图神经网络在实时风控中的应用

讲师简介

ebay msup

韩志超
高级机器学习工程师

eBay Payments and Risk 机器学习工程师，负责多个领域风险评估，包括实时交易，共谋和合规风险。入职 eBay 前有多年金融行业从业经验，包括对冲基金和投资银行。专注于时间序列决策问题。

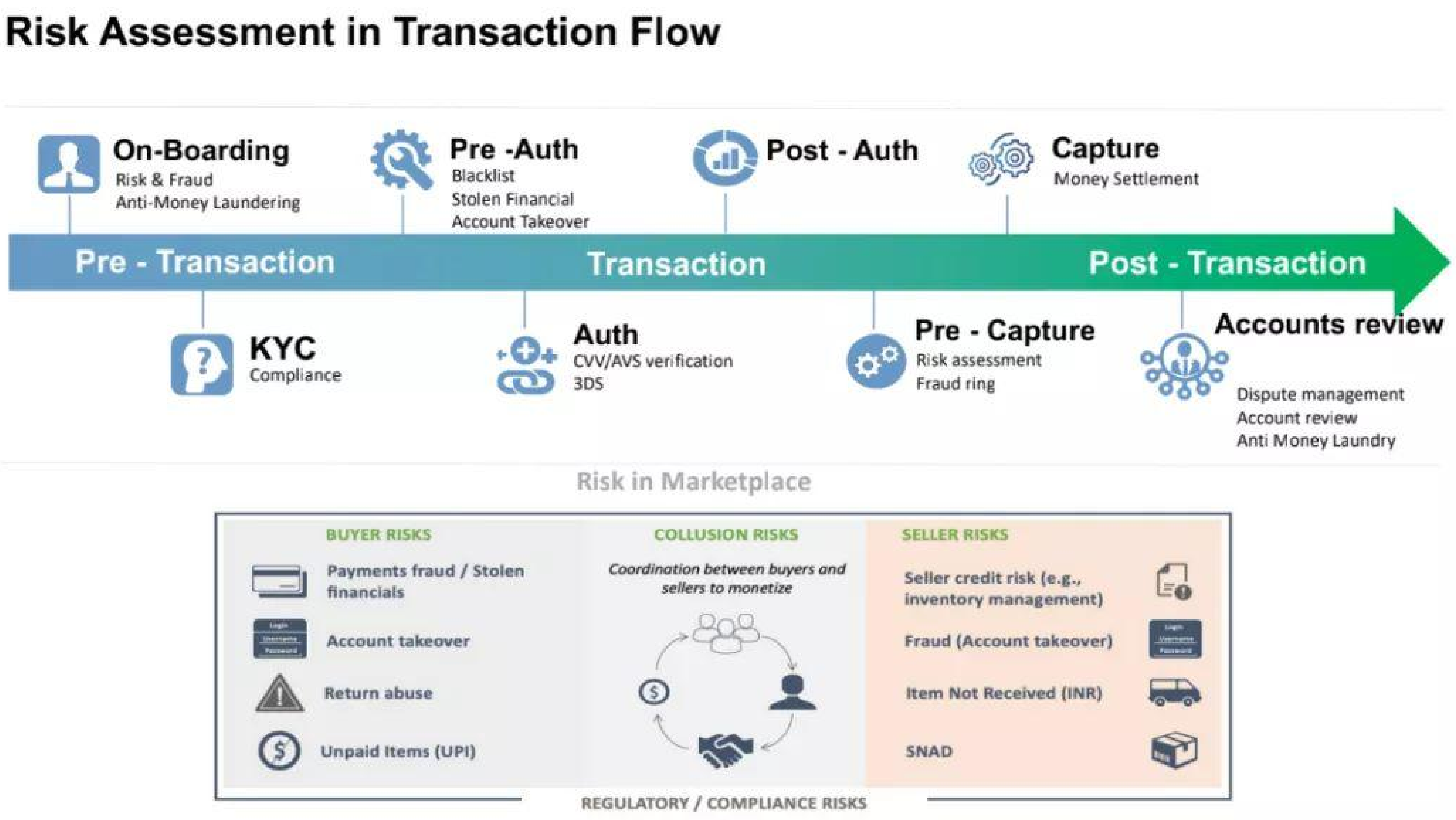
目录

1. 实时交易反欺诈场景介绍
2. 图神经网络在实时场景的挑战
3. 端到端的解决方案
4. 小结和展望

案例简介

- 实时交易的风险检测对模型推断的延迟要求较高
- 利用有向时序异构图方式构建动态图
  - 解决实验阶段信息穿越问题
  - 提高在线推断速度

实时交易反欺诈场景介绍



实时交易反欺诈场景介绍

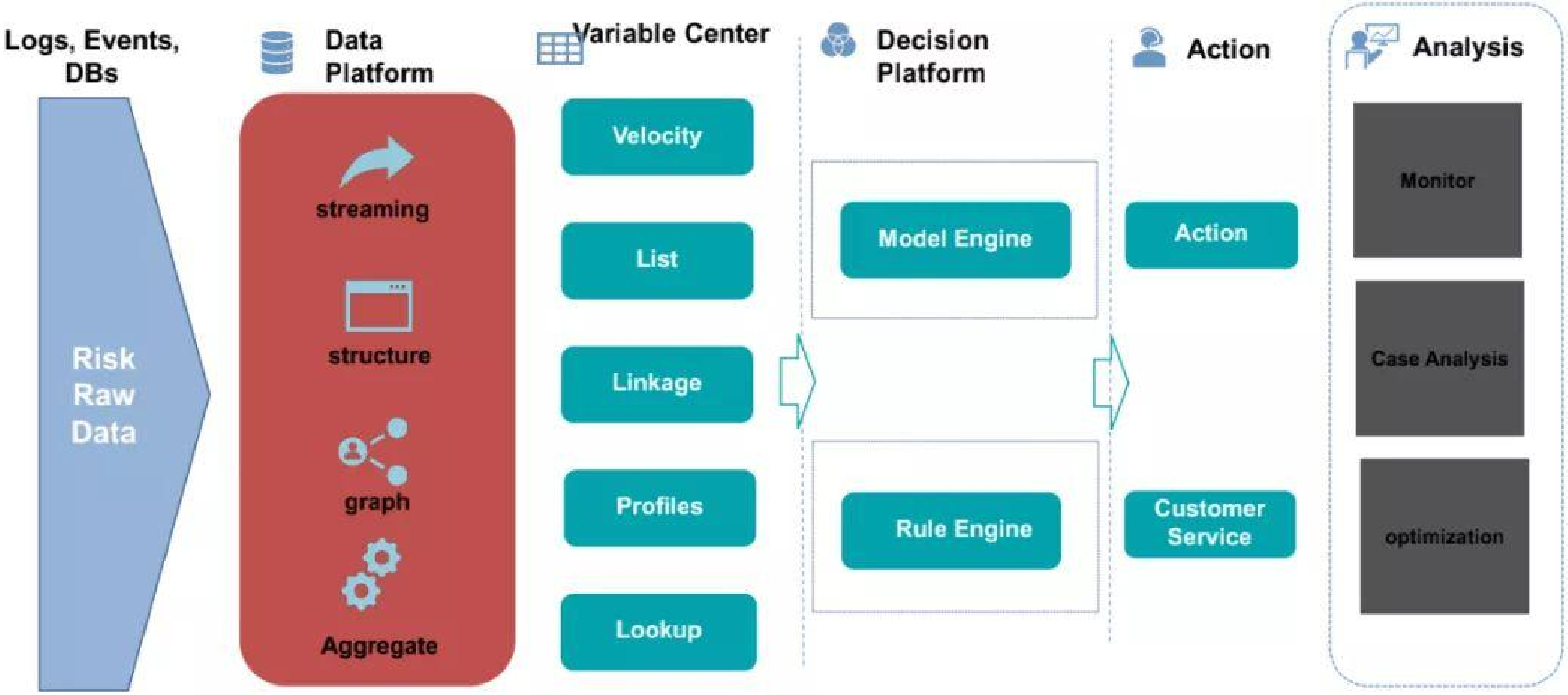
预授权评估	
盗号评估	盗卡评估

支付渠道评估
信息核对

交易评估	
是否标记（共谋）	是否拒绝（未授权交易）

实时交易反欺诈场景介绍

Transaction Risk Platforms



实时交易反欺诈场景介绍

Real-time Risk Evaluation: Supervised Machine-learned Models for Account Takeover

