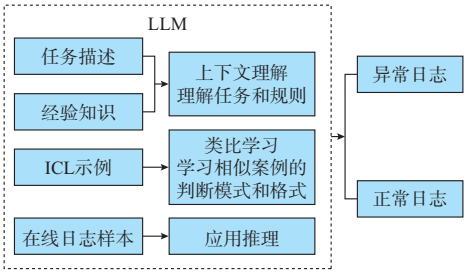


图4 LLM提示流程

Fig. 4 Process of LLM prompting

整个过程可分为3个主要步骤：上下文理解、类比学习和应用推理。在上下文理解阶段，LLM通过分析任务描述和经验知识，明确任务目标和注意事项。在类比学习阶段，LLM利用ICL示例学习相似日志的异常判断过程和输出格式。最后，在应用推理阶段，LLM将前两个阶段获得的知识应用于新的在线日志样本，进行实际的异常检测。这种设计使LLM能够灵活高效地将临时学到的相关知识应用于新的日志样本，实现更为准确的异常检测。

3 验证与应用

3.1 算例概况

为验证LogAdapt方案的有效性，本文选取3个具有代表性的数据集进行实验分析。其中，包括两个公开数据集BGL[29]和Spirit[30]，以及一个自主构

建的Electricbird数据集。BGL和Spirit可以反映通用系统的日志特征，提供日志异常检测任务基准。Electricbird数据集则专门为模拟配电主站场景而构建，是本文研究的核心数据集。

1）BGL数据集采集于Blue Gene/L超级计算机，包含4 747 963条日志消息，涵盖了从静态随机存取存储器（static random-access memory，SRAM）芯片奇偶校验错误到风扇故障等各种软硬件错误，由系统管理控制软件MMCS生成，有着严格的结构和格式，其中348 460条日志信息（7.34%）被标记为异常。

2）Spirit数据集采集于名为Spirit的Linux集群系统，包含超过1.72亿条异常日志消息，主要为磁盘相关的错误，如EXT3文件系统错误，由syslog-ng软件收集，数据量庞大但是有大量重复和冗余的信息。为确保研究的严谨性，与文献[31]保持一致，本文选取前500万行日志作为研究使用的数据集。其中764 500条为异常日志消息，占总数的15.29%。

Electricbird数据集采集于某配电主站系统，主要为网络通信相关的异常，同时包含一些文件存储、终端以及应用软件相关的错误。同时，主站系统接入的网络以及终端设备的多样化也导致了该数据集存在着分布不均衡的问题。鉴于实际日志标注的困难性，Electricbird融入了大量格式和内容高度吻合的Thunderbird[29]数据，以确保数据的规模和日志异常情况的多样性，数据集共包含1 000万条日志消息，其中4 934条为异常日志消息，占总数的0.049%。

结合以上3个数据集对LogAdapt进行评估，既满足了实际配电主站的具体需求，又确保了数据特征的丰富性，可以更全面地评估配电主站系统日志异常检测研究的适用性，并为后续LogAdapt推广应用到更多配电主站场景提供基础。

3.2　实验设置

3.2.1　运行环境

本文所有实验使用Python 3.8.11编写，深度学习框架选用PyTorch 1.13.1，CUDA版本为11.4。实验平台操作系统为Windows 10，使用3.70 GHz的Intel Core i9 10900K CPU和NVIDIA GeForce RTX 3090 GPU加速基准方法的训练和LLM的推理。

3.2.2　模型介绍

Mistral-7B[32]模型是由Mistral AI发布的一个拥有73亿参数的开源LLM，使用组查询注意力机制（grouped query attention，GQA）和滑动窗口注意力机制（sliding window attention，SWA）等技术，拥有更快的推理速度和更低的计算成本。并且Mistral-7B具有较为宽松的开源协议，允许商业使用、修改和分发，这也使其在配电主站的应用成为了可能。

3.2.3　基准方法

Deeplog[9]是一种半监督方法，其利用LSTM模型，从事先给定的事件序列预测下一个日志事件，从而学习日志序列的正常模式。该方法通过判断传入的日志事件是否与LSTM模型的预测结果一致来进行异常检测。

LogAnomaly[10]也是一种半监督方法，其采用日志计数向量和template2vec语义向量作为输入进行LSTM模型的训练。与DeepLog相似，模型被设计用于预测下一个日志事件，如果被检测的日志事件与预测结果不符，则标记为异常。

NeuralLog[11]是一种有监督异常检测方法，该方法借助BERT模型对原始日志进行特征表示，以规避预处理可能导致的信息损失，然后使用Transformer的编码器部分作为异常判别模型，展现出了卓越的日志异常检测性能。

LoRA[33]是一种高效的LLM微调方法，其通过添加低秩矩阵的方法来实现参数高效微调。本研究采用该方法训练LLM进行日志异常检测，作为探索LLM在此任务中潜力的基准方法之一。

LogPrompt[18]是一种基于LLM的日志异常检测方法。该方法使用自生成提示、思维链提示和上下文提示3种提示策略，实现了零样本和少样本学习设置下的日志异常检测。

LogGPT[19]提出了一个基于ChatGPT的日志异常检测框架。该方法结合了知识注入、上下文提示和隐式思维链策略，在零样本和少样本学习设置下提升了异常检测性能。

