LogBERT：基于BERT的日志异常检测

---

1 背景

1.1 问题重述

检测系统日志中可能出现的异常。

1.2 解决方案

传统方法四种：

1. 主成分分析
2. One Class SVM
3. 隔离森林
4. 聚类分析

基于学习的方法三种：

1. Log Anomaly：基于深度学习
2. DeepLog：基于RNN
3. LogBERT：基于BERT（本文的方法）

1.3 RNN的局限性

1. 传统RNN基于从左到右的循环训练，无法同时利用左右上下文。而在日志情景下同时利用左右上下文进行预测是有必要的。
2. 双向RNN可以捕获上下文信息，但它面临梯度消失或爆炸的问题，这意味着对长序列效果不佳，无法捕捉长期依赖性。
3. 当前基于RNN的模型仅考虑通过给定先前日志消息预测下一条日志，依赖于正常日志消息间的相关性，而本文提出了一种方法编码所有正常序列共享的公共模式。

2 介绍

2.1 基于学习的方法

典型流程：

1. 采用日志解析器将日志消息转化为日志键
2. 使用特征提取方法（例如TF-IDF）构建特征向量
3. 基于有监督学习或无监督学习方法进行训练。由于异常序列的稀缺，一般采用无监督学习方法

> **TF-IDF（词频-逆文档频率）**
>
> 词频：描述了词在文档中的出现次数
>
> 逆文档频率：描述了词在所有文档中的稀有程度
>
> $$\text{TF-IDF} = \text{TF} \times \text{IDF}$$

> **有监督学习**：提供已标注的异常数据，从正常和异常序列中共同学习
>
> **无监督学习**：不提供异常数据，从正常序列的行为中学习异常序列

2.2 双向编码表示Transformer (BERT)

BERT 独特的训练任务是 **Masked Language Model(MLM)** 。

- 在训练过程中，输入序列中的一些词会随机地被 Mask ，模型的根据未被 Mask 的上下文预测这些被掩盖的词。

特点：

1. **编码**：专注于Transformer的编码器部分，用于生成上下文表示
2. **双向**：在处理每个词时会同时考虑该词前面和后面的上下文信息。

3 LogBERT

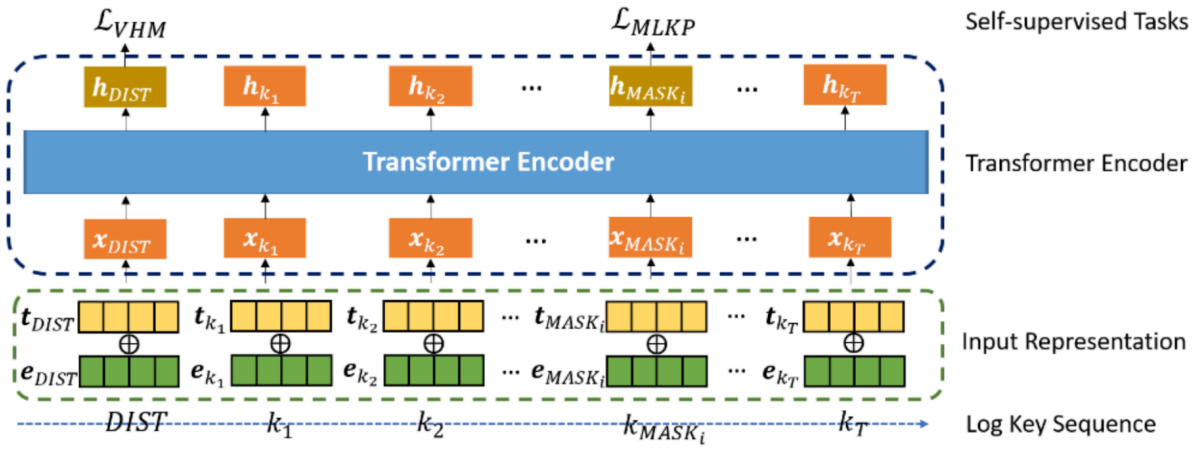

图 1：LogBERT 概述

训练流程：

1. 提取日志键：$S = \{k_1, \ldots, k_T\}$
2. 添加标记：$S' = \{\text{DIST}, k_1, \ldots, k_T\}$
3. 对顺序建模：$T_{t,2i} = \sin(t/10000^{2i/d}), T_{t,2i+1} = \cos(t/10000^{2i/d})$
4. 综合2,3步的结果：$x_{k^t} = e_{k^t} + t_{k^t}$
5. Transformer层：每层包括一个多头子注意力子层和位置前馈子层，然后进行层归一化
6. Transformer编码器：包括多个Transformer层

两个自监督训练任务：

1. 屏蔽日志键预测(MLKM)：
   - 将有随机Mask的日志作为输入，目标是准确预测被屏蔽的日志键
   - 旨在正确预测随机屏蔽的正常日志序列中的消息
2. 超球面体积最小化(VHM)
   - 将DIST的嵌入h_{DIST}视为整个日志序列的嵌入，目标是使每个嵌入与它们的平均值$c = Mean(h)$尽量接近
   - 旨在使正常日志序列在嵌入空间中彼此接近

TABLE II: Experimental Results on HDFS, BGL, and Thunderbird Datasets

Method	HDFS			BGL			Thunderbird		
	Precision	Recall	F-1 score	Precision	Recall	F-1 score	Precision	Recall	F-1 score
PCA	5.89	100.00	11.12	9.07	98.23	16.61	37.35	100.00	54.39
iForest	53.60	69.41	60.49	99.70	18.11	30.65	34.45	1.68	3.20
OCSVM	2.54	100.00	4.95	1.06	12.24	1.96	18.89	39.11	25.48
LogCluster	99.26	37.08	53.99	95.46	64.01	76.63	98.28	42.78	59.61
DeepLog	88.44	69.49	77.34	89.74	82.78	86.12	87.34	99.61	93.08
LogAnomaly	94.15	40.47	56.19	73.12	76.09	74.08	86.72	99.63	92.73
LogBERT	87.02	78.10	**82.32**	89.40	92.32	**90.83**	96.75	96.52	**96.64**

TABLE III: Performance of LogBERT base on One Self-supervised Training Task

	HDFS			BGL			Thunderbird		
	Precision	Recall	F-1 score	Precision	Recall	F-1 score	Precision	Recall	F-1 score
MLKP	77.54	78.65	78.09	93.16	86.46	89.69	97.07	95.90	96.48
VHM	2.43	39.17	4.58	71.04	43.84	54.22	56.58	43.87	49.42
Both	87.02	78.10	82.32	89.40	92.32	90.83	96.75	96.52	96.64

两个实验（结果见上图）：

1. 与其他方法对比
2. 消融训练任务的验证