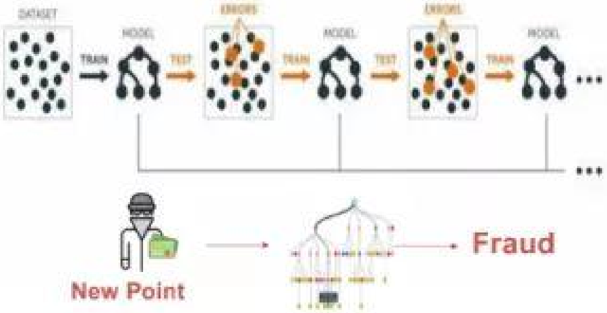
What is Account Takeover (ATO)?

- Account takeover fraud is a form of identity theft in which the fraudster gets access to a victim's eBay accounts
- A successful account takeover attack leads to fraudulent transactions and unauthorized shopping from the victim's compromised account

Data	• Gather data based on historical transactions • **Bad** (Unauthorized Claim with ATO Tag) or **Good** (No claim)
Ingestion	• Streaming • Batch
Data Processing	• Unified & **real-time** risk variable extraction
Model Training	• State-of-the-art classification model
Deployment	• **Predict** the possibility of ATO in real-time during transaction
Live System	• Model prediction affects decision of **Approve /Surface remedy / Decline** during the check out flow

Supervised Machine Learning

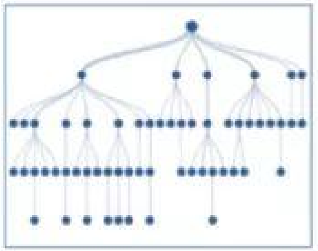
- Treat accounts/sample independently
- Capture known risk issue with good performance

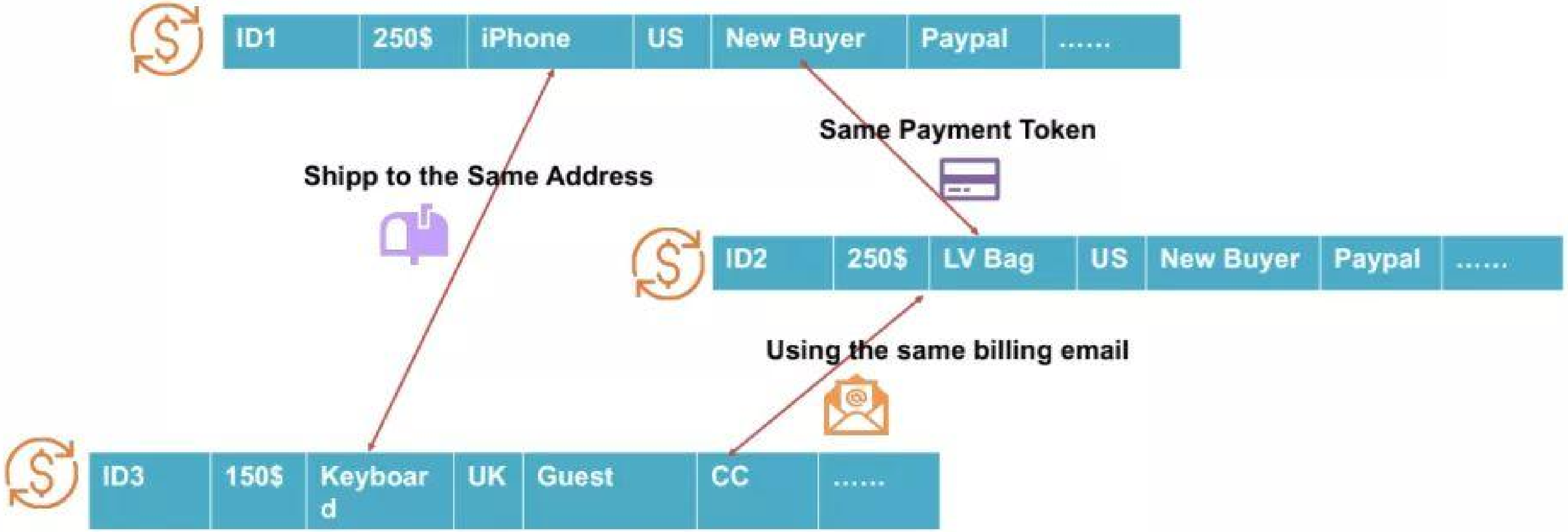


实时交易反欺诈场景介绍

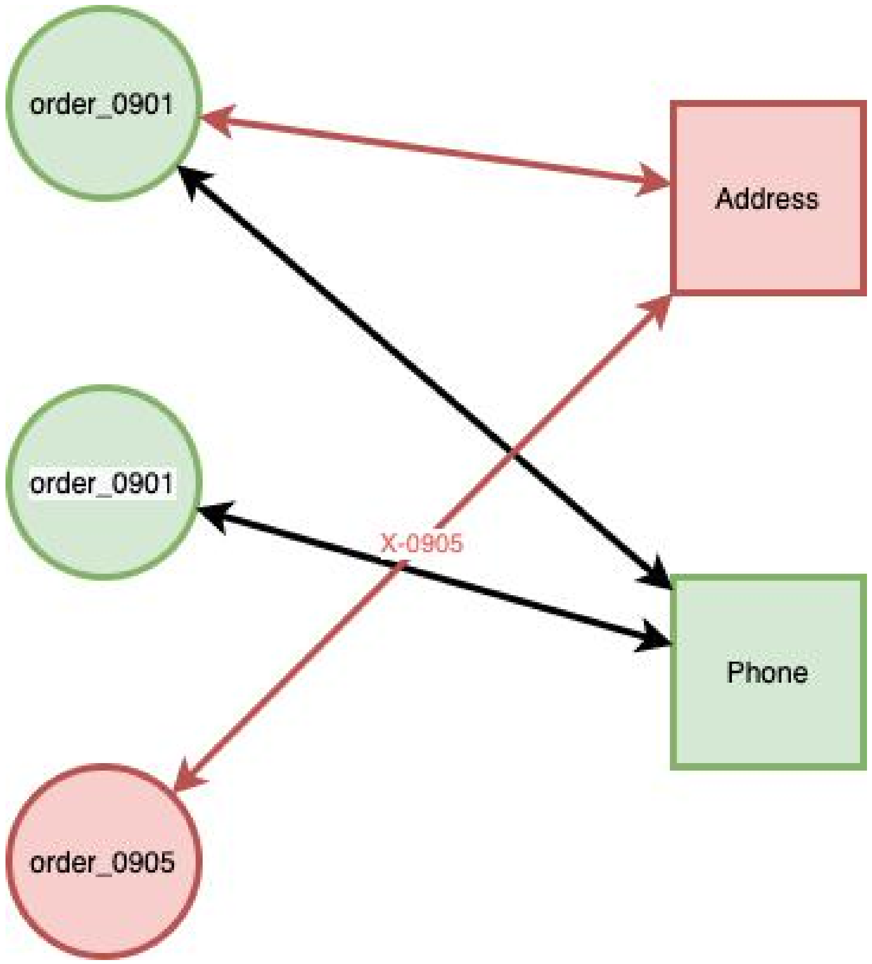
Current Machine Learning Model is not a powerful representation of context information!

