哈爾濱工業大學

组合优化与凸优化
阅读报告

题　　目	异常检测与根因定位的方法研究
学　　院	计算学部
专　　业	电子信息
学　　号	24S103184
学　　生	胡冠宇
任 课 教 师	刘绍辉

哈尔滨工业大学计算学部

2025.3

一、问题介绍

工业控制系统（Industrial Control Systems, ICS）和微服务架构作为现代工业与互联网应用的核心组成部分，其安全性与稳定性直接影响生产效率和用户体验。然而，这两大领域在实际运行中均面临复杂的挑战，亟需创新的优化与分析方法。

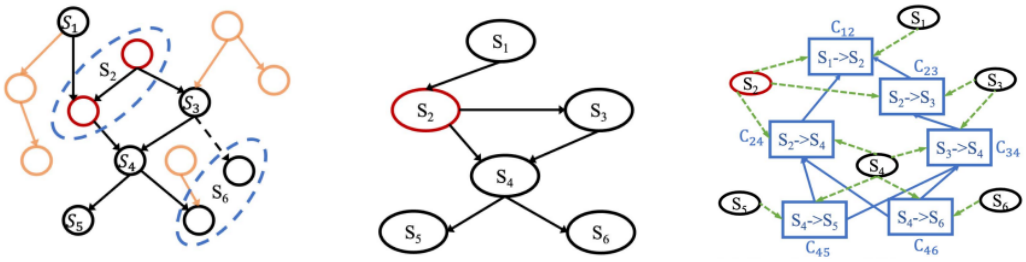
1.1 基于跨域表示学习的工业控制系统异常检测[1]

工业控制系统广泛应用于电力、制造、水利等关键基础设施，其安全性和稳定性至关重要。一旦 ICS 受到攻击，可能会造成严重的损坏。因此，针对 ICS 异常检测是十分重要的。传统异常检测方法主要关注单一域中的指标，如网络域中的网络流量或物理域中的传感器数据，但 ICS 中不同域（如传感器的物理状态、网络通信流量等）的行为存在强相关性，仅分析单一域难以全面识别异常。例如，网络攻击可能导致传感器数据异常，但某些攻击仅影响物理设备而不改变网络流量。现有的方法如基于 RNN[2]、GAN[3]或基于单域图的神经网络模型无法有效建模跨域关联，导致检测精度不足和误报率高。文章提出了一种基于跨域表示学习的 ICS 异常检测方法，该方法能够学习多域行为的联合特征，并在不同的域内进行异常检测。在构建跨域图来表示 ICS 中多个领域的行为之后，该方法可以利用图神经网络学习它们的联合特征。由于异常在不同领域表现不同，利用多任务学习方法分别识别不同领域的异常并进行联合训练。

1.2 基于异构图的大规模微服务系统性能问题诊断[4]

大规模微服务系统在运行过程中通常会产生大量的服务调用，尤其是在短时间内。这意味着当检测到性能问题时，系统周围可能会有大量的调用记录。这种海量的数据使得分析所有调用记录变得既低效又困难，而且根因定位的精度也会受到影响。主要原因在于，其中许多服务调用与性能问题并无直接关系。因此需要识别与性能问题有关的微服务调用。

大多数传统的方法识别异常的调用端口（即端口级），但 MicorDig 的目标是选择异常的微服务（即服务级）。为了更准确地定位异常微服务，文章首先选择构建端口级调用图，然后，通过在图上执行广度优先搜索（Breadth-First Search，BFS）和异常检测来保留相关调用和相应的微服务。如图 1a 所示，圆圈表示端口级节点，同一个虚线椭圆内的端口属于同一个微服务，橙色的端口级节点为问题无关节点，虚线表示的调用为非异常调用。最后，如图 1b 所示，将图中的端口级节点聚合到服务级。



(a) 端口级调用图　　(b) 服务级调用图　　(c) 服务级异构传播图

图 1　异构传播图的构造过程

二、相关求解算法简介及其实现

2.1　MGDN

文章提出一种跨域图表示学习方法。核心思想是通过构建多域图结构，将不同域的行为特征统一编码。该方法结合不同域的数据，利用注意力机制图卷积网络（Graph Convolutional Networks，GCN）学习共享与域特定特征。模型框架如图 2 所示。