3.2.4　评估指标

与已有研究保持一致[11,31]，本文使用精确率P，召回率R，F1分数F这3个评价指标来评估日志异常检测的性能。

$$P=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{P}}} \tag{2}$$

$$R=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}} \tag{3}$$

$$F=\frac{2PR}{P+R} \tag{4}$$

式中：T_{P}为预测正确的正（异常）样本数；F_{P}为预测错误的正样本数；F_{N}为预测错误的负样本数。

3.2.5 实验配置细节

1)数据预处理

首先,使用Drain日志解析方法完成日志解析;然后,将BGL、Spirit、Electricbird 3个数据集分别按照时间顺序进行排列;最后,使用窗口为20的固定窗口分组方法将数据集分割成日志序列。

2)实验场景设置

为模拟训练数据稀缺的场景,本文从每个数据集中选取前4 000个日志序列(8万条日志)模拟本地日志,同时也作为涉及训练基准方法的训练集。在余下的数据集中随机选取2 000个日志序列(4万条日志)模拟在线日志,作为日志异常检测的测试集。

3)LogAdapt的特殊配置

LogAdapt在本地日志中采样K个样本作为示例候选集。K值通过网格搜索方法在一个小的验证集上确定,在BGL、Spirit、Electricbird 3个数据集上的值分别为100、470、400。从示例候选集中为每一个在线日志选取N个ICL示例样本(主要分析了N=5和N=7两种情况),并与任务描述、经验知识、在线样本组装成文本提示输入到LLM中。

4)基准方法相关配置

LogPrompt方法的思维链提示和ICL示例数皆取原文最优值,其中ICL示例通过随机选取构建(N=20)。LogGPT的ICL示例通过人工构建(N=5),隐式思维链则配置为LLM判断的同时输出思考过程。为公平起见,LogPrompt和LogGPT其余配置皆与LogAdapt保持一致。

3.3 仿真验证

本节基于BGL和Spirit两个公开数据集进行仿真验证,以评估LogAdapt的有效性,实验结果见表1。

表1 LogAdapt与基准方法的性能对比分析

Table 1 LogAdapt vs. Baseline Methods: Performance Comparison Analysis

模型	是否训练	方法	BGL数据集			Spirit数据集			Electricbird数据集		
			F	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*
DNN	√	DeepLog	37.21	22.87	99.78	41.49	26.18	100	50.71	33.97	100
	√	LogAnomaly	38.53	23.87	100	38.93	24.17	100	42.81	27.24	100
	√	NeuralLog	78.62	80.58	76.75	65.30	48.60	99.50	26.99	53.93	18.00
LLM	√	LoRA	28.69	100	16.75	97.90	96.59	99.26	49.05	100	32.50
	×	$\text{LogPrompt}_{N=20}$	74.44	60.89	95.75	51.31	34.51	100	52.49	36.12	96.00
	×	$\text{LogGPT}_{N=5}$	76.21	65.82	90.50	56.63	39.62	99.25	59.65	46.56	83.00
	×	$\text{LogAdapt}_{N=5}$	82.04	76.46	88.50	92.34	86.15	99.50	84.14	77.87	91.50
	×	$\text{LogAdapt}_{N=7}$	82.50	80.82	84.25	95.04	92.04	98.25	81.11	73.00	91.25

在BGL数据集上,LogAdapt展现出了优异的性能。N=5和N=7两种情况下LogAdapt的F1分数分别达到82.04%和82.5%,明显优于其他方法。值得注意的是,监督学习方法NeuralLog在BGL数据集上也具备相对优异的性能,F1分数达到78.62%,仅次于LogAdapt。然而,半监督学习方法如DeepLog和LogAnomaly虽然具备很高的召回率,但其精确度却非常低,这意味着在训练数据稀缺的情况下,半监督方法容易产生大量假阳性,实际场景中很可能会造成操作员的警报疲劳。

在Spirit数据集上,LogAdapt展现出了更为优异的性能,N=5和N=7两种情况下LogAdapt的F1分数分别高达92.34和95.04%。值得注意的是,需要训练的LLM方法LoRA在Spirit数据集上表现优异,F1分数达到97.90%,略高于LogAdapt。经分析这是因为Spirit数据集主要包含磁盘相关的错误,异常式相对单一,LLM可以在相对较少的训练样本上快速学习到这些模式。然而,LoRA在BGL数据集上的表现却相对较差(F1分数仅为28.69%),反映出其在面对更复杂、多样化的异常类型时的不稳定性。

无须训练的LLM方法LogPrompt和LogGPT在两个数据集上都展现出一定潜力,但整体性能不及LogAdapt,仍有较大的提升空间。这是因为它们仅使用了简单的ICL示例构建策略,并未能充分释放LLM强大的潜能。同时,观察到LogGPT的性能略优于LogPrompt,证明了人工构建的ICL提示相比随机选取的ICL示例具备一定优势。

综合BGL和Spirit数据集的仿真验证结果可知,LogAdapt展现出了优异的性能和稳定性,不仅可以在不同类型和复杂度的日志异常检测任务中保持高性能,还一定程度上克服了传统方法和其他LLM方法的一些局限性,如数据依赖、计算资源消耗和性能不稳定等问题。这为LogAdapt在实际配

