

图平台在金融风控中的应用实践

2023.08

目录

CONTENTS

01

相关背景介绍

02

图平台产品介绍

03

应用案例

04

未来规划

PART I 第一部分

相关背景

可用数据、图平台需求、挑战、平台架构



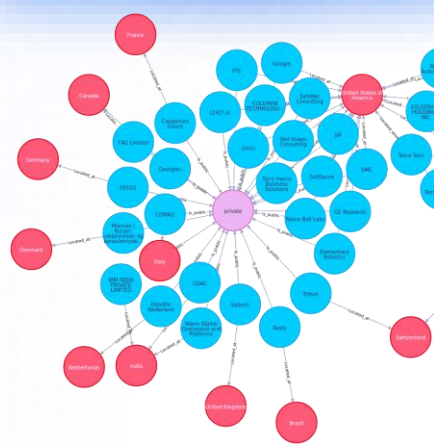
金融行业丰富的图数据



社交网络



工商信息



百亿节点



人



企



地址

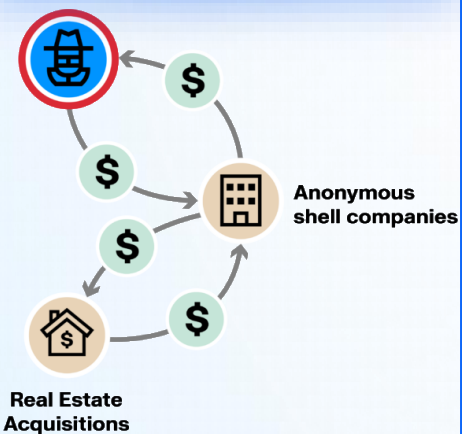


设备



网络

资金流向



通讯网络



千亿节点



人-(企)-人



人-(地)-人



人-(交易)-人



人-(设备)-人

不同用户及业务对图需求的多样性



智能风控

自动数据预处理

大规模分布式图计算

智能反欺诈

社团信息查看

经营地识别

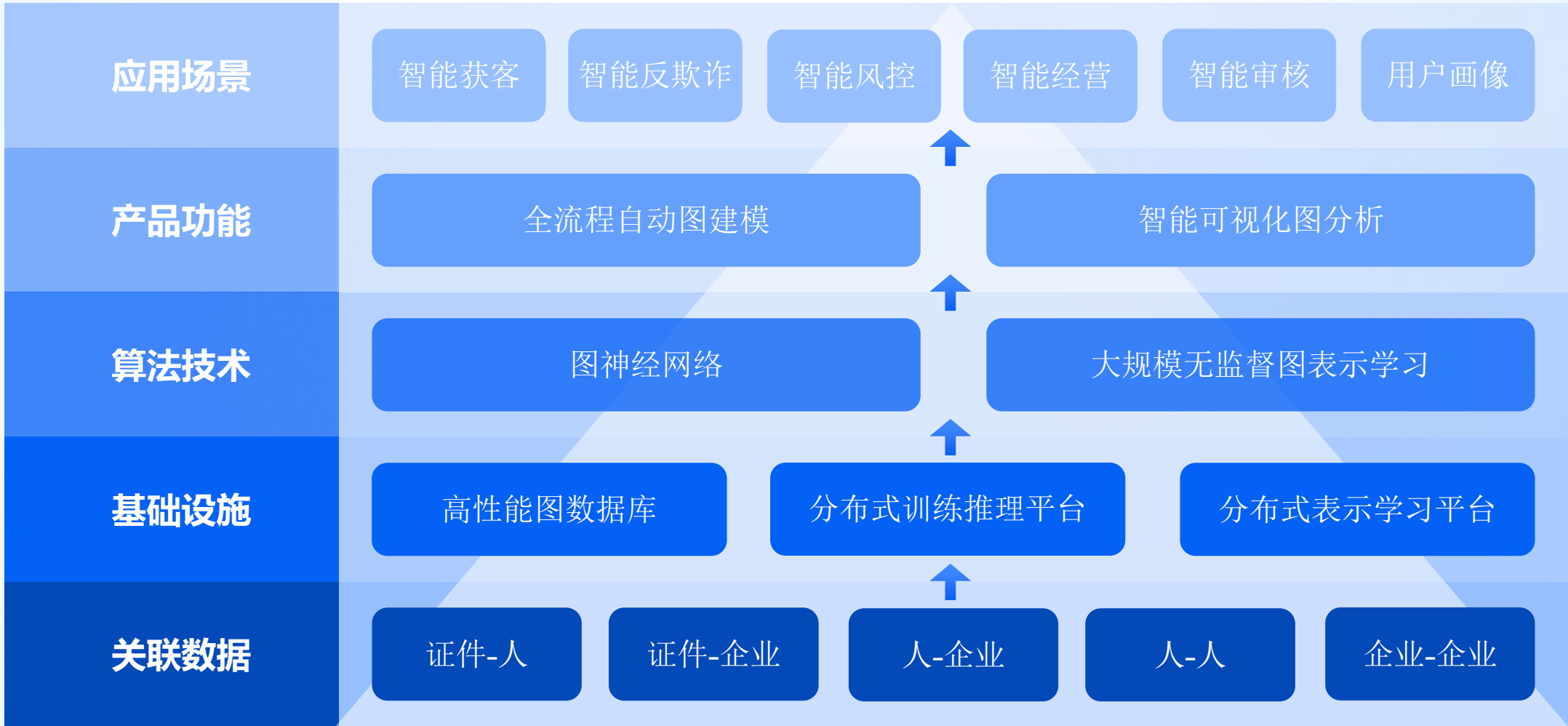
智能审核

人企关联审核

企业上下游关联审核

图平台面临的挑战





PART II 第二部分

