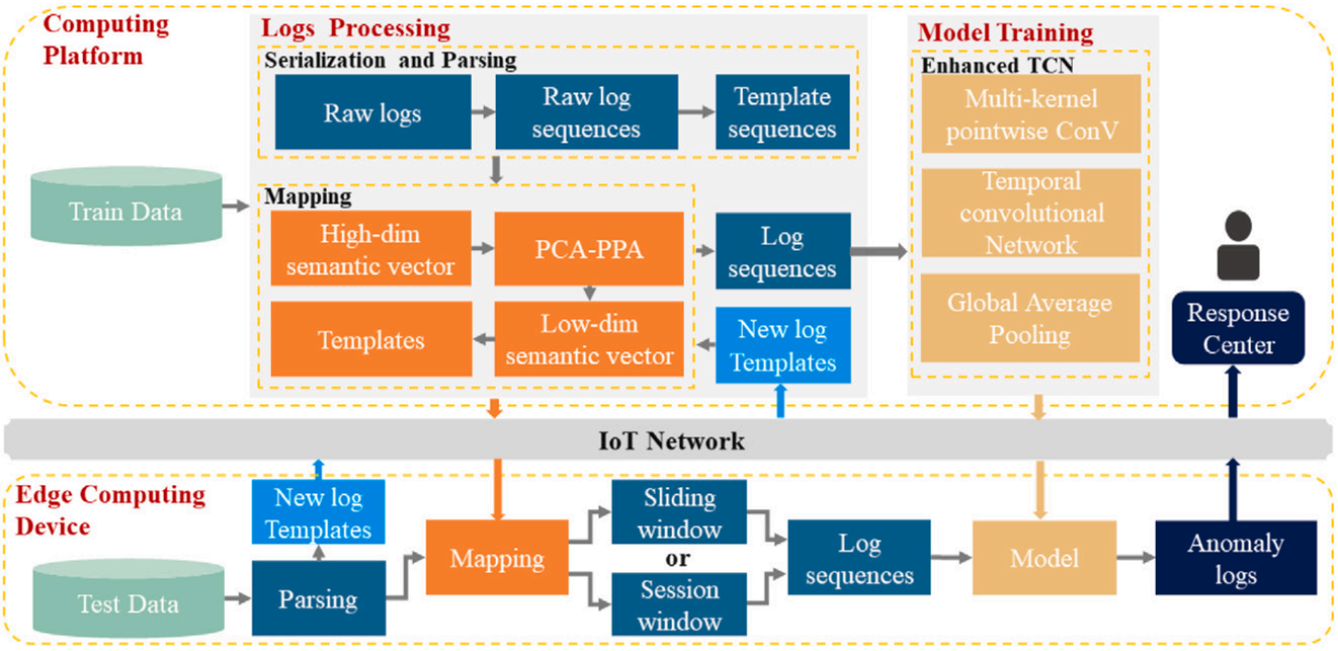
文献综述

相关主要研究工作

监督

LightLog[1]



日志模板解析之后使用wordvec进行语义编码，然后使用PCA和PPA将这些编码的多维向量进行降维，防止参数过多。

然后使用TCN模型进行训练。

解决的问题：日志异常检测轻量化。

Neruallog[2]

不使用日志解析，因为日志解析不准确且会造成噪声。

如下解释：

- **Anomaly:** ciod: LOGIN chdir(/p/gb1/stella/RAPTOR/2183) failed: ***Input/output error***
- **Normal:** ciod: LOGIN chdir(/home/bertsch2/src/bgl_hello) failed: ***Permission denied***
- **Parsed event:** ciod: LOGIN * failed: * *

(a) Errors introduced by Drain

- **Anomaly:** mptscsih: ioc0: ***attempting task abort!*** (sc=00000101bfc7a480)
- **Normal:** mptscsih: ioc0: ***task abort: SUCCESS*** (sc=00000101bfc7a480)
- **Parsed event:** mptscsih ioc0 * * * sc *

