

基于因果推断的商家经营 智能诊断实践

阿里巴巴集团 刘春辰
2022.11

目 录

- A 因果推断简述
- B 生意参谋之商家经营智能诊断
- C 关键技术节选
- D 效果展示

01

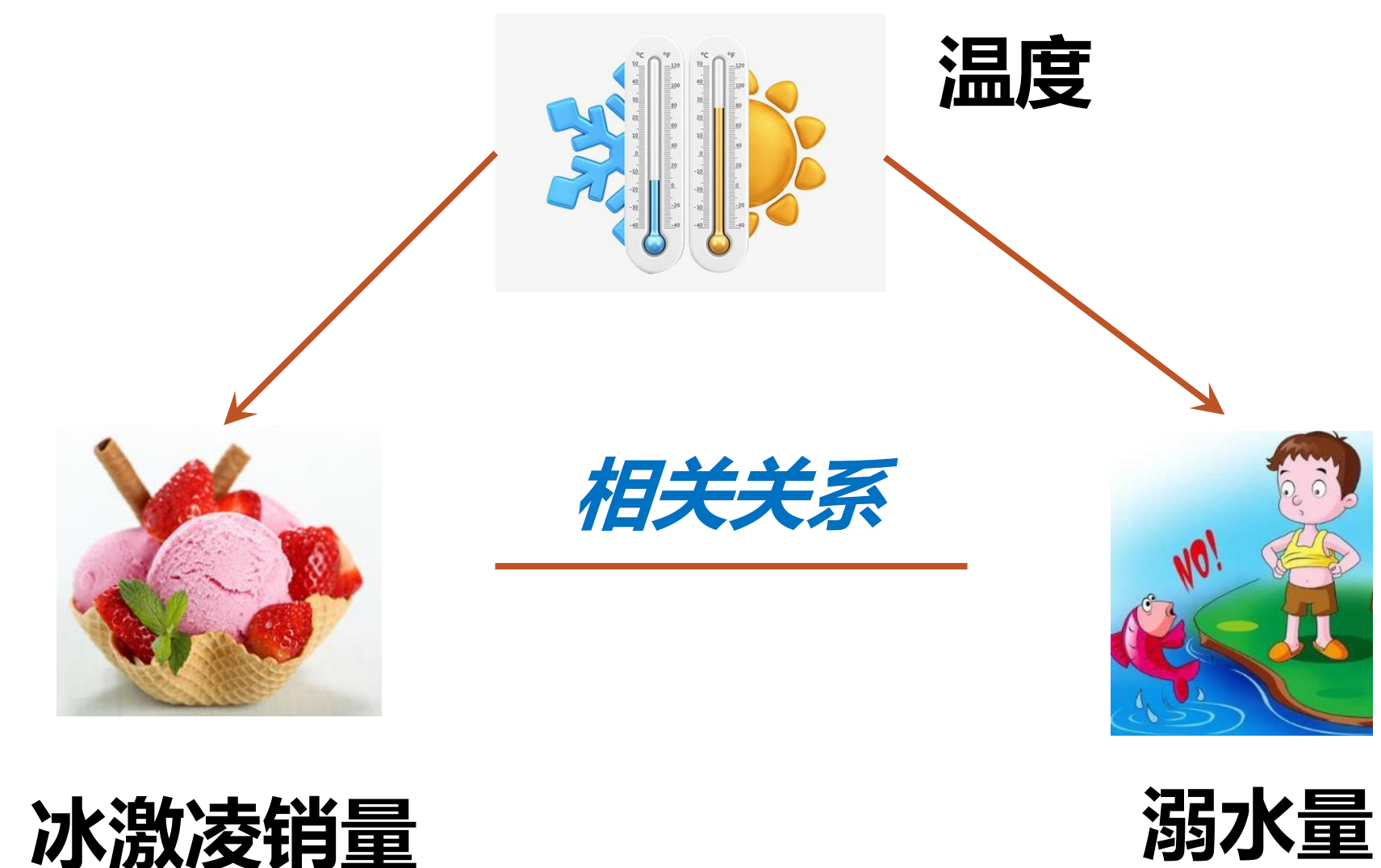
因果推断简介

什么是因果关系?