图平台产品介绍

平台架构、可视化功能



效果

分布式关联数据与宽表特征
存储平台

支持**千亿**数据规模
上千特征维度关联数据存储

高性能效率的
图查询、导入导出接口

毫秒级多跳查询
百亿数据**小时级**入库

集成图计算能力
为业务提供基础分析工具

15类图分析算法
分析效率提升**10倍+**

方案

- 海量数据规模
- 存储平台化

分布式图推理

- 优化点边索引
- 分布式建库

在线图推理

- 图分析算法
- 分布式计算

图分析能力



➤ 图采样占据大量训练时间

➤ 分布式预采样

➤ GPU加速图采样

➤ 采样速度 \uparrow 40倍

➤ 采样时间 \downarrow 80%

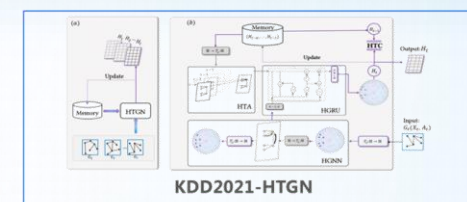
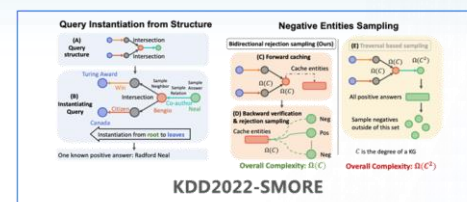
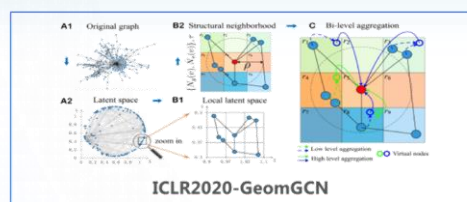
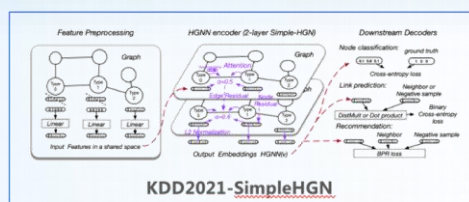
➤ 分布式分割算法决定训练规模上限