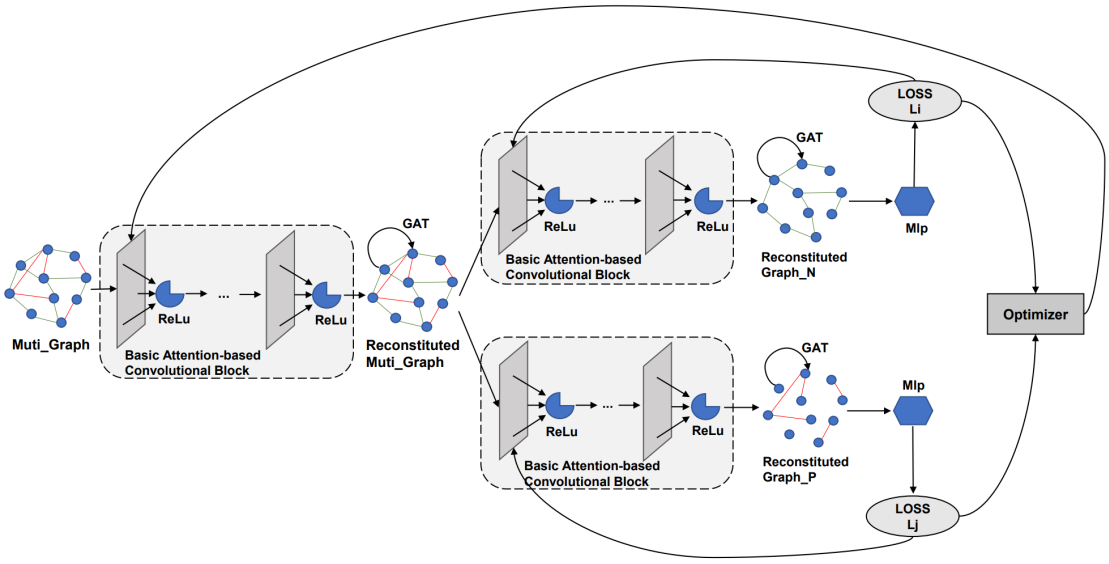


图 2　跨域表示学习的总体框架

2.1.1　多图构建

在该方法中，目标是从 ICS 的多个域（如物理域、网络域等）中提取节点特征，并构建统一的多图结构用于跨域建模。多图表示结构的构构建过程如图 3 所示。

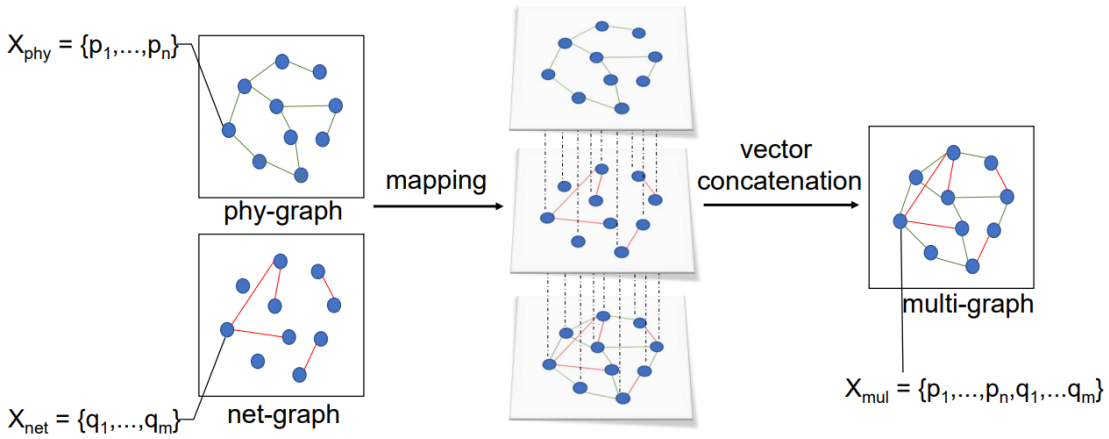


图 3　多图表示结构的构建过程

假设有 n 个节点（传感器和执行器），来自不同域的数据被统一为相同的时间粒度（如秒级粒度），每个节点在第 d 个域上的特征矩阵为

$$\boldsymbol{x}^{(d)} \in \mathbb{R}^{n \times t},$$

其中 t 为时间步长。

对每个域 $d \in \{1, 2, \cdots, D\}$, 构建一个无向加权图

$$\mathcal{G}_d = \langle \mathcal{V}, \mathcal{E}_d \rangle ,$$

其中 $\mathcal{V}$ 表示所有节点的集合，且 $|\mathcal{V}| = n$，$\mathcal{E}_d$ 为第 d 个域中的边集合。

节点之间的边权通过余弦相似度计算其嵌入向量之间的相似性。对于任意两个节点 i 和节点 j，在第 d 域的嵌入向量为 $\boldsymbol{v}_i^{(d)}$ 和 $\boldsymbol{v}_j^{(d)}$，则计算节点 i 到节点 j 的边权

$$e_{ij}^{(d)} = \frac{\left(\boldsymbol{v}_i^{(d)}\right)^{\mathrm{T}} \boldsymbol{v}_j^{(d)}}{\left\|\boldsymbol{v}_i^{(d)}\right\| \left\|\boldsymbol{v}_j^{(d)}\right\|},$$

之后使用 Top-k 策略筛选每个节点最相关的 k 个邻居构建图 $\mathcal{G}_d$，进一步融合成多图结构

$$\mathcal{G} = \left\langle \mathcal{V}, \bigcup_{i=1}^{d} \mathcal{E}_i \right\rangle ,$$