ID1	250$	iPhone	US	New Buyer	Paypal	……
ID2	250$	LV Bag	US	New Buyer	Paypal	……
ID3	150$	Keyboard	UK	Guest	CC	……



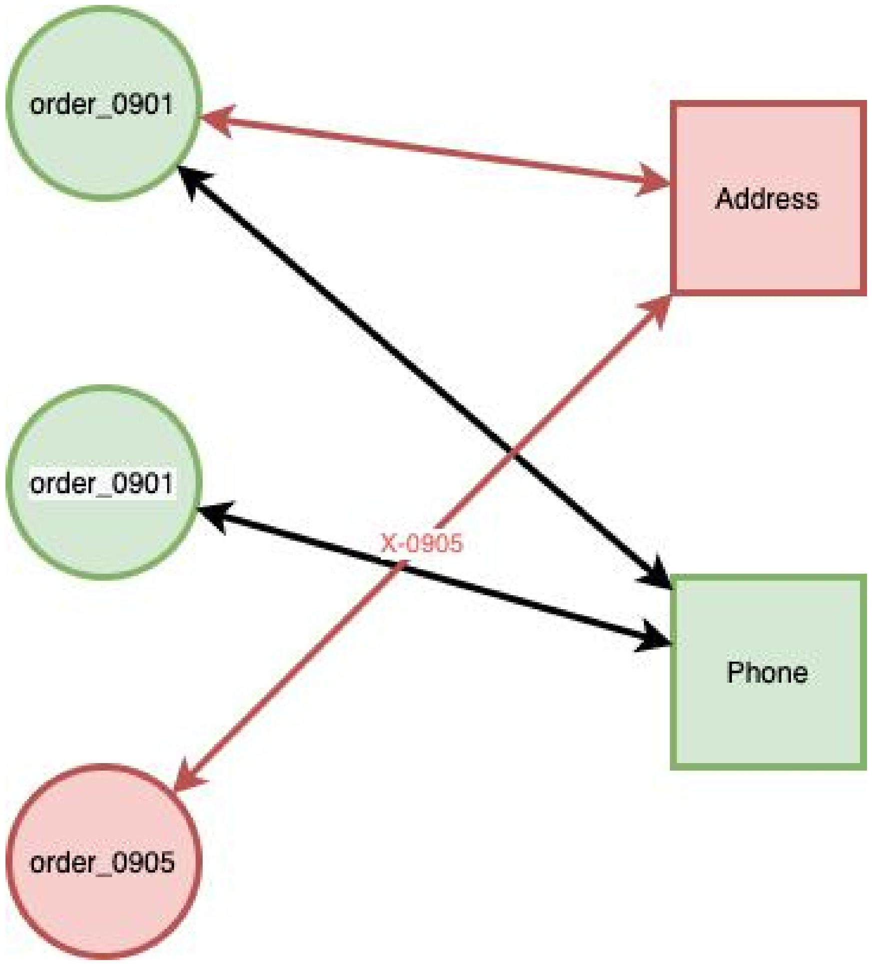


图神经网络在实时场景的挑战 – 二部图



- 事件节点和关联节点
- 避免了同质图的冗余
- 特征关联事件

图神经网络在实时场景的挑战 – 构图的时效性



关联节点涉及多个时间点

实例特征融合后可能有未来信息

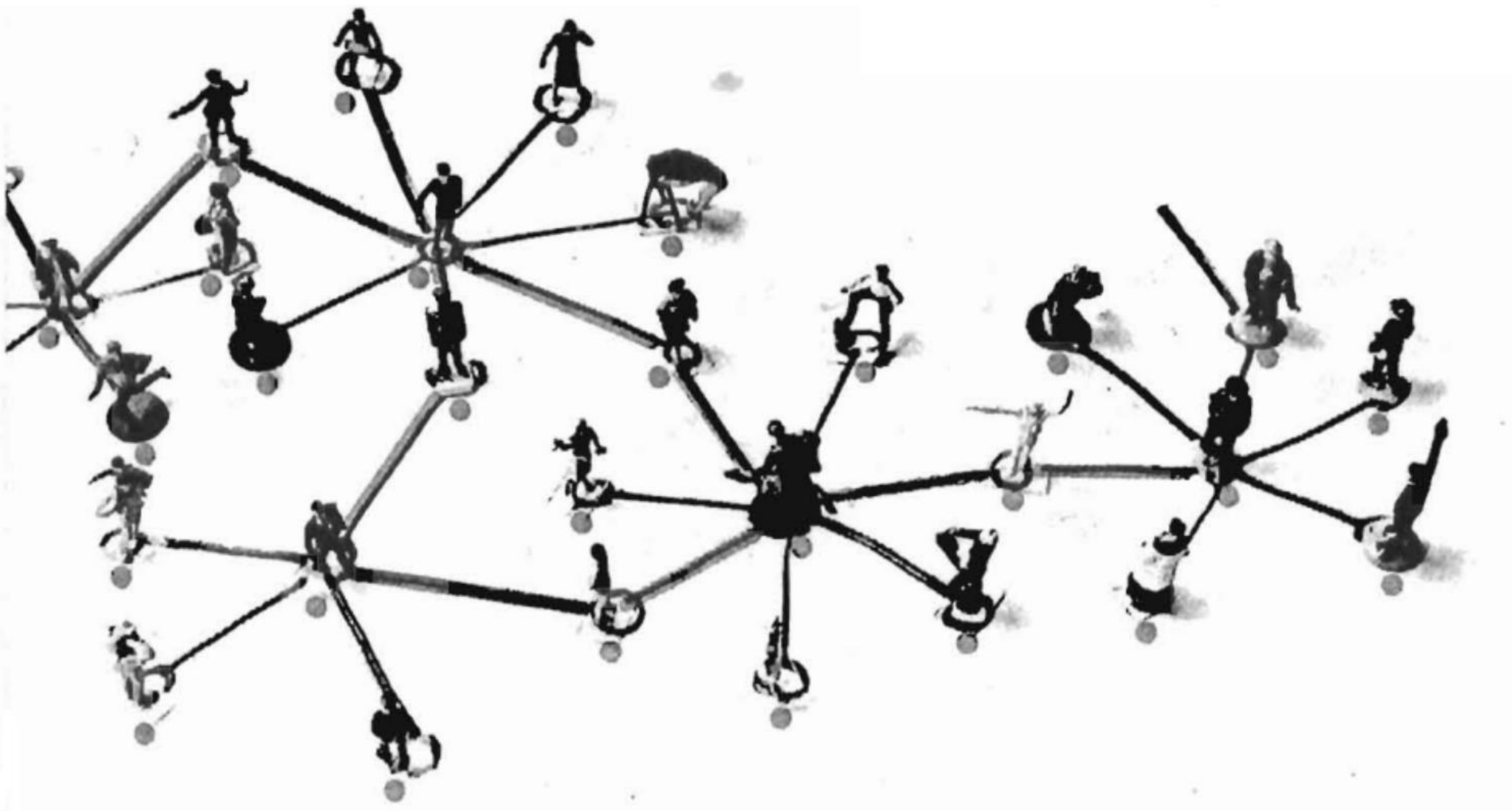
模型训练时具备推断时不可能有的“预测力”