➤ 线性分布式图分割映射表

➤ 优化异构图分割格式转换与ID映射

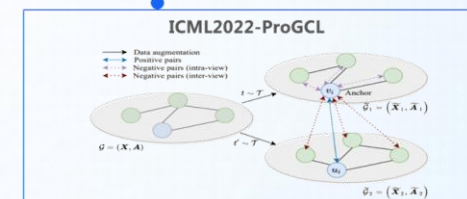
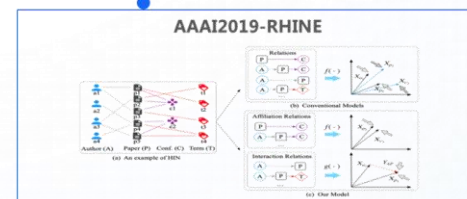
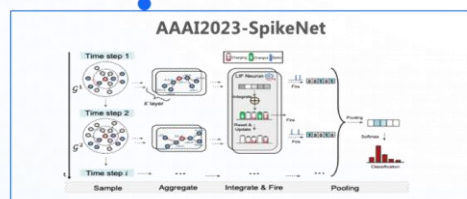
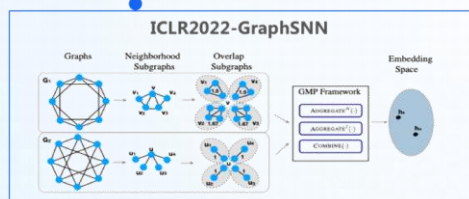
➤ 支持千亿以上规模大图训练