电主站环境中的应用奠定了坚实的理论基础。

3.4　工程验证

本节将LogAdapt应用于采集自某实际配电主站现场的Electricbird数据集，以评估LogAdapt在实际配电主站环境中的应用效果。同时，也部署了基准方法用以性能对比。实验结果见表1。从表1可以看出，LogAdapt的性能表现尤为突出，相比其他方法，实现了至少41%的性能提升（从59.65%提升到84.14%）。这证明了LogAdapt在处理实际配电主站数据时的卓越能力。同时，LogAdapt在精确度和召回率方面也达到了很好的平衡，这对于实际应用中减少误报和漏报至关重要。

而其他方法则普遍表现不佳，例如，监督学习方法NeuralLog的F1分数仅为26.99%，远低于其在BGL和Spirit数据集上的表现，反映出该方法对于数据集不同类别的分布具有很高的敏感性，而Electricbird数据集中异常日志仅占0.049%，使得该方法出现了欠拟合问题。

LLM方法在Electricbird数据集上的表现各异。需要训练的LoRA方法虽然精确度达到100%，但召回率仅为32.50%，导致F1分数只有49.05%，再次表明了LoRA在面对复杂异常模式时的局限性。同时也暴露了LLM训练过程中的一些难点，如需要充足的训练数据、对数据特征敏感、训练呈现出的不稳定性等，并且LLM庞大的参数量决定了模型更新、重训练的过程面临着更高的计算资源消耗。

无需训练的LLM方法LogPrompt和LogGPT表现相对较好，F1分数分别达到52.49%和59.65%，但仍明显落后于LogAdapt。同时，由于LogAdapt只需要输出单个token，其在推理速度上也表现出明显优势：经统计，LogAdapt推理速度比LogGPT快约5倍，比LogPrompt快约2倍。

LogAdapt在Electricbird数据集上的出色表现可归因于其几个关键优势：首先，它无须训练即可达到高性能，避免了在数据稀缺情况下的欠拟合或过拟合问题。其次，动态选择最优的ICL示例的能力使其能更好地适应不同类型的日志数据，这在复杂的实际环境中尤为重要。此外，LogAdapt的推理速度也具备优势，非常适合实时分析的需求，这对于配电主站的实时监控和快速响应至关重要。

在实际配电主站的工程验证中，LogAdapt的这些特性有望为配电主站的运维效率、可靠性和安全性带来的显著提升。高准确率和低误报率可以帮助运维人员更精准地识别异常情况，减少不必要的警报和检查。快速的推理能力则使系统能够及时发现和响应潜在的问题，提高整体运维效率。此外，LogAdapt的无需训练特性使其易于部署、更新和推广，可以快速适应不同配电主站环境的变化。

3.5　影响因素分析

3.5.1　消融实验

为了深入分析LogAdapt的有效性，本文在N=5的情况下进行了消融实验，以检验其不同组成部分的贡献。具体来说，首先分别去除文本提示中的任务描述、经验知识、ICL示例进行实验，分别记为“Non-描述”、“Non-经验”和“Non-示例”，以探究任务描述、经验知识、ICL示例3个模块的作用；然后针对ICL模块，去除聚类采样算法模块，直接从本地日志中执行筛选相似日志，记为“Non-采样”，将示例筛选算法替换为随机筛选，记为“Non-筛选”，以分别探究算法1、算法2的贡献，为了更直观的展现出两个算法的效用，本文设定了一组采用随机示例作为ICL示例的实验，记为“随机示例”。实验结果如表2所示。

表2　消融实验
Table 2　Ablation experiment

方法	BGL数据集			Spirit数据集			Electricbird数据集		
	F	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*
Non-描述	83.19	79.50	87.25	90.66	83.26	99.50	86.78	81.54	92.75
Non-经验	74.19	60.06	97.00	81.10	68.32	99.75	70.01	55.09	96.00
Non-示例	47.3	31.57	94.25	47.29	31.04	99.25	46.26	31.16	89.75
随机示例	51.12	34.49	99.50	43.89	28.40	96.50	39.38	24.65	97.75
Non-采样	42.09	54.58	34.25	92.59	87.36	98.50	20.07	31.38	14.75
Non-筛选	59.51	42.78	97.75	45.33	29.30	100	55.93	39.54	95.5
LogAdapt	82.04	76.46	88.50	92.34	86.15	99.50	84.14	77.87	91.5

通过对比“Non-描述”、“Non-经验”和“Non-示例”3组实验，可以发现，去除经验知识和ICL示例都会使LogAdapt的性能有较为明显的降低，并且去除ICL示例会对实验结果造成更大的影响，由此可

验证，经验知识和ICL示例均能提升LLM解决日志异常检测任务的能力，其中，ICL示例起着更为关键的作用。而任务描述对性能提升作用并不明显，本文推测这是因为部分情况下任务描述作用会被经验知识和ICL示例覆盖。方案中保留该部分的原因在于任务描述的存在可以使方案应对更极端的场景，增强LogAdapt在实际场景中的通用性。