(b) Errors introduced by Spell

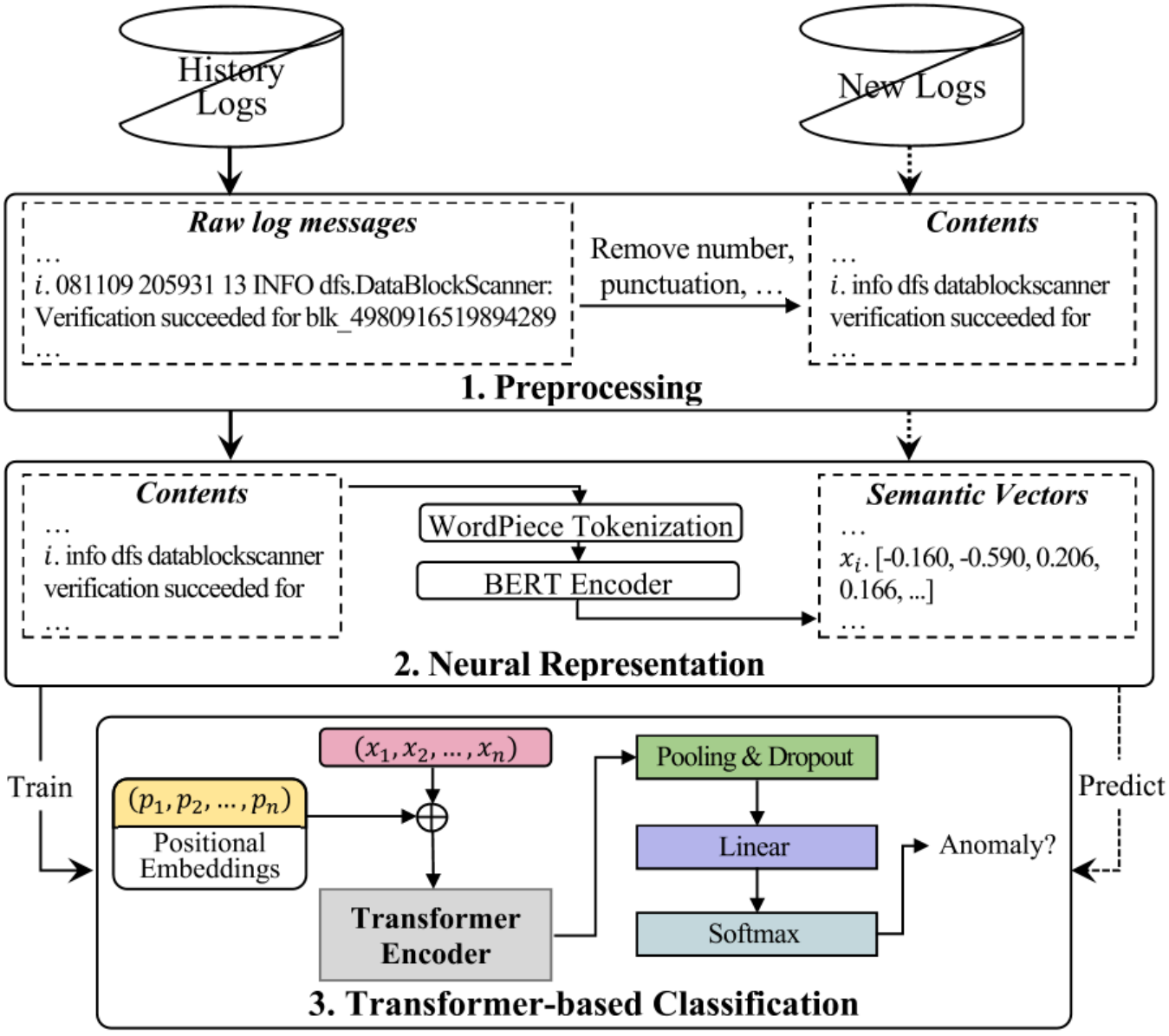
- **Anomaly:** floating point ***unavailable interrupt***
- **Normal:** floating point ***instr. enabled.....1***