图神经网络在实时场景的挑战 – 在线推断的延迟

图查询延迟

关联特征的拼接

在线推断



图神经网络在实时场景的挑战 – VS GBDT

- Dataset: House insurance
  - Feature – 66 columns
    - Boolean 7
    - Integer 18
    - String 11
    - Other 7
  - Label - Resiliated
- Winner - XGBoost
  - Best metric in test
  - Efficient in Training Time
  - More Understandable

	ROC AUC	F1 score	Time (sec)
XGBoost	0.7706	0.5591	500
MLP	0.7514	0.5458	184
TabNet without pretrain	0.7579	0.5529	1464
TabNet with pretrain	0.7524	0.5484	2370

Source: https://www.kaggle.com/kyosukemorita/deep-learning-vs-gbdt-model-on-tabular-data/notebook

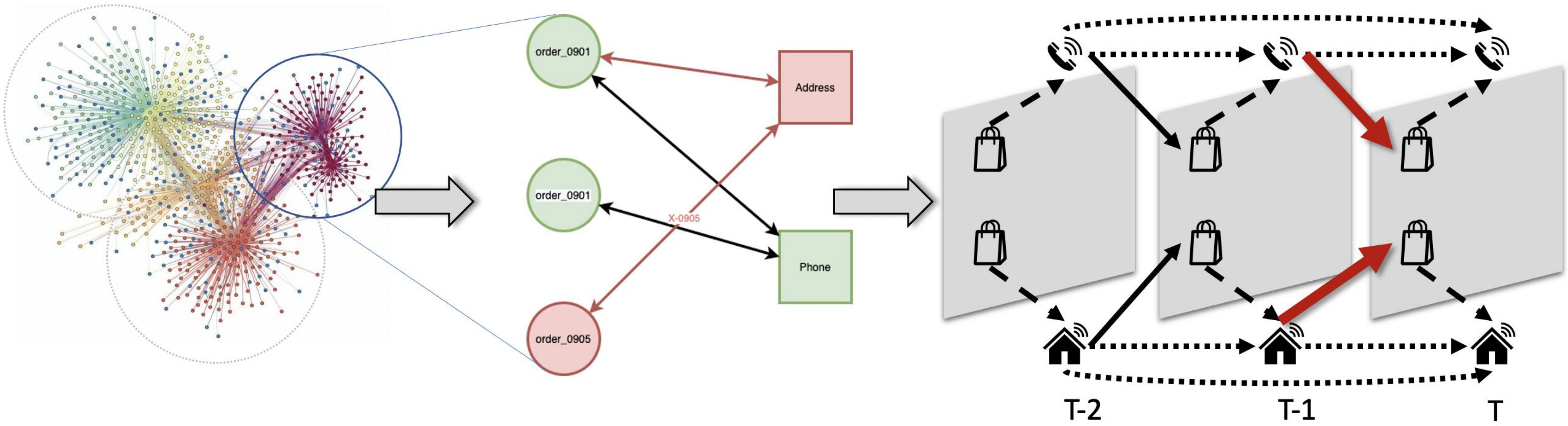
构图的时效性

- 有向动态切片图 Directed Dynamic Snapshot Graph

推断低延迟

- Lambda 架构的网络结构

端到端方案 – 有向动态切片图

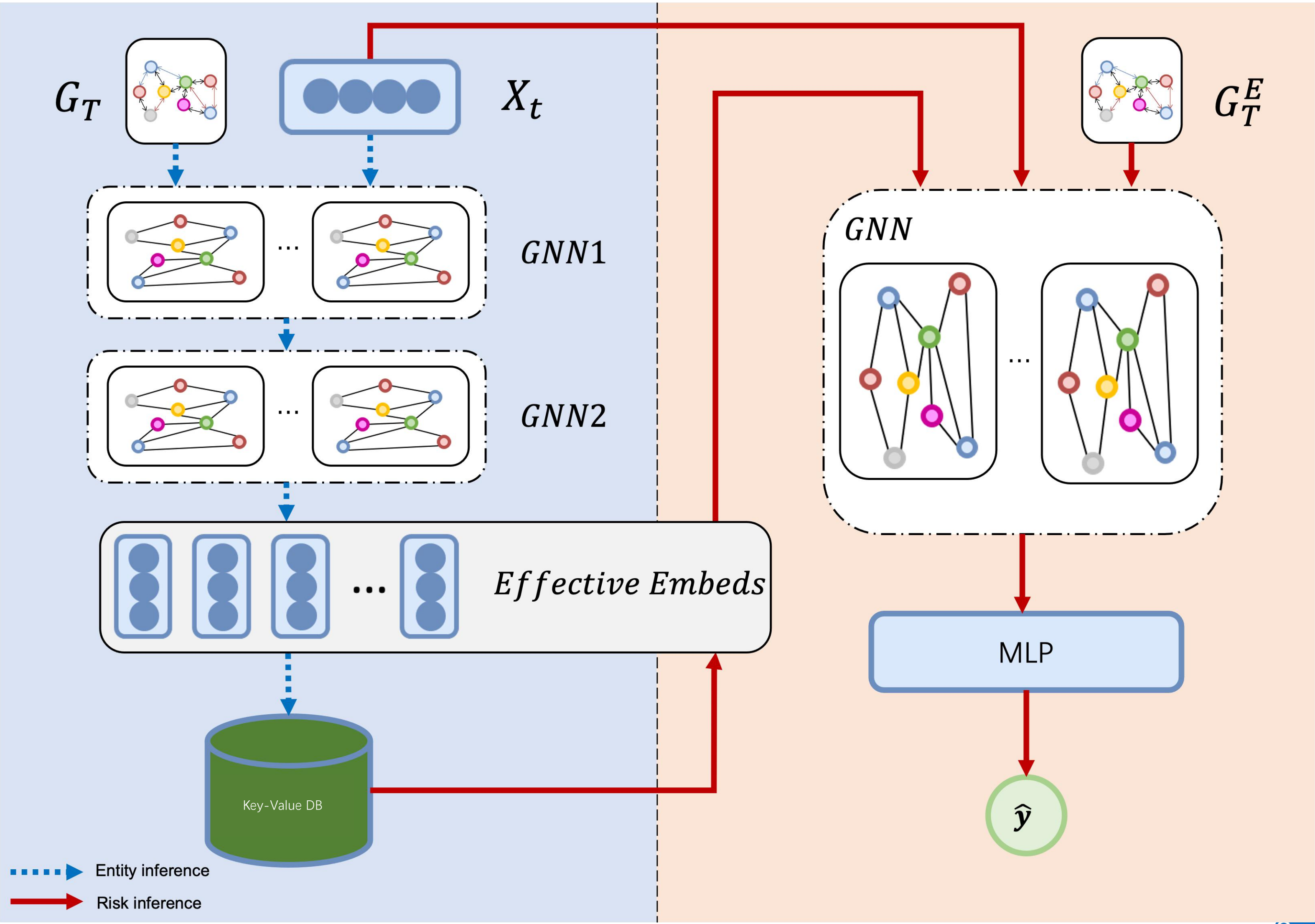


端到端方案 – 有向动态切片图

- DDSL

  Directed Dynamic Snapshot Lambda

- 运用在分区切片图上
  - 类ClusterGCN
- 堆叠的通用GNN层
  - 类DeepGCN
- 推断的最后一跳
  - 类双塔模型



Source: Graph Neural Networks in Real-Time Fraud Detection with Lambda Architecture. *arXiv*