通过对比“随机示例”、“Non-采样“、”Non-筛选”和LogAdapt四组实验，可以发现，逐一剔除聚类采样算法和示例筛选算法，或同时剔除两种算法，都会显著降低LogAdapt的性能。由此可验证，保证ICL候选示例的多样性和ICL示例与待查询日志的相似性，都有助于提供更高质量的ICL示例，从而进一步提升LLM在解决日志异常检测任务时的能力。同时，注意到去除算法1后，LogAdapt在3个数据集上的表现差异较为明显，其中在BGL和Electricbird两个数据集上，尤其是在Electricbird数据集上性能有着显著的下降，而在Spirit数据集上其性能保持基本不变，这与3个数据集的分布特性相呼应：Electricbird数据集的数据分布不均衡，省略算法1之后，算法2很大概率会筛选到单一形式的ICL示例，从而限制了LLM可从ICL示例中学到的知识，若查询日志和ICL示例形式不一致，便很可能导致大模型推理错误。而Spirit数据集由于日志形式本身就很单一，所以即便ICL示例形式单一，也不会对LLM推理产生有太大影响。

3.5.2 示例数量的影响

本节探究了示例数量对实验结果的影响，具体而言，分别将LogAdapt的示例个数N设置为1、3、5、7、9，并执行相应的实验，实验结果如图5所示。

通过观察图5，可以发现随着示例数量的增加示例提供的信息也更加丰富，F1、精确度和召回率三个指标在整体上也呈现出了一种增长的趋势。然而，值得注意的是，当示例数量超过7时，LogAdapt在Spirit数据集的性能表现反而有了下降的趋势。结合LLM上下文推理的类比学习的核心思想，推测这是因为过多的示例条目，可能会引入一些弱相关甚至不相关信息，对最终的推理造成干扰。这也启示工作人员在实际应用中不能盲目地增加ICL示例的数量，高质量的ICL示例才是决定任务执行效果的关键因素。

更多关于具体技术选择的详细分析见附录C，探讨了不同聚类采样算法（如K-means、谱聚类、GMM等）对示例候选集构建的影响和不同词嵌入方法（如TF-IDF、GloVe、BERT等）对日志特征表示的影响。这些分析结果表明，基于距离的聚类方法和基于预训练的词嵌入模型在本方案中具有更好的性能表现。

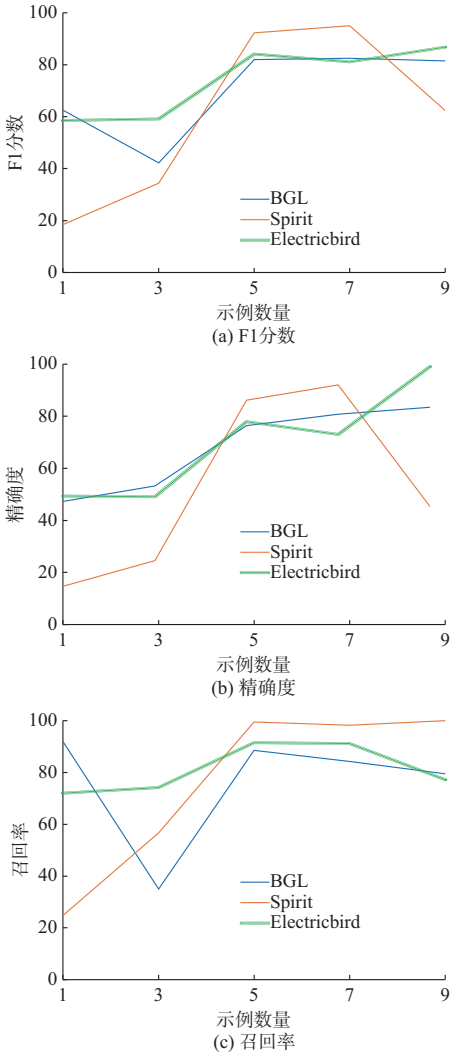


图5 不同示例数量对实验结果的影响

Fig. 5 Impact of different sample sizes on experimental results

4 结语

作为监控软件系统运行并识别异常行为的关键技术，日志异常检测的性能表现高度依赖于大量域内训练数据。然而，在现实场景中，配电主站系统的需求和运维业务是不断变化的，系统的升级导致了日志形式的不稳定性，也使得域内训练数据变得稀缺，传统异常检测方法的实际性能大幅下降。为解决这一挑战，本文设计并实现了一种基于自适应提示学习的配电主站日志异常检测方案LogAdapt，自动为每一条测试日志定制文本提示，精准引导LLM解决下游任务，在训练数据稀缺的情景下实现了高效的配电主站日志异常检测。在多个数据集上的实验测试结果显示，LogAdapt相比当下主流的日志异

常检测方法具备更为优异的性能，展现出了更高的灵活性和泛化性。这些发现表明，LogAdapt的引入有望对提升配电主站系统的稳定性和安全性产生积极影响。

然而，借助LLM解决配电主站系统日志异常检测的问题也存在着较多局限性，如计算资源消耗较大、不可解释性、幻觉问题、LLM能力受限于数据质量等问题。针对这些问题，未来的研究可以从以下几个方面展开：

1）可解释性增强：探索结合检索增强生成（RAG）技术，使模型能够基于可靠的外部知识库进行推理，提高输出的可解释性。

2）数据合成：利用生成式模型创建大规模的模拟数据，并结合小模型过滤掉低质量数据，以提供充足的优质训练数据。

3）模型优化：考虑蒸馏技术，将大型语言模型的知识转移到较小的模型中，以降低计算资源消耗，提高实际应用中的效率。

4）框架改进：探索Agent框架，将日志异常检测的各个步骤工具化（如日志解析工具、提示构建工具等）。通过Agent的规划与工具调用行为，使得LLM解决问题的过程更加透明、可控，并进一步提升日志异常检测性能。