- **Parsed event:** floating point * *

(c) Errors introduced by AEL

- **Anomaly:** ciod: Error creating node map from file ***map.dat: No child processes***
- **Normal:** ciod: Error creating node map from file ***map.dat: Bad file descriptor***
- **Parsed event:** ciod Error creating node map from file * * * *

(d) Errors introduced by IPLoM

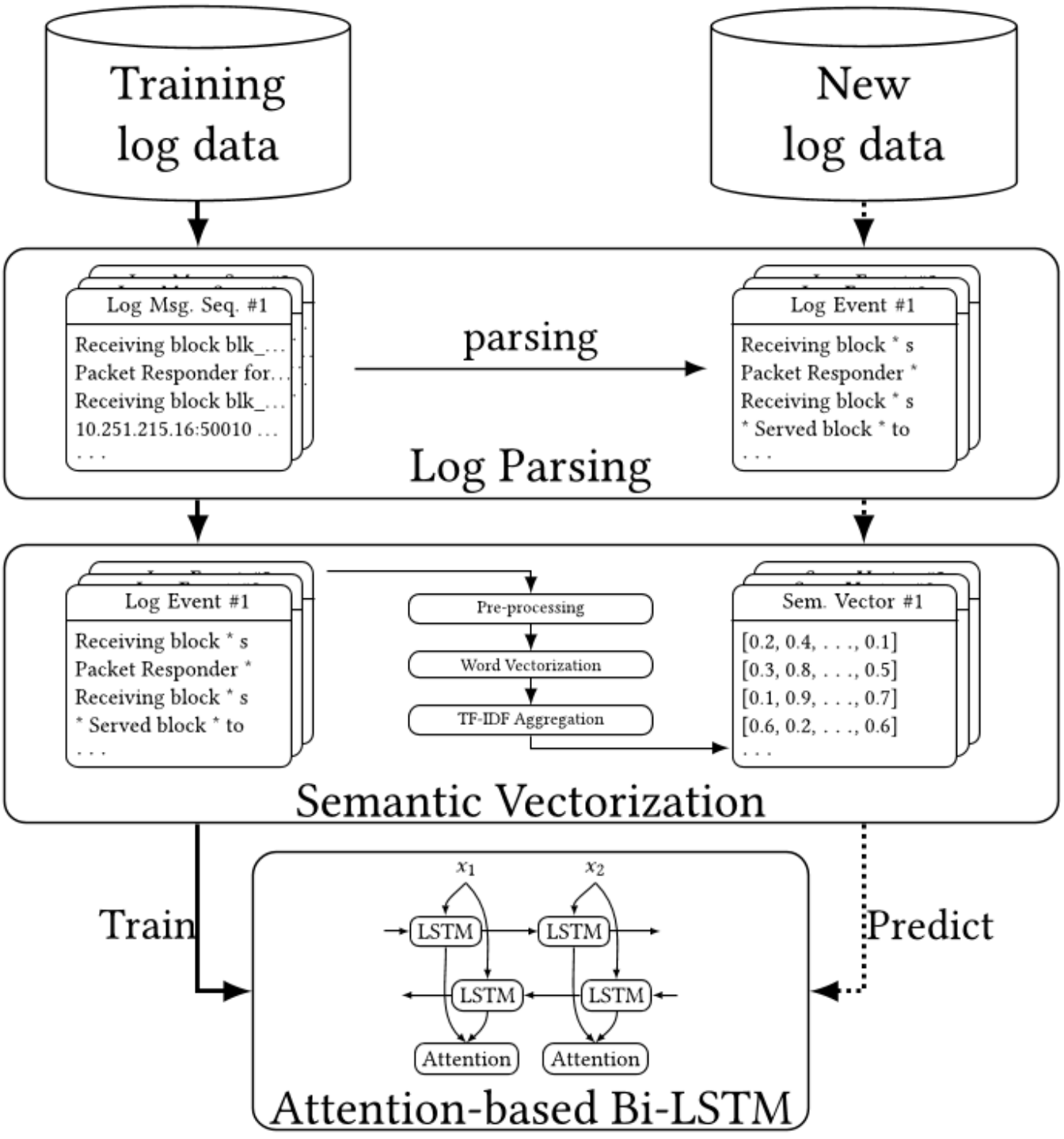


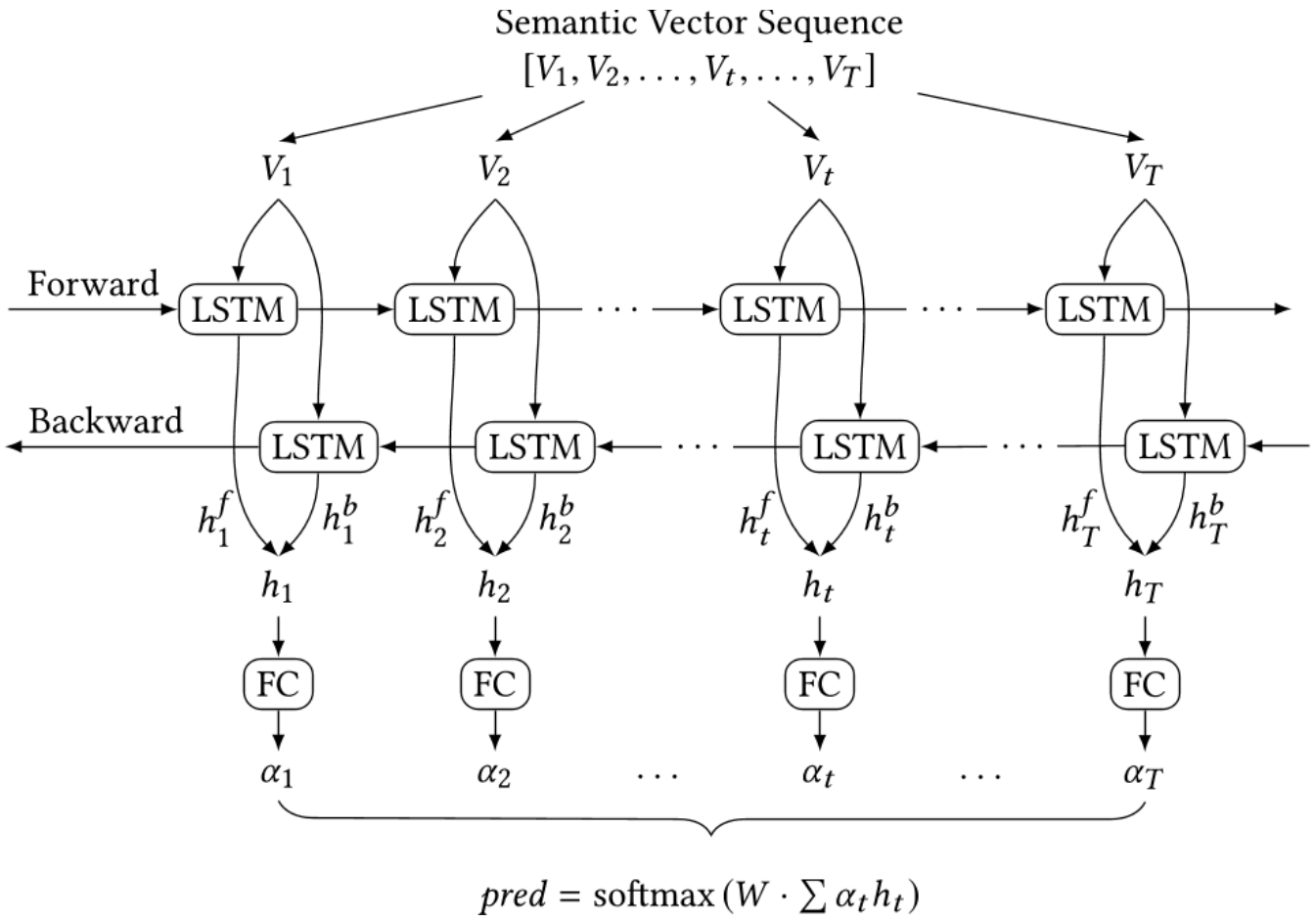
解决的问题：日志不经过日志解析器而进行日志异常检测。

LogRubost[3]