离线学习

分区构图	编码学习

在线推断

实体表述	订单评估

离线学习

分区构图	编码学习

在线推断

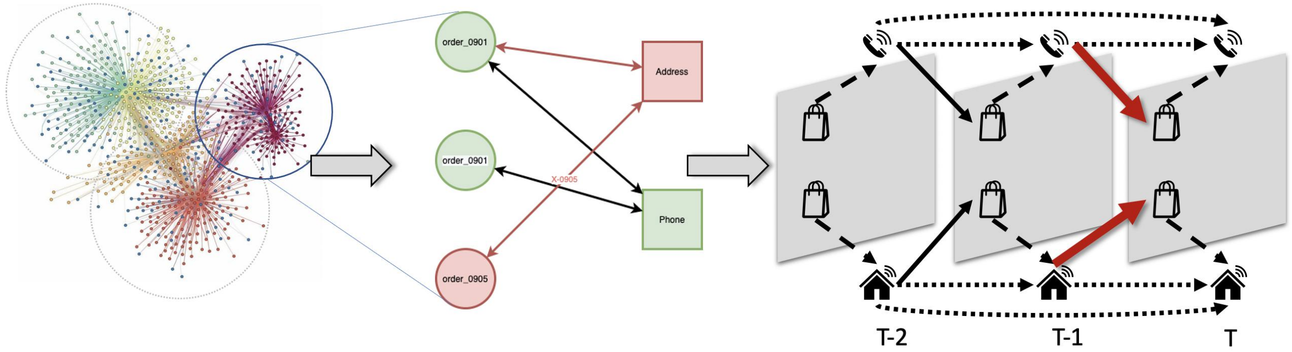
实体表述	订单评估

离线学习

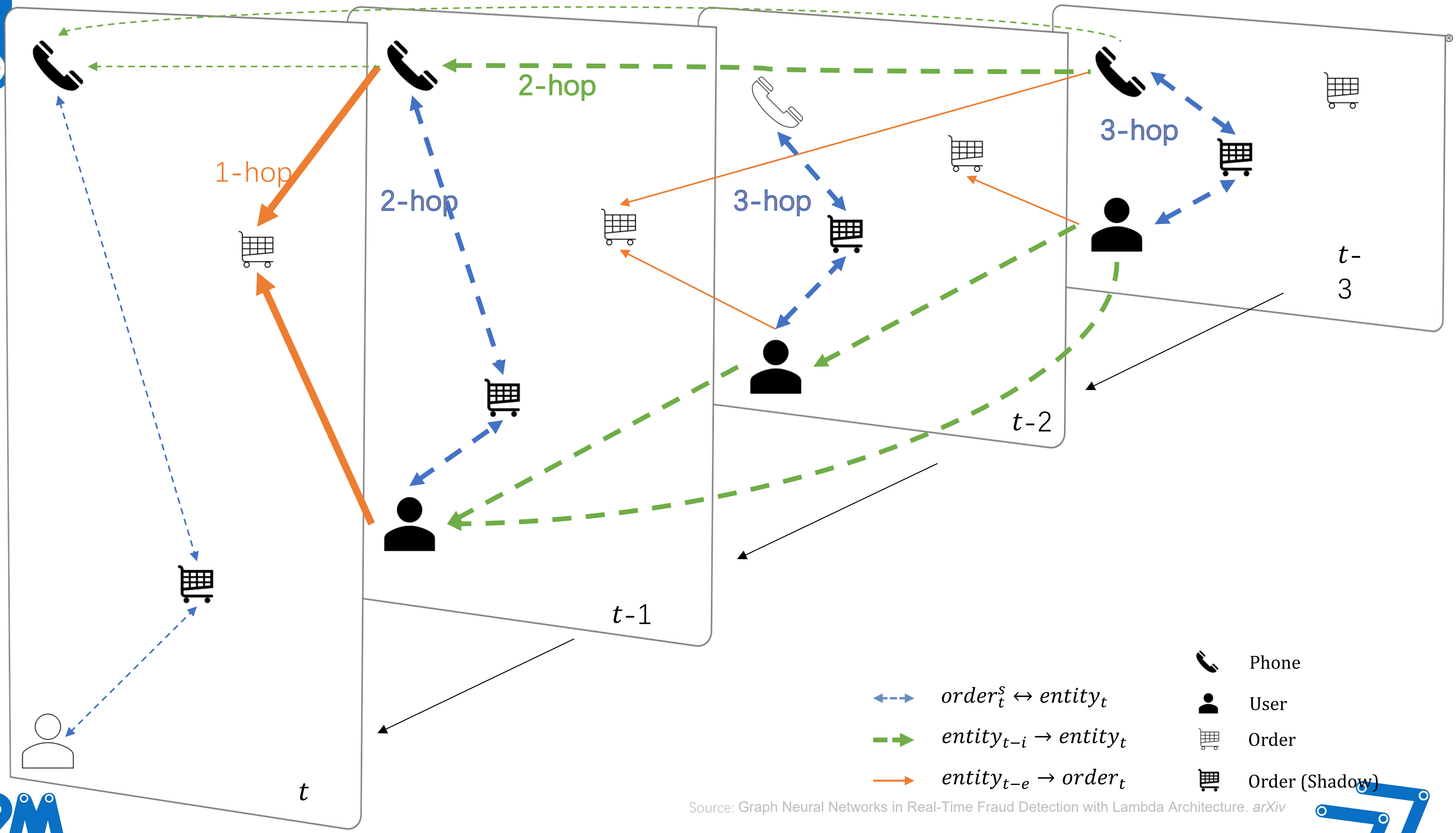
分区构图 | 编码学习

在线推断

实体表述 | 订单评估



- $Order_t^s \rightarrow Entity_t$
  - 实体表述
- $Entity_{t-i} \rightarrow Entity_t$
  - 实体表述
- $Entity_{t-e} \rightarrow Order_t$
  - 订单评估



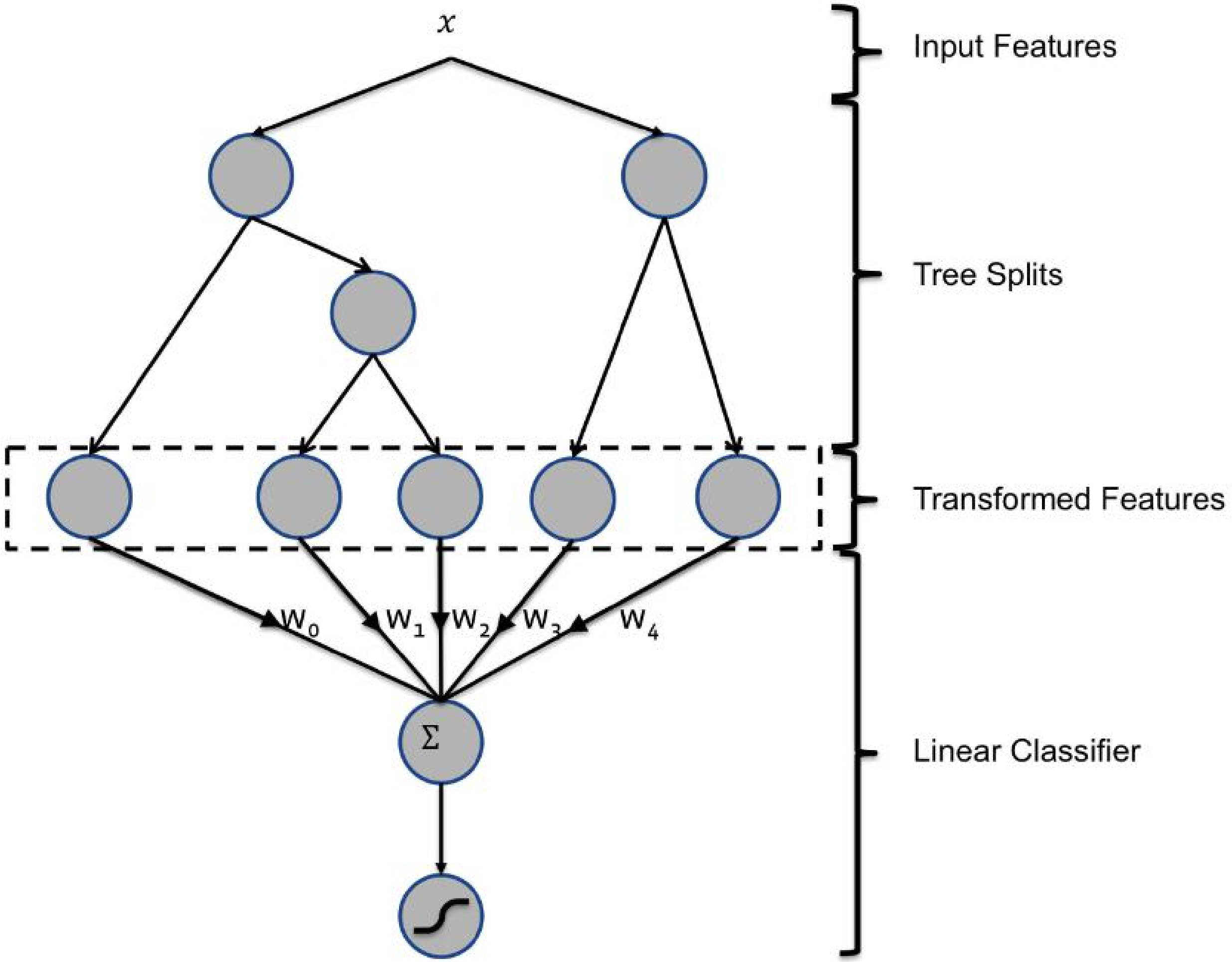
Source: Graph Neural Networks in Real-Time Fraud Detection with Lambda Architecture. *arXiv*