节点 i 的跨域特征向量表示为

$$\boldsymbol{v}_i = \boldsymbol{v}_i^{(1)} \oplus \boldsymbol{v}_i^{(2)} \oplus \cdots \oplus \boldsymbol{v}_i^{(D)}。$$

2.1.2 基于注意力的图卷积建模

该部分通过引入图注意力机制（Graph Attention Network, GAT）[5]对节点信息进行聚合，捕捉局部邻居中的非均匀关系。

首先，定义节点 j 对节点 i 的注意力权重

$$\alpha_{ij} = \text{Softmax}\left(\text{LeakeyReLU}\left(\boldsymbol{a}^{\mathrm{T}}(\boldsymbol{v}_i \oplus \boldsymbol{v}_j)\right)\right),$$

其中 $\boldsymbol{a}$ 为可学习的向量。则节点 i 的表示更新为

$$\boldsymbol{v} = \text{ReLU}\left(\alpha_{ii}\boldsymbol{W}\boldsymbol{v}_i + \boldsymbol{W}\sum_{j\in\mathcal{N}_i}\alpha_{ij}\boldsymbol{v}_j\right),$$

其中 $\boldsymbol{W}$ 为可学习的矩阵。

2.1.3 多目标优化

为了在多个域（如物理域、网络域等）上同时优化预测性能，文章使用了多任务学习（Multi-Task Learning, MTL）方法。其目标是联合优化每个任务的损失

$$\min_{\boldsymbol{W}_s,\boldsymbol{W}_1,\boldsymbol{W}_2,\cdots,\boldsymbol{W}_D}\sum_{d=1}^{D}\gamma_d\mathcal{L}_d(\boldsymbol{W}_s,\boldsymbol{W}_d),$$

其中 $\boldsymbol{W}_s$ 为共享层参数，$\boldsymbol{W}_d$、$\mathcal{L}_d$ 和 γ_d 分别为第 d 域的专属参数、损失函数及任务权重。