下一步将继续深入研究LLM的特性，并对LogAdapt进行改进，以应对更加复杂和多样化的业务场景，进一步提升配电主站系统的安全性和可靠性。

本文算法代码及数据集已共享，可在本刊网站支撑数据处下载（http://www.aepsinfo.com/aeps/article/abstract/20240523001）。

附录见本刊网络版（http://www.aeps-info.com/aeps/ch/index.aspx），扫英文摘要后二维码可以阅读网络全文。

参 考 文 献

[1] 刘健，张志华，陈宜凯，等．适用于含DG配电网故障处理性能测试的主站注入测试技术[J]．电力系统自动化，2017，41(13)：119-124.
LIU Jian，ZHANG Zhihua，CHEN Yikai，et al. Host injection test technology for fault handling performance test of power distribution network with DG[J]. Automation of Electric Power Systems，2017，41(13)：119-124.

[2] 范闻博，关石磊，符金伟，等．配电自动化潮流计算测试平台设计[J]．电力系统保护与控制，2018，46(8)：117-123.
FAN Wenbo，GUAN Shilei，FU Jinwei，et al. Test bench design for distribution automation power flow calculation[J]. Power System Protection and Control，2018，46(8)：117-123.

[3] 王宗耀，苏浩益．配网自动化系统可靠性成本效益分析[J]．电力系统保护与控制，2014，42(6)：98-103.
WANG Zongyao，SU Haoyi. Cost-benefit analysis model for reliability of distribution network automation system[J]. Power System Protection and Control，2014，42(6)：98-103.

[4] 周劼英，张晓，邵立嵩，等．新型电力系统网络安全防护挑战与展望[J]．电力系统自动化，2023，47(8)：15-24.
ZHOU Jieying，ZHANG Xiao，SHAO Lisong，et al. Challenges and prospects of cyber security protection for new power system[J]. Automation of Electric Power Systems，2023，47(8)：15-24.

[5] 严康，陆艺丹，于宗超，等．配电网用户侧异构电力物联设备运行安全管控分析[J]．电力系统自动化，2023，47(8)：53-61.
YAN Kang，LU Yidan，YU Zongchao，et al. Analysis on operation security management and control for user-side heterogeneous power internet-of-things devices in distribution network[J]. Automation of Electric Power Systems，2023，47(8)：53-61.

[6] AHMED A，KRISHNAN V V G，FOROUTAN S A，et al. Cyber physical security analytics for anomalies in transmission protection systems[J]. IEEE Transactions on Industry Applications，2019，55(6)：6313-6323.

[7] LI X Y，CHEN P F，JING L X，et al. SwissLog：robust anomaly detection and localization for interleaved unstructured logs[J]. IEEE Transactions on Dependable and Secure Computing，2023，20(4)：2762-2780.

[8] QI J X，LUAN Z Z，HUANG S H，et al. LogEncoder：log-based contrastive representation learning for anomaly detection[J]. IEEE Transactions on Network and Service Management，2023，20(2)：1378-1391.

[9] DU M，LI F F，ZHENG G N，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，October 30-November 3，2017，Dallas，USA：1285-1298.

[10] MENG W B，LIU Y，ZHU Y C，et al. Loganomaly：Unsupervised detection of sequential and quantitative anomalies in unstructured logs[C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence，August 10-16，Macao，China：4739-4745.

[11] LE V H，ZHANG H Y. Log-based anomaly detection without log parsing[C]// 2021 36th IEEE/ACM International Conference on Automated Software Engineering（ASE），November 15-19，2021，Melbourne，Australia：492-504.

[12] HE S L，ZHU J M，HE P J，et al. Experience report：system log analysis for anomaly detection[C]// 2016 IEEE 27th International Symposium on Software Reliability Engineering（ISSRE），October 23-27，2016，Ottawa，Canada：207-218.

[13] 万龙．自动化配电主站系统设计及其在工程改造中应用[D]．南昌：南昌大学，2019.
WAN Long. Design of automatic distribution master station system and its application in engineering transformation[D]. Nanchang：Nanchang University，2019.

[14] 闫力,夏伟.基于机器学习的日志异常检测综述[J].计算机系统应用,2022,31(9):57-69.
YAN Li, XIA Wei. Survey on log anomaly detection based on machine learning[J]. Computer Systems & Applications, 2022, 31(9): 57-69.

[15] 赵俊华,文福拴,黄建伟,等.基于大语言模型的电力系统通用人工智能展望:理论与应用[J].电力系统自动化,2024,48(6):13-28.
ZHAO Junhua, WEN Fushuan, HUANG Jianwei, et al. Prospect of artificial general intelligence for power systems based on large language model: theory and applications [J]. Automation of Electric Power Systems, 2024, 48(6): 13-28.

[16] 江秀臣,臧奕茗,刘亚东,等.电力设备ChatGPT类模式与关键技术[J].高电压技术,2023,49(10):4033-4045.
JANG Xiuchen, ZANG Yiming LIU Yadong, et al. Power equipment ChatGPT-type model and key technologies[J]. High Voltage Engineering, 2023, 49 (10): 4033-4045.

[17] 孙永辉,孟雲帆,葛磊蛟,等.人工智能赋能微电网运行优化的应用及展望[J].高电压技术,2023,49(6):2239-2252.
LIU Yonghui, MENG Yunfan, GE Leijiao, et al. Application and prospect of microgrid operation optimization enabled by artificial intelligence [J] High Voltage Engineering, 2023, 49 (6): 2239-2252.