LogRobust 双向LSTM+Attention进行编码分类，对原始日志的模版进行word vector的向量化送入模型进行分类。

]





解决的问题：

- **指出了基于日志的异常检测中存在的日志不稳定性问题。我们还对现实世界系统的日志不稳定性进行了实证研究，以证实我们的发现。**
- **日志格式和日志中的语句会变化，模型提高检测的鲁棒性**

亮点：

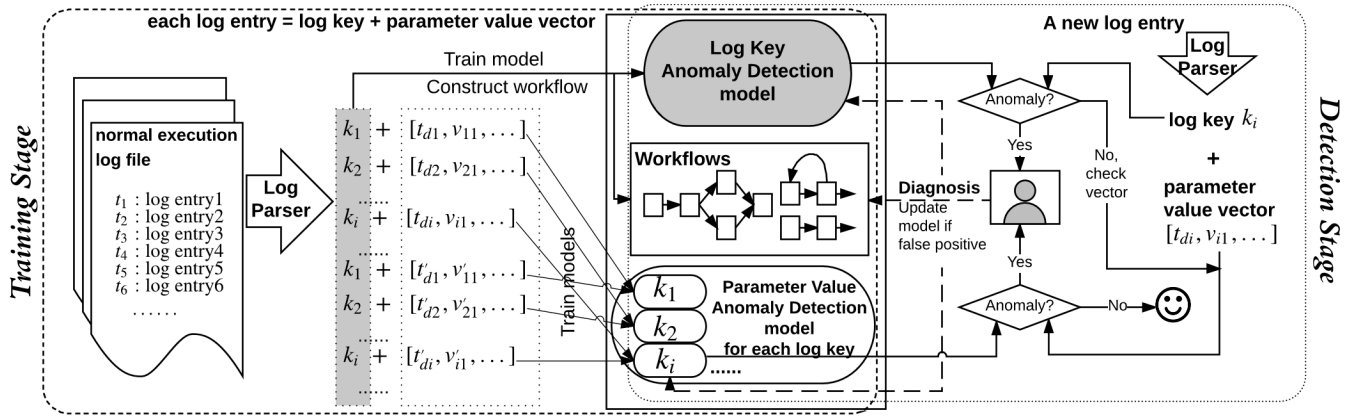
- 首次提出了TF-IDF编码。
- 论文写的非常好。
- 提出了如何构建不稳定数据集的方法。

半监督

DEEPLOG[4]

DeepLog 是一种基于深度学习的日志分析方法，旨在检测系统日志中的异常行为。DeepLog 在日志分析领域引入了循环神经网络（RNN）和长短时记忆网络（LSTM）的应用，有效地捕获了日志序列中的依赖关系和模式。

在 DeepLog 中，系统日志首先经过预处理，将日志中的文本转换为离散的日志键值，从而降低了输入数据的维度。接下来，DeepLog 利用 LSTM 网络对日志序列进行建模。LSTM的设计使其能够捕获长距离依赖关系，从而有效地挖掘日志序列中的潜在规律。在训练过程中，**DeepLog 通过学习正常日志序列的模式，从而具备了区分正常和异常日志序列的能力。** 在异常检测阶段，DeepLog 根据训练好的 LSTM 模型为每个日志事件生成一个预测概率分布，并计算实际日志事件与预测概率分布之间的交叉熵损失。如果损失超过预先设定的阈值，则将相应的日志事件标记为异常。