为了解决梯度冲突问题，引入了多梯度下降算法（Multiple Gradient Descent Algorithm，MGDA）[6]，其基本思想是寻找一组权重 $\{\gamma_d\}$，使得多个损失函数在共享参数 $\boldsymbol{W}_s$ 上的梯度方向可以共同优化，即

$$\begin{cases}\min\limits_{\gamma_1,\gamma_2,\cdots,\gamma_D} & \left\|\sum\limits_{d=1}^{D}\gamma_d\nabla_{\boldsymbol{W}_s}\mathcal{L}_d(\boldsymbol{W}_s,\boldsymbol{W}_d)\right\|^2 \\ \text{s.t.} & \sum\limits_{d=1}^{D}\gamma_d = 1, \\ & \gamma_1,\gamma_2,\cdots,\gamma_D \geq 0,\end{cases}$$

该优化问题保证在共享参数更新中不会偏向某一特定任务。

2.2 MicroDig

为了捕捉性能问题的传播模式并进一步缩小候选根本原因的范围，文章对关联图中的每一条边执行异常检测，使用一种高效且广泛使用的异常检测方法——k-sigma 方法。该方法从历史数据中学习参数 μ 和 σ，并将超出 $(\mu - k\sigma, \mu + k\sigma)$ 范围的值视为异常。

文章基于端口级数据进行异常检测，而不是基于服务级数据，因为端口级的异常可能意味着包含该端口的服务存在异常行为，如果基于从端口级数据汇总的服务级数据进行异常检测，可能会被这些异常淹没。根据上述步骤，文章保留了关联图中存在异常的边。由于通过上述步骤得到的图是端口级关联图，而文章的目标是定位异常服务，因此需要合并端口的调用数据，并构建服务级关联图。

令 p 为端口级关联图中的一个节点，用 $P(S)$ 表示给定服务 S 的所有端口。为了构建服务级图，需要将同一服务 S 的所有端口级节点合并为一个节点，记为 $S = \{p \in P(S)\}$。在服务级图上，边 $S \to^{call} S'$ 的异常率 $R(S, S')$ 整合了每个相关端口级边 $p \to^{call} p'$ 的异常调用数 $F(p, p')$ 和总调用数 $N(p, p')$，即对于时间节点 t，

$$R_t(S, S') = \frac{\sum_{p \in S, p' \in S'} F_t(p, p')}{\sum_{p \in S, p' \in S'} N_t(p, p')}$$

通过上述步骤，可以构建一个服务级关联图，其中的节点都与问题服务相关。同时，可以获得边 $S = \{p \in P(S)\}$ 的异常率时间序列：

$$R(S, S') = (R_{t-\varphi}, R_{t-\varphi+1}, \cdots, R_{t+\varphi})$$

该时间序列表示在时间区间 $[t - \varphi, t + \varphi]$ 之间中边 $S \to^{call} S'$ 的异常率变化。MicroDig 的总体框架如图 4 所示。

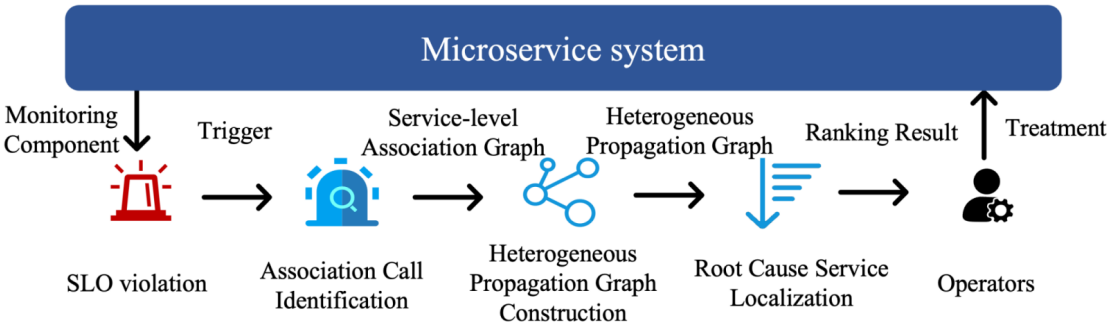


图 4　MicroDig 的总体框架

2.2.1 异构传播图的构造

构建的关联图是有向图，其边的方向表示服务之间的调用关系。然而，调用关系并不等同于因果关系。因此，关联图不能直接用作根本原因定位的因果图。文章提出构建一个异构传播图，以反映调用和服务之间的因果关系，基于关联图进行构建。