[18] LIU Y L, TAO S M, MENG W B, et al. LogPrompt: prompt engineering towards zero-shot and interpretable log analysis [C]// Proceedings of the 2024 IEEE/ACM 46th International Conference on Software Engineering, April 14-20, 2024, Lisbon, Portugal: 364-365.

[19] QI J X, HUANG S H, LUAN Z Z, et al. LogGPT: exploring ChatGPT for log-based anomaly detection [C]// 2023 IEEE International Conference on High Performance Computing & Communications, Data Science & Systems, Smart City & Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys), December 17-21, 2023, Melbourne, Australia: 273-280.

[20] HE P J, ZHU J M, ZHENG Z B, et al. Drain: an online log parsing approach with fixed depth tree [C]// 2017 IEEE International Conference on Web Services (ICWS), June 25-30, 2017, Honolulu, USA: 33-40.

[21] ZHU J M, HE S L, LIU J Y, et al. Tools and benchmarks for automated log parsing [C]// 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), May 25-31, 2019, Montreal, Canada: 121-130.

[22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Advances in neural information processing systems, December 4-9, 2017, Long Beach, USA.

[23] GARG S, KAUR K, KUMAR N, et al. A hybrid deep learning-based model for anomaly detection in cloud datacenter networks [J]. IEEE Transactions on Network and Service Management, 2019, 16(3): 924-935.

[24] XIAO R Z, CHEN H, LU J T, et al. AllInfoLog: robust diverse anomalies detection based on all log features[J]. IEEE Transactions on Network and Service Management, 2023, 20 (3): 2529-2543.

[25] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of NAACL-HLT, June 2-7, 2019, Minneapolis, USA: 4171-4186.

[26] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners [C]// Advances in neural information processing systems, December 7-12, 2020, Online: 1877-1901.

[27] LIU J C, SHEN D H, ZHANG Y Z, et al. What makes good In-context examples for GPT-3? [C]// Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, May 27, 2022, Dublin, Ireland and Online: 100-114.

[28] YE J, WU Z, FENG J, et al. Compositional exemplars for in-context learning [C]//International Conference on Machine Learning. PMLR, July 23-29, 2023, Hawaii, USA: 39818-39833.

[29] ZHU J M, HE S L, HE P J, 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), October 9-12, 2023, Florence, Italy: 355-366.

[30] OLINER A, STEARLEY J. What supercomputers say: a study of five system logs [C]// 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'07), June 25-28, 2007, Edinburgh, UK: 575-584.

[31] LE V H, ZHANG H Y. Log-based anomaly detection with deep learning: how far are we?[C]// Proceedings of the 44th International Conference on Software Engineering, May 22-27, 2022, Pittsburgh, USA: 1356-1367.

[32] Jiang A Q, Sablayrolles A, Mensch A, et al. Mistral 7B[EB/OL]. arXiv: 2310.06825[2024-05-23]. https://arxiv.org/abs/2310.06825.

[33] hu e j, wallis p, allen-zhu z, et al. LoRA: low-rank adaptation of large language models [C]//Proceedings of the 10th International Conference on Learning Representations, April 25-29, 2022, Vienna, Austria.

王　申(1999-),男,硕士研究生,主要研究方向为:人工智能在电力系统中的应用。E-mail:wangshen@ncepu.edu.cn

魏兴慎(1986-)男,硕士,高级工程师,主要研究方向为:网络安全、图神经网络。E-mail:weixshen@gmail.com

朱卫平(1983-),男,博士,高级工程师,主要研究方向为:配电自动化。E-mail:jszhuweipin@163.com

关志涛(1979-)男,通信作者,教授,博士生导师,主要研究方向为:电力信息安全。E-mail:guan@ncepu.edu.cn

(编辑　代长振)

Log Abnormity Detection for Distribution Master Station Based on Large Language Models

WANG Shen[1], *WEI Xingshen*[2], *ZHU Weiping*[3], *ZHU Daohua*[4], *GUAN Zhitao*[1]

(1. School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China;
2. Nanjing NARI Information and Communication Technology Co., Ltd., Nanjing 211106, China;
3. State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210024, China;
4. Electric Power Research Institute of State Grid Jiangsu Electric Power Co., Ltd., Nanjing 211103, China)

Abstract: Log abnormity detection is a critical technology for monitoring the operation of distribution master station systems and identifying anomalous behaviors. Existing deep learning-based log abnormity detection methods rely on large amounts of labeled training data, yet the lack of annotated training data in power distribution master station systems leads to significant performance degradation in log abnormity detection. Based on the contextual reasoning capabilities of large language model (LLM), this paper proposes LogAdapt—a training-free log abnormity detection scheme for distribution master stations. The Proposed in-context learning (ICL) example filtering algorithm is designed to dynamically select a number of high-quality ICL examples from a small amount of locally labeled online logs tailored to different types of logs. By integrating task descriptions and human expert knowledge, it automatically constructs text prompts to guide LLM in completing the task of abnormity detection in distribution master station logs. The experimental results show that the proposed scheme has better performance compared to existing schemes.

This work is supported by State Grid Corporation of China (No.5400-202340217A-1-1-ZN).