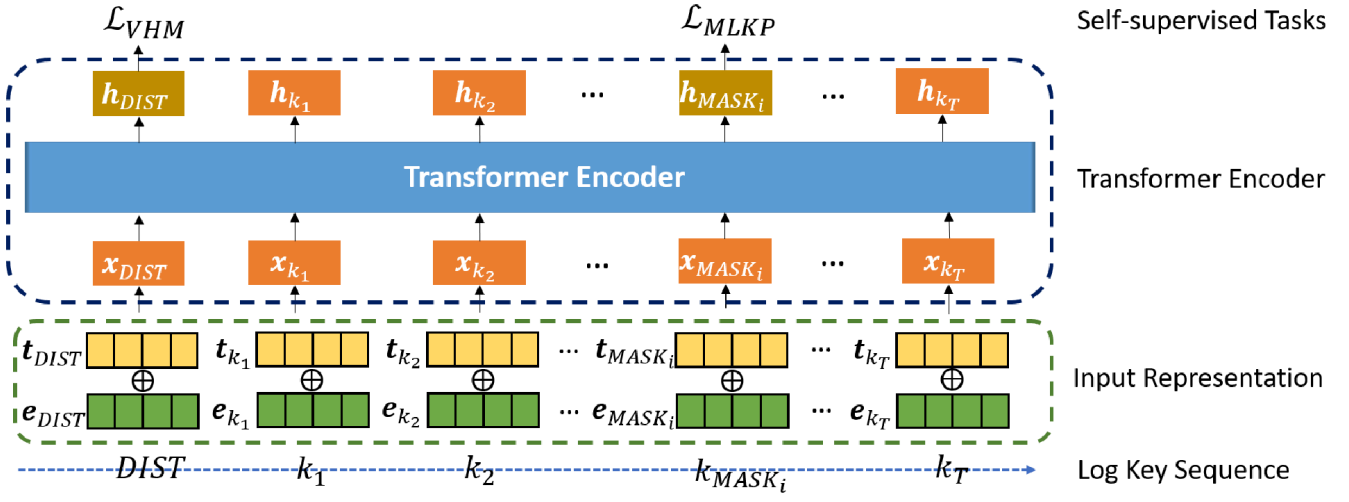
解决问题：日志异常检测的开山之作，首次将深度学习引入这个领域。

LogBERT[5]

第一个任务是Masked Log Key Prediction (MLKP)。在这个任务中，LogBERT 通过在日志序列中随机替换一定比例的日志键值为MASK token，使模型能够预测被MASK的日志键值，从而捕捉日志序列的上下文信息。

第二个任务是 Volume of Hypersphere Minimization (VHM)。在 VHM任务中，使用**所有正常日志序列**的中心点作为超球的中心，训练 LogBERT，使其编码能够更好地聚集在超球的表面附近。

这两个任务共同目标是训练 LogBERT 来捕捉正常日志序列的模式，以便检测出与正常日志序列模式有所偏差的异常事件。



PLELOG[6]