➤ 支持百亿以上规模异构图训练



同构图 ➤ 异构图 ➤ 时序图 ➤ 几何图

网络表示 ➤ 知识表示 ➤ 图对比学习 ➤ 动态图表示



图神经网络算法

无监督图表示学习算法



批量图推理

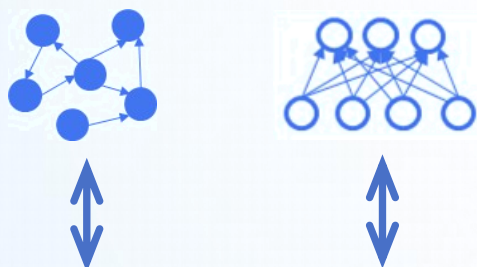
计算规模大
计算效率低

➤ 分布式图推理

10分钟内完成亿次推理计算

➤ GPU加速

统一平台，任务分钟级部署



训练

推理

在线图推理

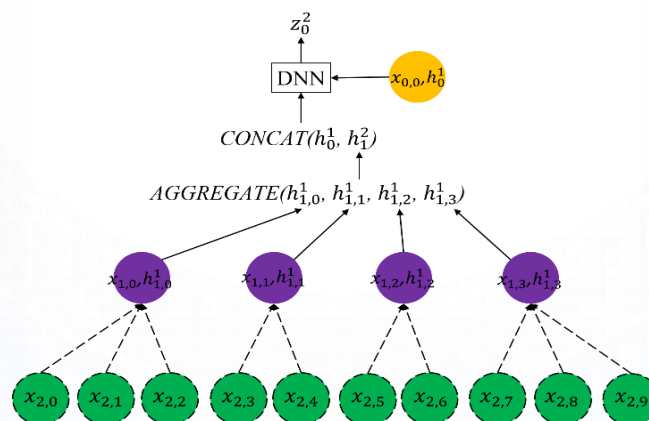
GNN推理速度慢
计算时效低

➤ 多层聚合采样

多层图神经网络毫秒级计算延迟

➤ 时效性补偿

子图采样更新达到秒级时效水平



计算一致性

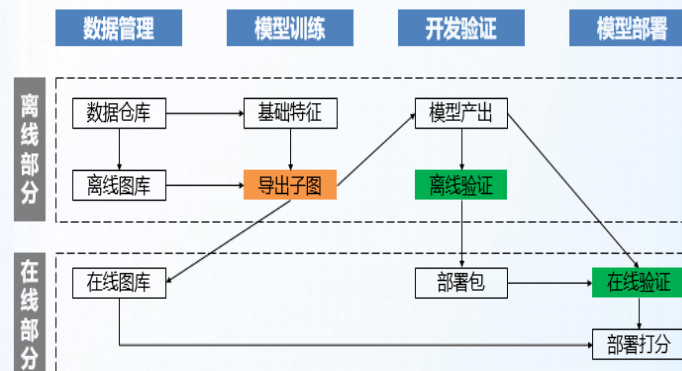
模型验证与质量保障

➤ 训练推理一致性

所训即所得，保障模型上线效果

➤ 在离线模型一致性

多场景模型打分计算结果保障





关联风险分析

用户等待时间长体验差

PART III 第三部分

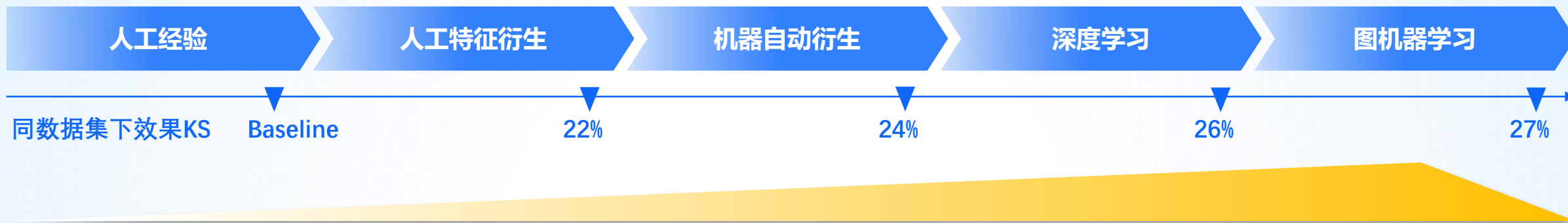
应用案例

应用于征信报告挖掘案例、分阶段算法优化方案介绍

图机器学习应用于征信解读，获吴文俊人工智能科技进步奖



征信解读的五个发展阶段



图算法进展和效果

阶段1：单报告内部关系图

以查询/借还款行为和机构为节点构图，捕捉同样行为在不同关系场景下的不同含义



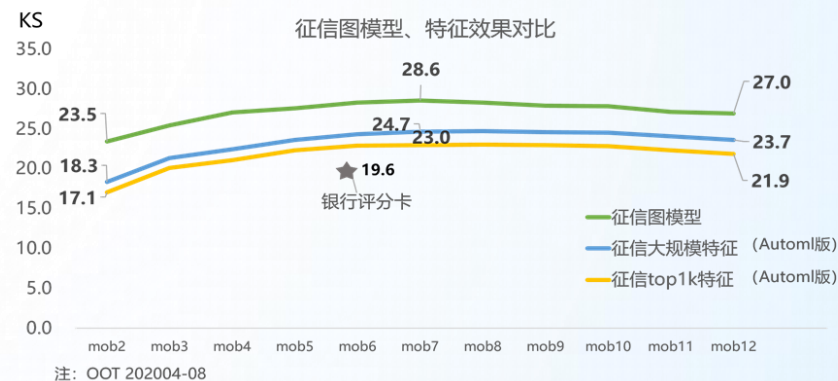
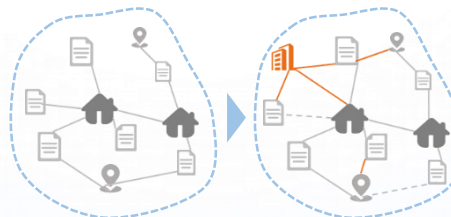
阶段2：多个报告之间关系图