Key words: log abnormity detection; distribution master station; large language model; prompt engineering; in-context learning; deep learning

附录 A

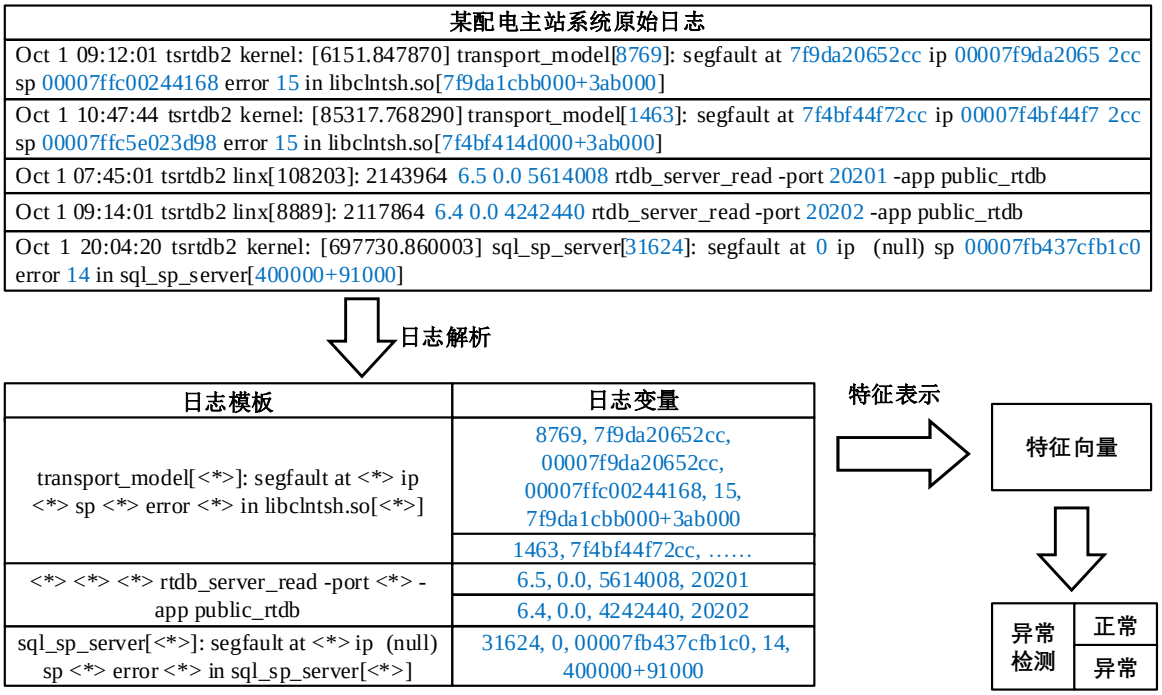


图 A1　配电主站日志异常检测流程

Fig. A1　General process of anomaly detection in distribution master station log

附录 B

表 B1　基于聚类的候选集采样

Table B1　Cluster-Based Candidate Set Sampling

算法 1 基于聚类的候选集采样	
输入: 本地日志集合 X, 词嵌入方法 $Emb(\cdot)$, 聚类方法 $Cluster(\cdot)$, 候选集合大小 K	
输出: C: 包含 K 个样本的候选集	
1 $C=\Phi$, Emb = BERT, $Cluster$ = K-means (初始化)	
2 $V=Emb(X)$	
3 $clusters=Cluster(V, K)$	
4 while$K>0$ do	
5	$K=K$-1
6	$C=C\cup\{\text{random } c\in clusters_K\}$
7 end while	
8 Return$C$	