离线学习

分区构图 | 编码学习

在线推断

实体表述 | 订单评估



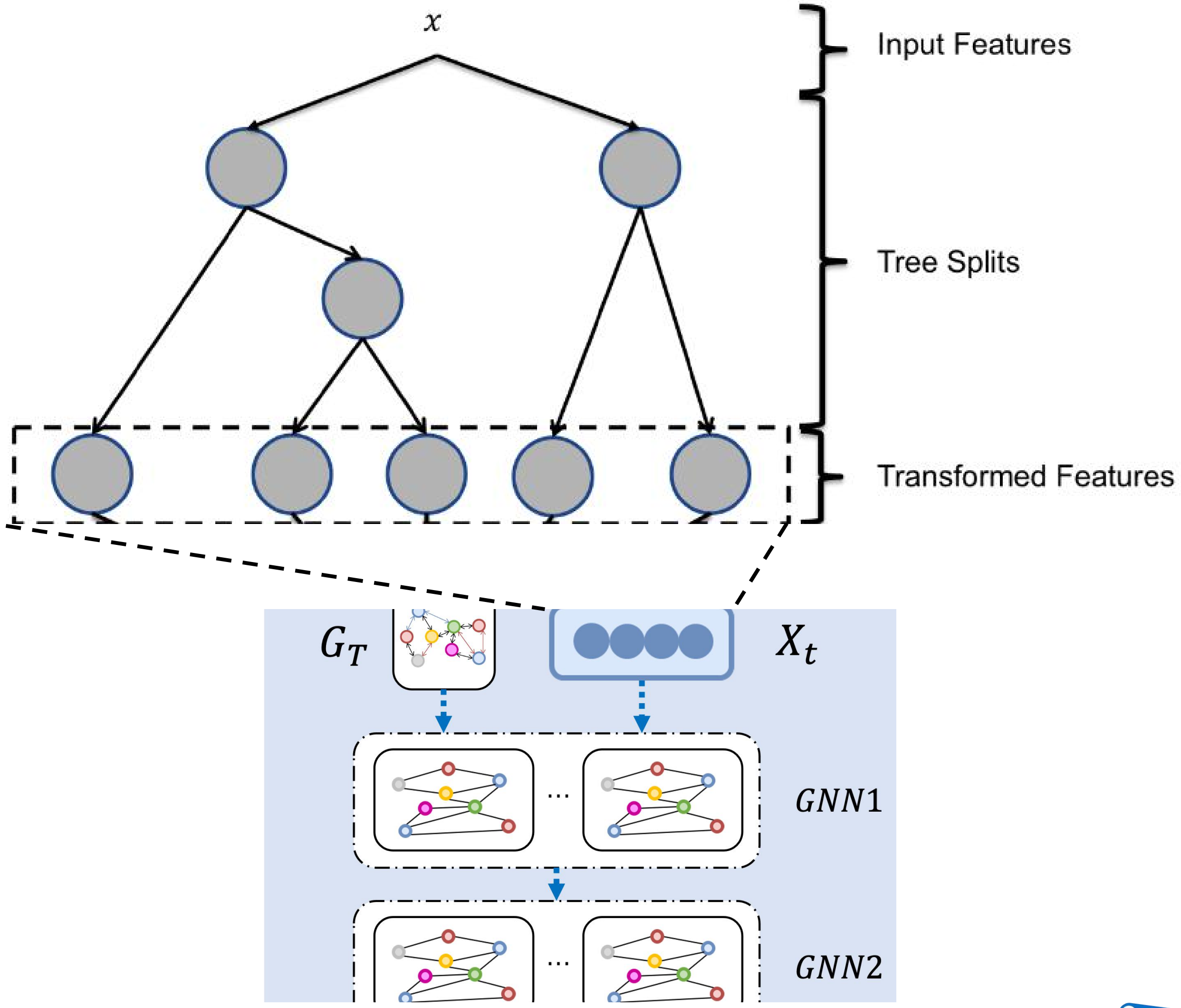
Source: Practical Lessons from Predicting Clicks on Ads at Facebook