算法 1 异构传播图构造

$G_h \leftarrow$ 初始化异构传播图
G_h.addNodes(G.edges())
for all $S \in G$.nodes() **do**
    for all $C_{out} \in G$.outEdges(S) **do**
        for all $C_{in} \in G$.inEdges(S) **do**
            G_h.addEdge(C_{out}, C_{in})
G_h.addNodes(G.nodes())
for all $C \in G$.edges() **do**
    $caller$, $callee \leftarrow$ splitCall(C)
    G_h.addEdge($caller$, C)
    G_h.addEdge($callee$, C)

2.2.2 根因定位

为了在异构传播图中进行根因定位，文章提出了一种新颖的方法，称为异构导向随机游走（Heterogeneity-Oriented Random Walk, HORW）。该方法充分考虑了异构性的特点，并在转移概率的计算上进行了创新。HORW 方法的核心思想是通过模拟在异构传播图中的随机游走过程来识别根本原因。与传统的随机游走方法不同，HORW 不仅考虑了图中的节点和边的结构信息，还结合了不同类型的节点（如服务节点、调用节点）和边（如服务间调用、调用节点之间的因果关系）的异构特性。通过对每种类型的边和节点赋予不同的权重，HORW 能够更加精确地捕捉性能问题的传播路径，从而准确定位到根本原因服务。该方法的创新之处在于转移概率的计算，即根据图中节点的类型和关联的边的性质来调整转移概率，而不是简单地依赖于图的结构。这使得 HORW 能够更有效地识别和定位性能问题的根源，尤其是在复杂和动态的微服务环境中。

三、最新发展、数据集与 SOTA 结果

3.1 MGDN

近年来，工业控制系统的异常检测研究逐渐从单领域分析转向多模态数据融合。传统方法如基于统计过程控制（SPC）和自回归模型（ARIMA）虽能捕捉单维度异常，但难以应对跨领域攻击（如同时篡改传感器数据与网络流量的协同攻击）。最新进展集中在图神经网络（GNN）与多任务学习的结合上，通过建模设备间的物理拓扑与网络交互关系提升检测能力。

论文采用 SWaT（Secure Water Treatment）数据集，该数据集由新加坡理工大学发布，模拟真实水处理系统的运行场景。数据集包含 51 个物理传感器（如流量计、阀门状态）和 16 个网络特征（如数据包数量、协议类型），时间跨度为 11 天，前 7 天为正常操作，后 4 天注入 36 种攻击（包括传感器欺骗、命令注入、网络泛洪等）。例如，攻击 A21 通过修改化学药剂投加量触发水质异常，而攻击 A35 则通过高频 Modbus 请求干扰控制指令。实验结果如表 1 所示。

表 1　MGDN 与基线方法的准确性（在数据集 SWaT 下）

Method	FPR (%)	Precision (%)	Recall (%)	F1 (%)
DTAAD	13.33	59.88	99.99	74.90
GDN	10.70	64.91	99.45	78.55
LSTM-AD	13.33	59.88	99.99	74.90
MAD-GAN	13.57	59.45	99.99	74.57
MSCRED	13.33	59.89	99.99	74.91
MTAD-GAT	13.39	59.78	99.99	74.83
OmniAnomaly	13.36	59.83	99.99	74.87
TranAD	13.35	59.85	99.99	74.88
USAD	13.26	60.02	99.99	75.01
MGDN	**3.07**	**84.65**	**85.12**	**84.88**

3.2 MicroDig

在评估 MicroDig 的性能时，文章构建了三个数据集，分别是来自腾讯的真实世界性能问题案例（数据集 $\mathcal{A}$）、开源微服务系统 Train-Ticket 中的注入问题（数据集 $\mathcal{B}$），以及中国建设银行电子商务系统的模拟问题（数据集 $\mathcal{C}$）。这些数据集涵盖了不同规模和类型的微服务系统，为全面评估 MicroDig 提供了丰富的场景。实验旨在回答三个研究问题：MicroDig 的诊断准确性（RQ1）、诊断效率（RQ2）以及核心组件对性能的贡献（RQ3）。