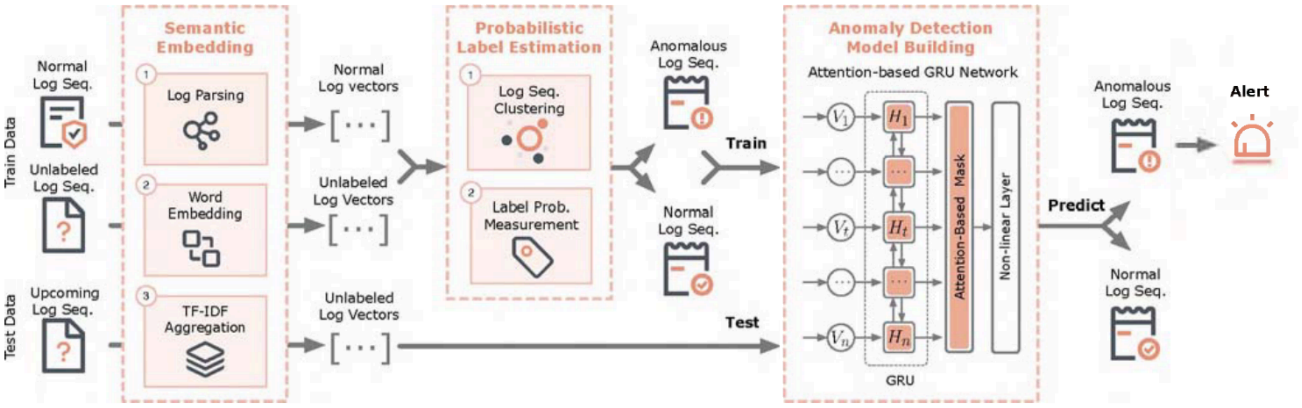


Fig. 2: Overview of PLELog

向量化后聚类，然后根据聚类之后的指标打上伪标签，然后将这些数据传入GRU模型中进行分类。

这个论文结合了半监督的和监督的方法。

输入的数据都是正常日志，当异常日志与正常日志发生偏移的时候才会捕获异常。

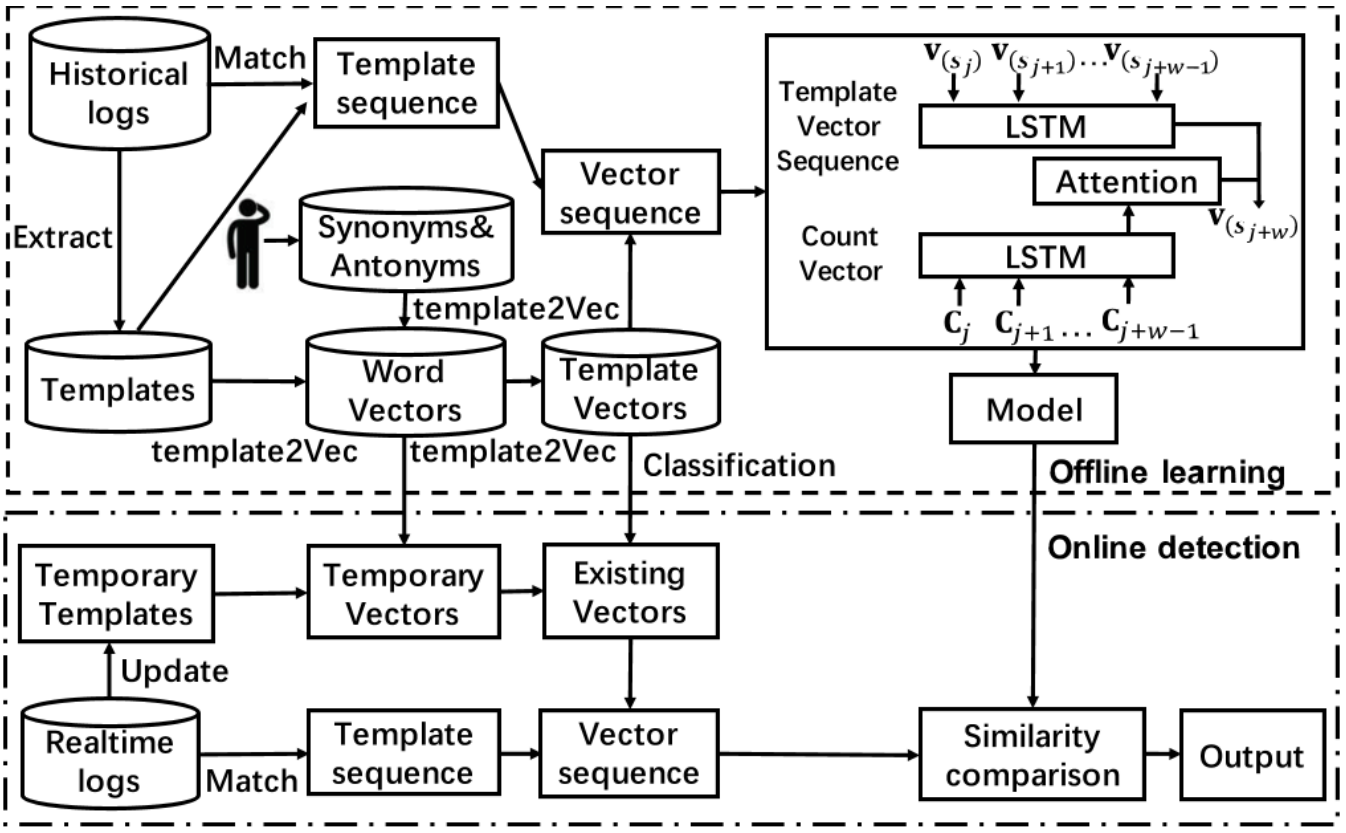
解决问题：

- 出了一种实用且鲁棒的基于日志的异常检测方法PLELog，该方法是半监督的，并通过概率标签估计结合了历史异常的知识。此外，PLELog能够通过语义嵌入对不稳定的日志数据保持免疫，并通过设计基于注意力的GRU神经网络高效地检测异常。
- 本文提出了一种基于已知正常日志序列的聚类方法来估计未标记对数序列的标记。特别地，我们设计了概率标签估计，减少聚类产生的噪声的影响。

无监督

LogAnomaly （2019）[7]

在 LogAnomaly 中，首先对原始日志数据进行预处理，提取有意义的日志键值，并将其转换为数值表示。然后，将处理后的日志序列输入到一个 LSTM网络中，以学习日志数据的时间特性和潜在结构。为了进行异常检测，LogAnomaly 采用了一种基于异常分数的方法。**在训练阶段，使用正常日志序列训练 LSTM网络。**在检测阶段，将新的日志序列输入到训练好的 LSTM 网络中，计算每个日志事件的异常分数。异常分数反映了输入日志事件与正常日志模式之间的差异程度。若异常分数超过预先设定的阈值，则将相应的日志事件标记为异常。



解决的问题：

- 受词嵌入的启发，提出template2vec，来捕获包括同义词和反义词在内的语义信息。
- 设计一种无需操作员反馈就可以合并新的日志模板的机制。
- 提出LogAnomaly，一个端到端使用LSTM网络自动同时检测**序列和定量**异常的框架。

> DeepLog的PLUS版本

解决问题

日志的标签获取难度大，且需要专家的先验知识--半监督或者弱监督

日志会随着时间发生变化-- 模型需要鲁棒性--这里我使用对比学习的方法解决

日志具有不同特征--这里我使用加权或者注意力机制来融合多个特征

参考文献

[1] Wang, Z., Tian, J., Fang, H., Chen, L. & Qin, J. (2022). LightLog: A lightweight temporal convolutional network for log anomaly detection on the edge. Computer Networks, 203(), 108616. https://doi.org/10.1016/j.comnet.2021.108616