利用工作单位、居住地址信息，结合股权关系等外部数据，将不同征信报告关联成一张图



阶段3+动态图、时序图

动态预测节点之间的关系，图网络结构不断演进，节点与边的属性实时更新



图模型效果超大规模特征，特征效果超银行评分卡

征信报告变成一张图谱



个人征信报告信息



节点



边

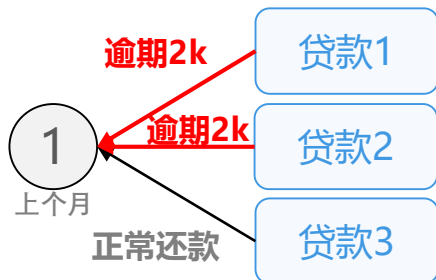


主要问题

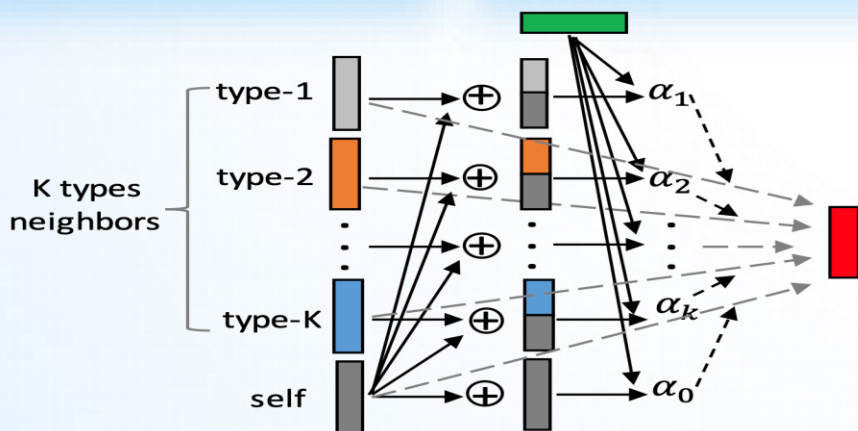
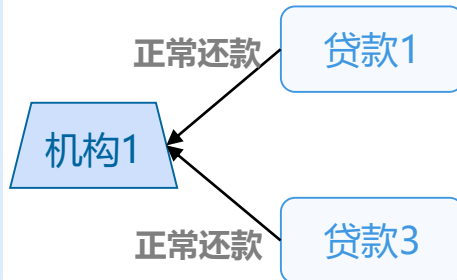
征信报告的非结构化特性导致难以高效挖掘有用的风险信息

时间和机构视角的风险自适应聚合

上个月用户的风险?

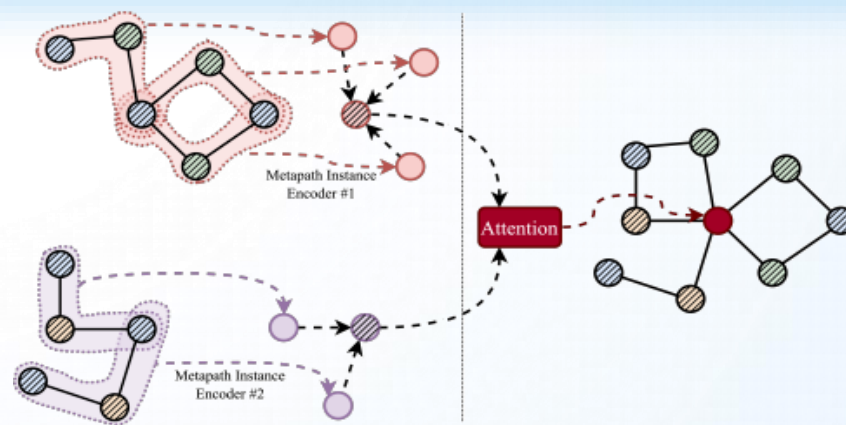
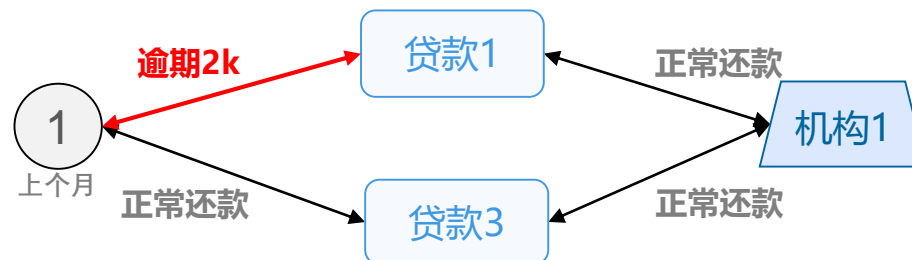


XX机构用户的风险?