附录 C

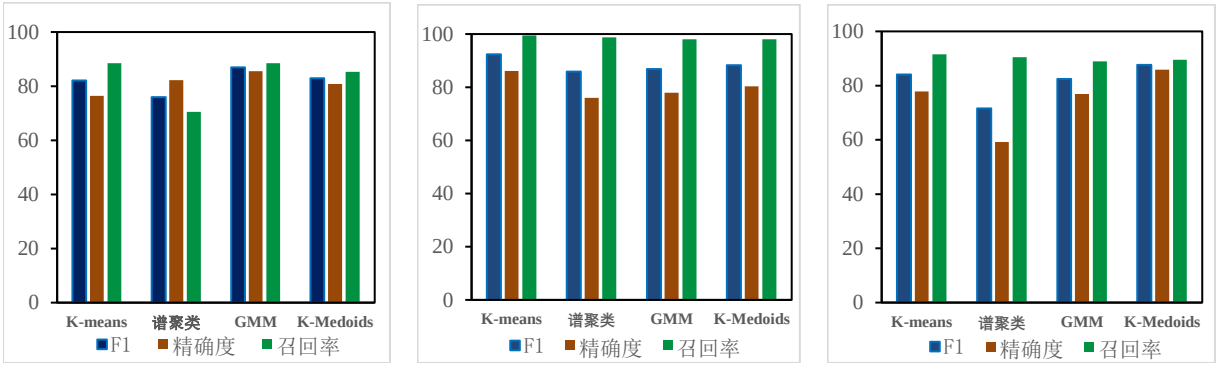
C1 聚类采样方法的影响

聚类采样是本文方案中用于构建ICL候选集合的关键技术,其决定了候选样本的数量和分布,进而影响基于上下文学习的日志异常检测的效果。

本小节选取了K-means聚类、谱聚类(Spectral Clustering)、K-Medoids聚类和高斯混合模型(GMM)聚

表 B2　基于 *k*NN 的 ICL 样本筛选
Table B2　kNN-Based ICL Sample Selection

算法 2 基于 kNN 的 ICL 样本筛选	
输入：在线日志 x_q，候选集 C，词嵌入方法 $Emb(\cdot)$，余弦相似度计算函数 $cos\_sim(\cdot)$，ICL 样本数量 k。	
输出：Θ：包含 n 个样本的 ICL 示例集合	
1 $\Theta=\Phi$，$Emb=$ BERT（初始化）	
2 $V_q = Emb(x_q)$	
3 for $c \in C$ do	
4	$V_c = Emb(c)$
5	$s_c = cos\_sim(V_q, V_c)$
6 end for	
7 选择相似度最大的 n 个候选下标 s_c（按照升序排列）的下标 $\{\delta(1), \cdots, \delta(k)\}$	
8 $\Theta = \{C_{\delta(1)}, \cdots, C_{\delta(k)}\}$	
9 Return Θ	



(a)　BGL 数据集　　(b)　Spirit 数据集　　(c)　Electricbird 数据集

图 C1　不同聚类采样方法对实验结果的影响
Fig. C1　Influence of different clustering sampling methods on the experimental results

类四种常见的聚类算法进行实验与探究，实验结果如图 C1 所示。其中，K-means 是一种基于距离的聚类算法，其通过迭代地将数据点分配到最近的聚类中心来获得聚类结果；谱聚类是一种基于图论的聚类算法，通过构建并切割数据点之间的相似性图来进行聚类；K-Medoids 与 K-means 类似，但其使用数据点而非聚类中心来进行聚类，从而减少了噪声和异常值的影响；GMM 聚类假设数据由多个高斯分布混合而成，并通过拟合这些分布进行聚类。

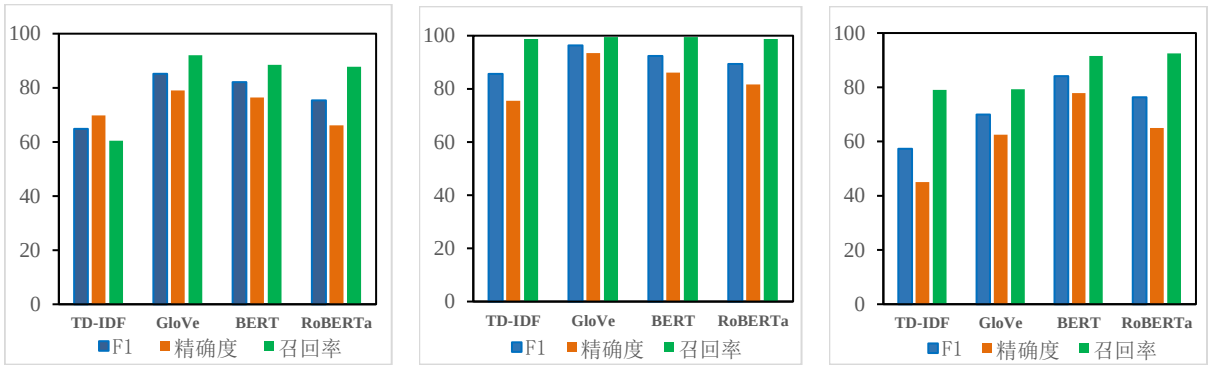
通过观察图 C1，可以发现四种聚类方法在三个数据集上均能表现出较好的日志异常检测性能。其中 K-means 和 K-Medoids 具备更为稳定的性能，推测这是因为经过预处理的日志数据通常具有相对规则的结构，较为适合使用基于距离的聚类方法。而谱聚类性能相对较弱，GMM 聚类在特定数据集如 BGL 上表现良好，在另外两个数据集上表现一般，推测这是因为这两种方法对数据分布都有较强的假设，当数据的实际分布与这些假设不完全匹配时，聚类效果便会受到影响。

C2 词嵌入方式的影响

词嵌入技术是本方案用到一项重要技术，在聚类采样和相似示例筛选两个关键环节都需要用到该技术进行日志的特征表示，是决定示例样本多样性和相似性的重要影响因素。

本小节选取了四种常见的词嵌入方法 TF-IDF、GloVe、BERT 和 RoBERTa 进行实验与探究，实验结果如图 C2 所示。其中 TF-IDF 是一种基于统计的词嵌入方法，通过计算词频和逆文档频率来生成词向量；GloVe 是一种全局向量表示方法，通过对词汇-上下文共现矩阵进行分解来得到词向量；而 BERT 和 RoBERTa 则是基于 Transformer 的预训练模型，通过大规模的预训练学习到了丰富的语义知识，进而提取出词向量。

观察图 C2，可以发现四种词嵌入方法在三个数据集上均能表现出较好的日志异常检测性能。在这四种



(a) BGL 数据集　(b) Spirit 数据集　(c) Electricbird 数据集

图 C2　不同词嵌入方法对实验结果的影响

Fig. C2　Influence of different word embedding methods on the experimental results

方法中，TF-IDF 的性能相对较弱，推测这是因为在相似示例筛选阶段，文本数量较少，导致基于统计的词嵌入方法 TF-IDF 无法充分发挥作用。而相比之下，GloVe、BERT 和 RoBERTa 这三种在大型语料库上进行过预先训练的词嵌入方法则能够保持相对稳定的性能表现。