离线学习

分区构图 | 编码学习

在线推断

实体表述 | 订单评估



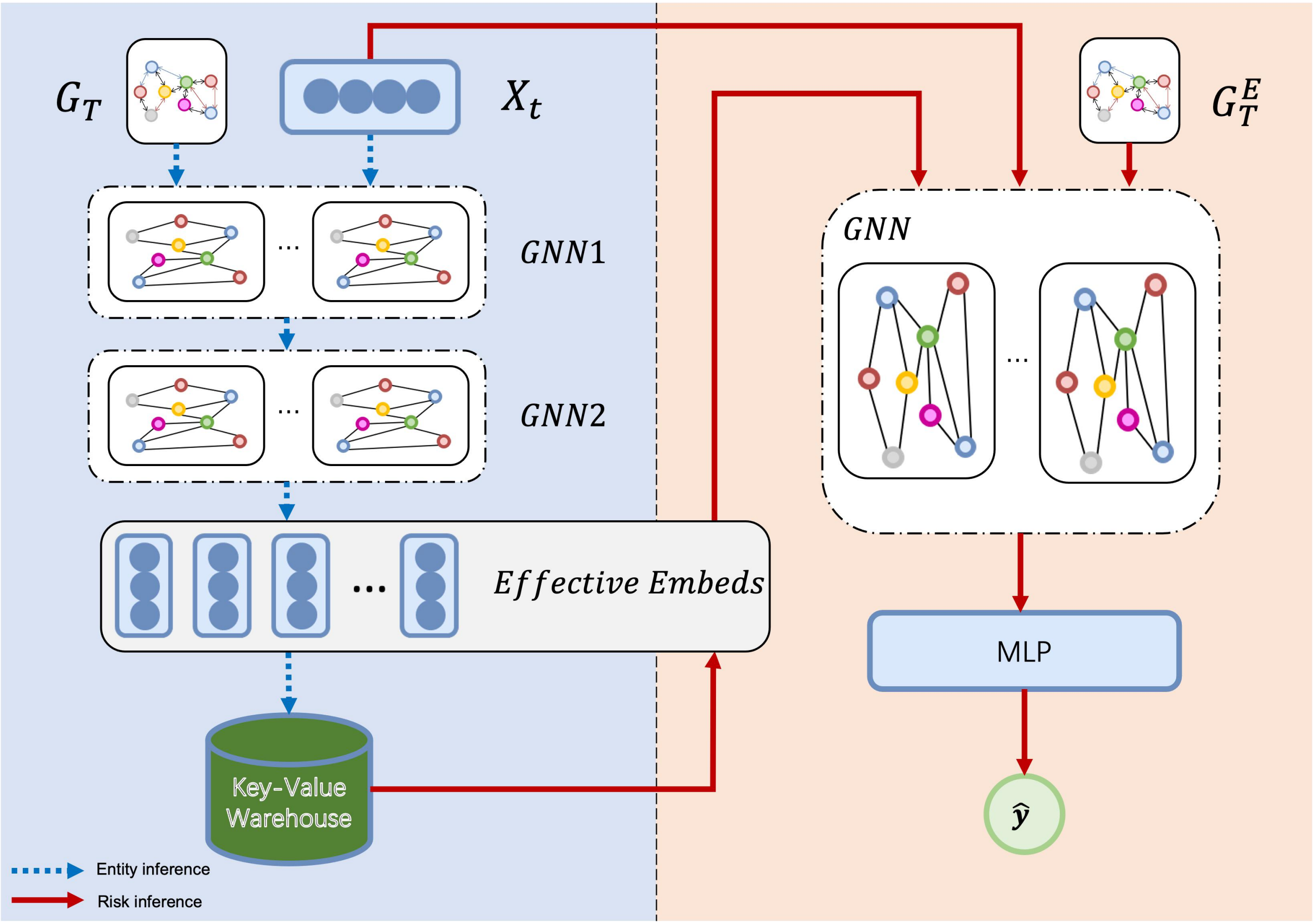
Source: Practical Lessons from Predicting Clicks on Ads at Facebook

离线学习

分区构图 | 编码学习

在线推断

实体表述 | 订单评估



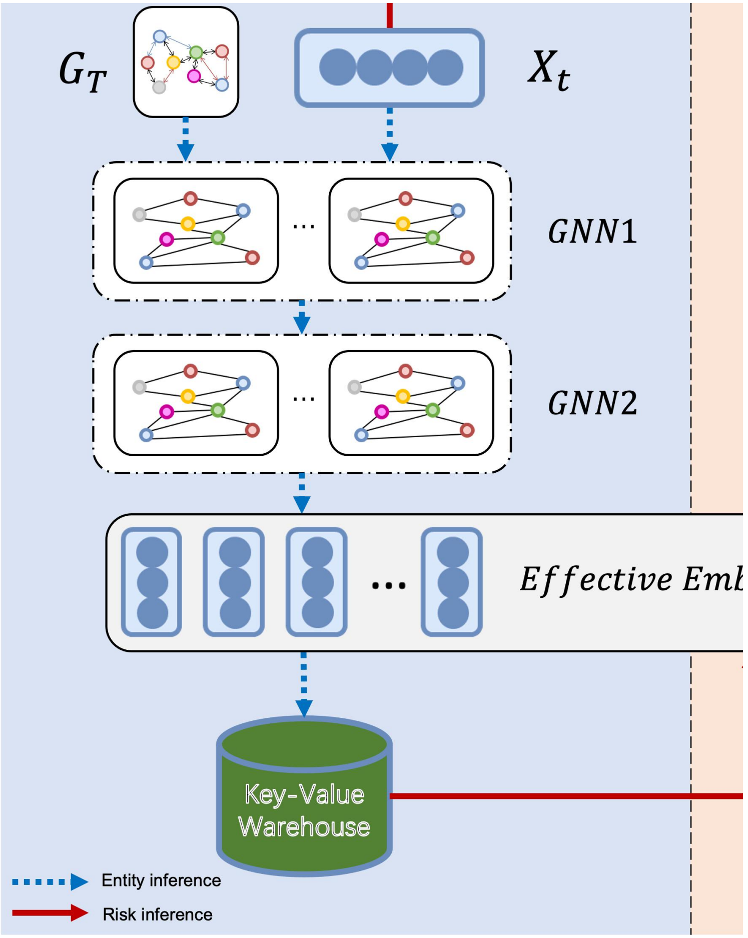
Source: Graph Neural Networks in Real-Time Fraud Detection with Lambda Architecture. *arXiv*

离线学习

分区构图 | 编码学习

在线推断

实体表述 | 订单评估



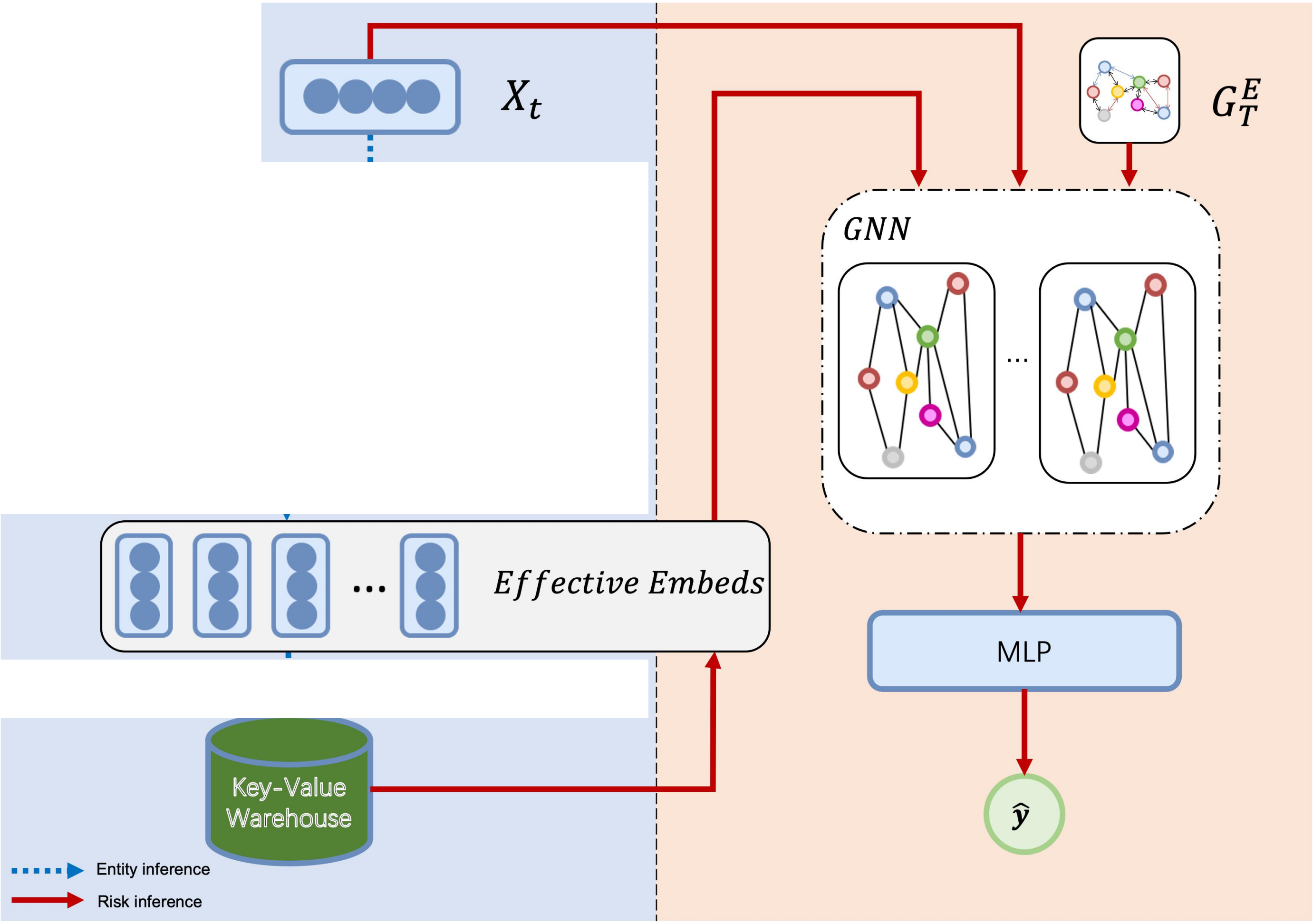
Source: Graph Neural Networks in Real-Time Fraud Detection with Lambda Architecture. *arXiv*