用户行为风险自适应聚合

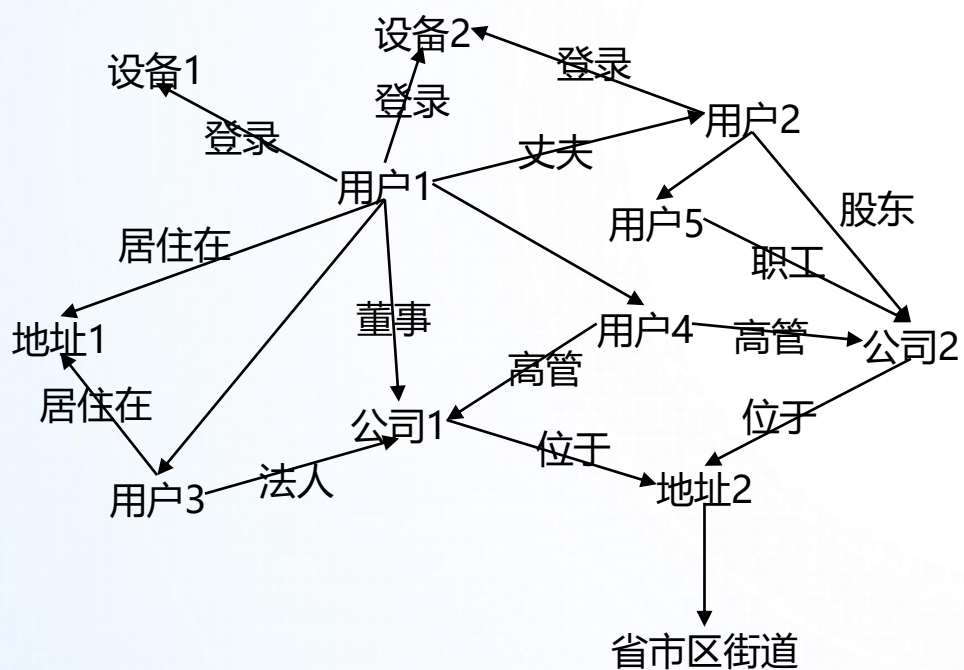
上个月用户的某个行为会带来多少风险?



