

开题报告

课程名称：DevOps理论与实践

题目：基于轻量级TextCNN的OpenStack日志异常检测研究

39组

组员姓名：陈展,张耀宇,周林辉,贾亦宸

一、研究背景与意义

随着云计算技术的快速发展，OpenStack作为开源云平台的核心基础设施，其稳定性与可靠性直接影响上层服务的质量。在实际运维中，系统日志是反映平台运行状态的关键数据源，然而传统基于关键字匹配或规则引擎的监控方法存在明显局限性：一方面无法有效识别未知异常模式，另一方面产生大量误报，增加了运维负担。例如，内存泄漏等潜在故障在爆发前往往在日志中呈现特定序列模式，但传统方法难以提前捕捉这类细微异常[1]。

当前基于深度学习的日志异常检测方法（如BERT、LSTM等）虽在准确率上有所提升，但是模型复杂度太高、计算资源需求大，难以在资源受限的中小型集群中实际部署[2]。尤其是在DevOps实践中，轻量级、可扩展的监控方案是实现CI/CD的必要条件。

我们小组的研究旨在探索轻量级TextCNN模型在OpenStack日志异常检测中的应用，通过模型架构优化与日志模板化处理，在保证检测准确性的同时显著降低资源消耗，为中小规模云平台提供符合DevOps理念的自动化运维解决方案。

二、研究问题与方法

我们的核心研究问题是如何在有限计算资源约束下实现OpenStack日志的准确、高效异常检测，并满足DevOps对实时性与可扩展性的要求？

我们采用“模板化预处理+轻量级TextCNN”的两阶段方法。在特征工程方面，

日志解析与模板生成

我们将基于改进的Drain算法[3]，对半结构化OpenStack日志进行解析，保留关键参数（如实例ID、错误码）的同时生成标准化模板。例如，原始日志条目：

```
"Instance 78dc1847-8848-49cc-933e-9239b12c9dcf power on failed"
```

将被归一化为模板：

"Instance * power on failed"

该方法减少数据稀疏性，提升模型泛化能力。

对于模型方面，

在Yoon Kim的TextCNN架构基础上[4]，我们准备进行三方面优化：嵌入维度从300维压缩至100维，减少参数规模；采用单层卷积结构，卷积核数量控制在64个；使用全局最大池化替代全连接层，能够进一步降低计算复杂度。

我们使用公开OpenStack日志数据集。这个数据集里面包含normal1.log/abnormal.log等日志文件，总大小约在50MB左右。我们准备按7:2:1划分训练集、验证集与测试集。除准确率、召回率和F1值外，我们会重点评估模型在推理延迟、资源占用等方面的性能，并与支持向量机等机器学习模型及LSTM等深度学习模型进行对比。

三、研究计划与进度安排

- **第一阶段（11.6-11.10）：数据预处理模块开发**

完成日志解析工具链搭建，实现模板自动生成与验证

- **第二阶段（11.11-11.20）：模型训练与调优**

完成TextCNN模型实现与基线测试，通过网格搜索优化超参数

- **第三阶段（11.21-11.30）：报告撰写与答辩准备**

整理实验数据与可视化图表，完成终版报告与模拟答辩

四、小组分工

- 陈展：总体方案设计、轻量化模型优化、答辩主讲；
- 张耀宇：日志解析算法实现、数据预处理与质量验证；
- 周林辉：TextCNN模型训练、超参数调优与消融实验设计；
- 贾亦宸：系统集成、API开发、性能测试与对比分析。

协作采用Git分支开发模式，代码需经双人评审后方可合并；

参考文献

- [1] He S, Zhu J, He P, et al. Loghub: A Large Collection of System Log Datasets for AI-driven Log Analytics[C]//2023 IEEE International Symposium on Software Reliability Engineering (ISSRE). IEEE, 2023: 1-12.
- [2] Pang G, Shen C, Cao L, et al. Deep Learning for Anomaly Detection: A Review[J]. ACM Computing Surveys, 2020, 1(1): 1-38.

- [3] He P, Zhu J, Zheng Z, et al. Drain: An online log parsing approach with fixed depth tree[C]//2017 IEEE International Conference on Web Services (ICWS). IEEE, 2017: 33-40.
- [4] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.
- [5] OpenStack Log Dataset. Logpai/loghub[EB/OL]. <https://github.com/logpai/loghub>