[2] Le, V. & Zhang, H. (2021). Log-based Anomaly Detection Without Log Parsing. 2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE), null(null), 492-504. https://doi.org/10.1109/ASE51524.2021.9678773

[3] Zhang, X., Xu, Y., Lin, Q., Qiao, B., Zhang, H., Dang, Y., Xie, C., Yang, X., Cheng, Q. & Li, Z. (2019). Robust log-based anomaly detection on unstable log data. Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, (), 807-817. https://doi.org/10.1145/3338906.3338931

[4] Du, M., Li, F., Zheng, G. & Srikumar, V. (2017). DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, (), . https://doi.org/10.1145/3133956.3134015

[5]Guo, H., Yuan, S. & Wu, X. (2021). LogBERT: Log Anomaly Detection via BERT. 2021 International Joint Conference on Neural Networks (IJCNN), (), 1-8. https://doi.org/10.1109/IJCNN52387.2021.9534113

[6]Yang, L., Chen, J., Wang, Z., Wang, W., Jiang, J., Dong, X. & Zhang, W. (2021). Semi-Supervised Log-Based Anomaly Detection via Probabilistic Label Estimation. , (), 1448-1460. https://doi.org/10.1109/ICSE43902.2021.00130

[7]Meng, W., Liu, Y., Zhu, Y., Zhang, S., Pei, D., Liu, Y., Chen, Y., Zhang, R., Tao, S. & Sun, P. (2019). LogAnomaly: Unsupervised Detection of Sequential and Quantitative Anomalies in Unstructured Logs. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, (), 4739-4745. https://doi.org/10.24963/ijcai.2019/658