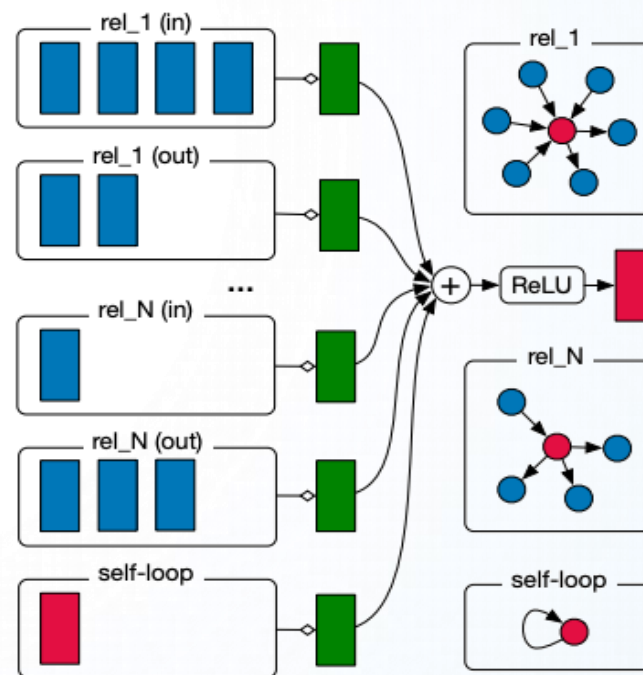


主要问题 征信报告只能反应用户自身的信用行为所带来的风险，无法捕获其他用户对用户的风险影响

结合征信信息与外部信息建立异构大图



基于Attention机制的多关系大图模型

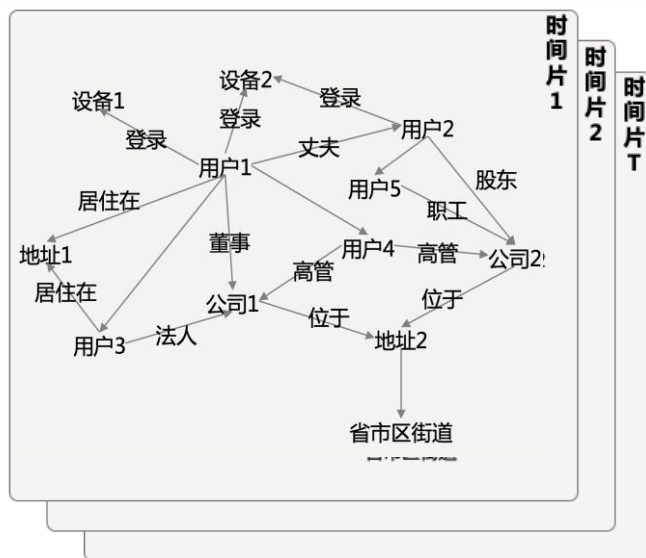




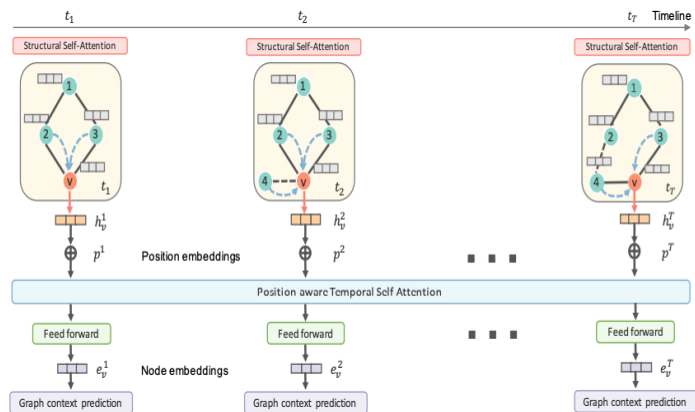
主要问题

征信知识图谱无法及时感知到“环境变化”，难以处理用户风险随时间动态漂移的问题

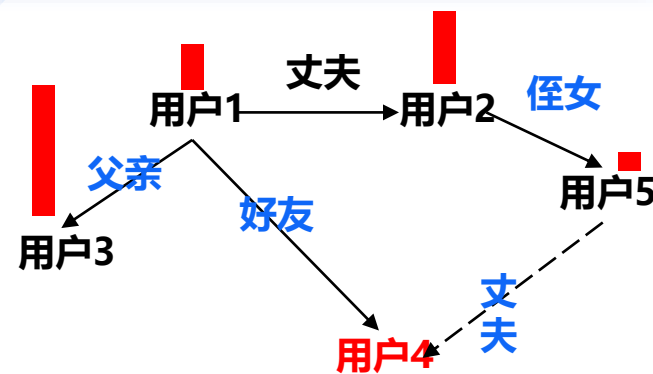
时序大图



动态知识表示



关系推断与风险预测



逾期风险大小?

节点特征为征信图子分值越大逾期风险越高

PART IV 第四部分

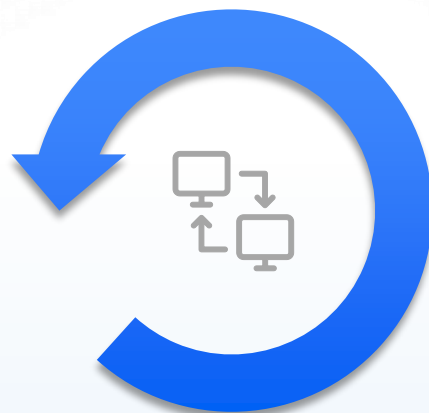
未来规划





更高效的基础服务

- 实时/时序图库
- 通用时序图神经网络建模



更精深的算法技术

- 动态图预训练
- 大图-征信图联合优化



探索与大模型结合

- 大图强化学习预训练
- 面向风险的图Prompt设计



1. Relation Structure-Aware Heterogeneous Information Network Embedding, AAAI 2019
2. Scaling Up Dynamic Graph Representation Learning via Spiking Neural Networks, AAAI 2023
3. Geom-GCN: Geometric Graph Convolutional Networks, ICLR 2020
4. A New Perspective on “How Graph Neural Networks Go Beyond Weisfeiler-Lehman?” , ICLR 2022
5. ProGCL: Rethinking Hard Negative Mining in Graph Contrastive Learning, ICML 2022
6. Are we really making much progress? Revisiting, benchmarking, and refining heterogeneous graph neural networks, KDD 2021
7. SMORE: Knowledge Graph Completion and Multi-Hop Reasoning in Massive Knowledge Graphs, KDD 2022
8. Discrete-time Temporal Network Embedding via Implicit Hierarchical Learning in Hyperbolic Space, KDD 2021
9. Heterogeneous Graph Neural Network, KDD 2019
10. MAGNN: Metapath Aggregated Graph Neural Network for Heterogeneous Graph Embedding, KDD 2020
11. Modeling Relational Data with Graph Convolutional Networks, ESWC 2018
12. Dynamic Graph Representation Learning via Self-Attention Networks, LoG 2019

感谢聆听

**Thanks for
listening**

2023.08