表 2　MicroDig 与基线方法的准确性（在数据集 $\mathcal{A}$ 下）

Method	AC@1 (%)	AC@2 (%)	AC@3 (%)	Avg@3 (%)	MRR
ServiceRank	50.8	55.7	57.4	54.6	0.55
MonitorRank	49.2	61.9	71.4	60.8	0.62
TraceRCA	61.5	72.7	75.8	70.0	0.70
TraceRank	16.9	20.3	20.3	19.2	0.20
Microscope	50.8	70.4	75.4	65.5	0.64
MicroHECL	61.9	73.8	76.2	70.6	0.71
MicroDig	**64.4**	**87.3**	**94.1**	**81.9**	**0.78**

文章选用了 Top-k 准确率（AC@k）、平均 Top-k 准确率（Avg@k）和平均倒数排名（MRR）作为主要的性能评估指标。这些指标能够全面反映 MicroDig 在定位性能问题根源时的准确性和效率。为了进行公平比较，文章选择了六种具有代表性的基线方法，包括 Microscope、ServiceRank、MicroHECL、MonitorRank、TraceRCA 和 TraceRank，这些方法在各自的领域都展现出了优越的性能。

实验结果表明，MicroDig 在所有三个数据集上都显著优于基线方法。如表 2 所示，在腾讯的真实世界数据集上，MicroDig 在 AC@1、AC@2 和 AC@3 上分别比其他方法高出 4%、18.3% 和 23.5%。此外，MicroDig 的平均诊断时间最短，为 24.72 秒/案例，显示出良好的效率。消融实验进一步证实了异构传播图、超参数 β 和异常检测对 MicroDig 性能的显著贡献。

四、结论

工业控制系统与微服务架构的异常检测与根因分析是保障现代工业与互联网应用稳定性的核心问题。两篇论文通过跨领域图表示学习与潜在空间干预识别，为解决复杂系统中的多维度、异构性问题提供了创新方案。

ICS 异常检测模型 MGDN 通过物理域与网络域数据的联合建模，克服了单领域分析的局限性。注意力机制与多任务优化的结合，使模型在保留领域特征的同时捕捉跨域关联。

文章[4]在微服务系统性能诊断领域做出了重要贡献。它不仅提供了一种新的根因定位方法，还通过实际案例展示了该方法的有效性。尽管存在一些局限性，但文章也已经详尽写出，这位未来的研究工作进行了初步的规划和奠定了基础。MicroDig 的出现，无疑为微服务系统的运维和性能优化提供了新的工具和思路，对于提高系统的可靠性和用户体验具有重要意义。此外，这项研究也为学术界和工业界提供了新的研究方向，特别是在微服务架构日益普及的背景下，如何有效地诊断和解决性能问题，成为了一个亟待解决的挑战。MicroDig 的成功

应用，为这一挑战提供了一个有力的解决方案，也为未来的研究和实践提供了宝贵的经验和启示。

五、参考文献

[1] ZHAN D, ZHANG W, YE L, et al. Anomaly detection in industrial control systems based on cross-domain representation learning[J]. IEEE Transactions on Dependable and Secure Computing, 2024.

[2] MANDIC D P, CHAMBERS J. Recurrent neural networks for prediction: learning algorithms, architectures and stability[M]. John Wiley & Sons, Inc., 2001.

[3] CRESWELL A, WHITE T, DUMOULIN V, et al. Generative adversarial networks: An overview[J]. IEEE signal processing magazine, 2018, 35(1): 53-65.

[4] TAO L, LU X, ZHANG S, et al. Diagnosing performance issues for large-scale microservice systems with heterogeneous graph[J]. IEEE Transactions on Services Computing, 2024.

[5] DENG A, HOOI B. Graph neural network-based anomaly detection in multivariate time series[C]//Proceedings of the AAAI conference on artificial intelligence: Vol. 35. 2021: 4027-4035.

[6] DÉSIDÉRI J A. Multiple-gradient descent algorithm (mgda) for multiobjective optimization[J]. Comptes Rendus Mathematique, 2012, 350(5-6): 313-318.