离线学习

分区构图 | 编码学习

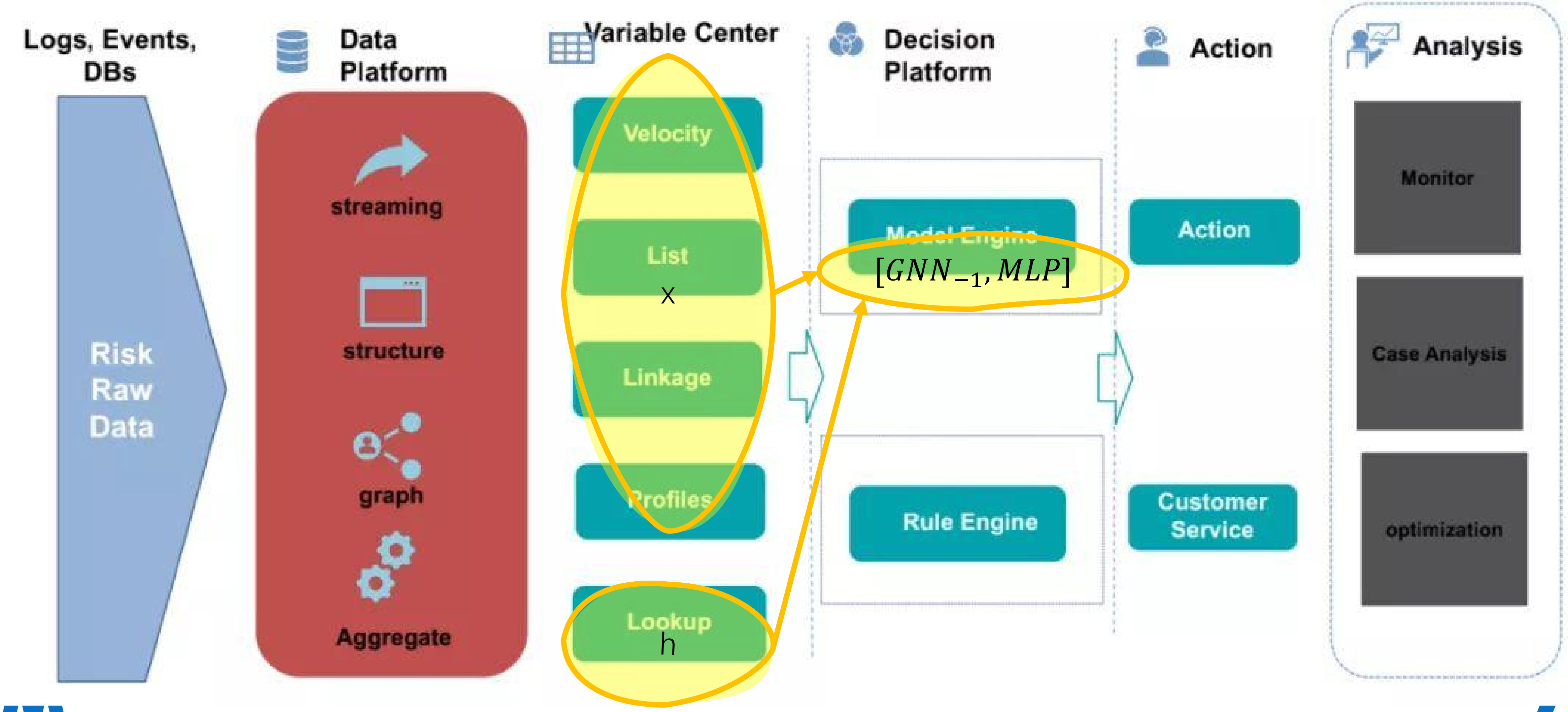
在线推断

实体表述 | 订单评估



Source: Graph Neural Networks in Real-Time Fraud Detection with Lambda Architecture. *arXiv*

Transaction Risk Platforms



端到端方案 – 实验对比

交易图

- 数月交易数据构建
- 90天实体关联窗口

数据

- 训练 时间切片前80%
- 验证 时间切片中10%
- 测试 时间切片后10%

关联实体

- 账户, 电子邮箱, 设备, 收货地址, IP

交易订单节点标签分布

- 正常:异常 - 20:1

端到端方案 – 实验对比

Model	Average Precision	ROC AUC
MLP	0.3917±0.0030	0.9209±0.0016
LGB	0.4081±0.0096	0.9317±0.0005
DDSL (GCN)	**0.4928±0.0021**	**0.9428±0.0006**
DDSL (GAT)	0.4796±0.0144	0.9387±0.0004

2020-11 Guest

Model	$ Fraud Precision	$ Fraud Recall
mlp	16.67%	12.50%
bst	16.67%	10.80%
gcn	16.67%	40.20%

2020-11 New (first payment dof < 90)

Model	$ Fraud Precision	$ Fraud Recall
mlp	16.67%	38.50%
bst	16.67%	51.20%
gcn	16.67%	68.99%

小结

- 端到端的图神经网络实时反欺诈检测的方案
  - 图切分使得分布式训练和推断可行
  - 有效的实验切片构图规避未来信息
  - 高效的在线推断 P99<80ms
  - 相比GBDT有显著提升 AP +20%

展望

- 对比引入RNN机制的TGN
- 社区分区大小对模型的影响
- 深度模型替换GBDT编码
- 图神经网络的预训练

参考资料

1. Lu, M., Han, Z., Zhang, Z., Zhao, Y. Shan, Y. (2021). Graph Neural Networks in Real-Time Fraud Detection with Lambda Architecture. *arXiv preprint arXiv:2110.04559*.
2. Chiang, W. L., Liu, X., Si, S., Li, Y., Bengio, S., & Hsieh, C. J. (2019, July). Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 257-266).
3. Li, G., Muller, M., Thabet, A., & Ghanem, B. (2019). Deepgcns: Can gcns go as deep as cnns?. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9267-9276).
4. He, X., Pan, J., Jin, O., Xu, T., Liu, B., Xu, T., ... & Candela, J. Q. (2014, August). Practical lessons from predicting clicks on ads at facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising* (pp. 1-9).
5. Deep Learning vs GBDT model on tabular data, https://www.kaggle.com/kyosukemorita/deep-learning-vs-gbdt-model-on-tabular-data/notebook

关注msup公众号

获取更多AI落地实践

麦思博(msup)有限公司是一家面向技术型企业的培训咨询机构，携手2000余位中外客座导师，服务于技术团队的能力提升、软件工程效能和产品创新迭代，超过3000余家企业续约学习，是科技领域占有率第1的客座导师品牌，msup以整合全球领先经验实践为己任，为中国产业快速发展提供智库。