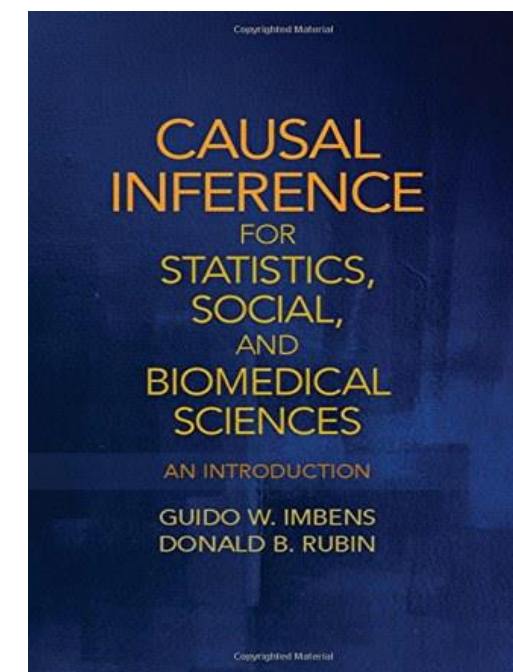
干预控制变量，目标变量的取值或者分布发生变化

Golden Metric: Randomized Control
Practical Way: Observational Studies



因果推断最主要的两大基础研究：潜在结果模型和因果网络模型。

- **潜在结果模型**：该类模型给出了因果作用的形式化数学公式，*主要用在原因和结果变量已知的前提下，定量评价原因变量对结果变量的因果作用。*

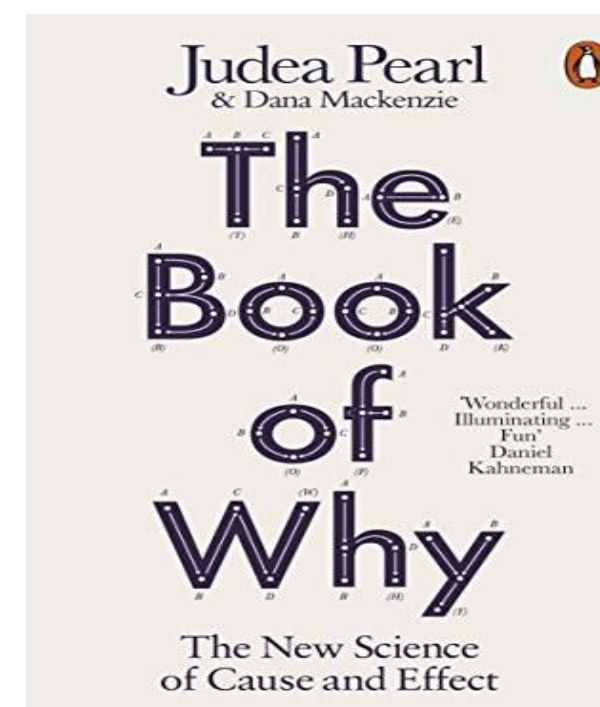


《causal inference for statistics, social, and biomedical science》，
2015



Donald B. Rubin
Department of statistics,
Harvard university, 清华

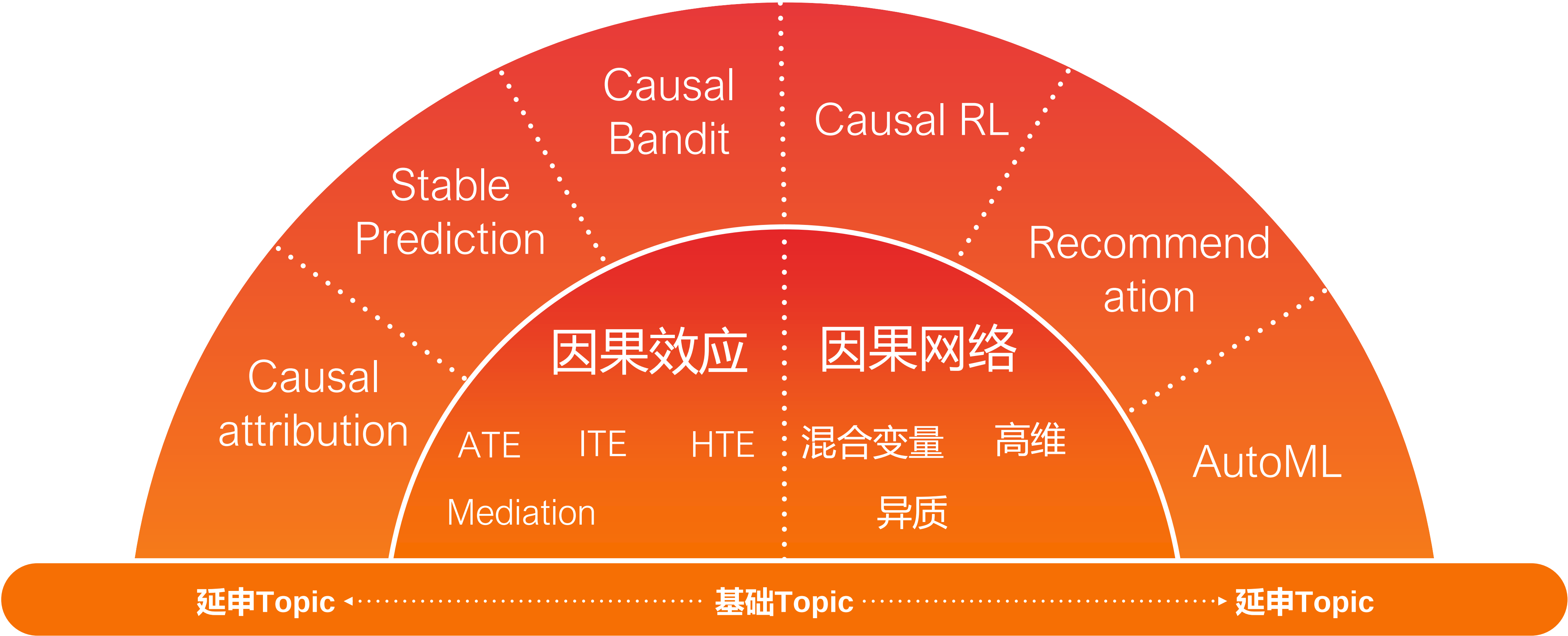
- **因果网络模型**：该类模型是描述数据产生机制和外部干预的形式化语言，*不仅能用于识别因果关系是否存在以及因果方向，也能定量评价因果作用。*



《The Book of Why》，
2018



Judea Pearl
Computer science,
UCLA



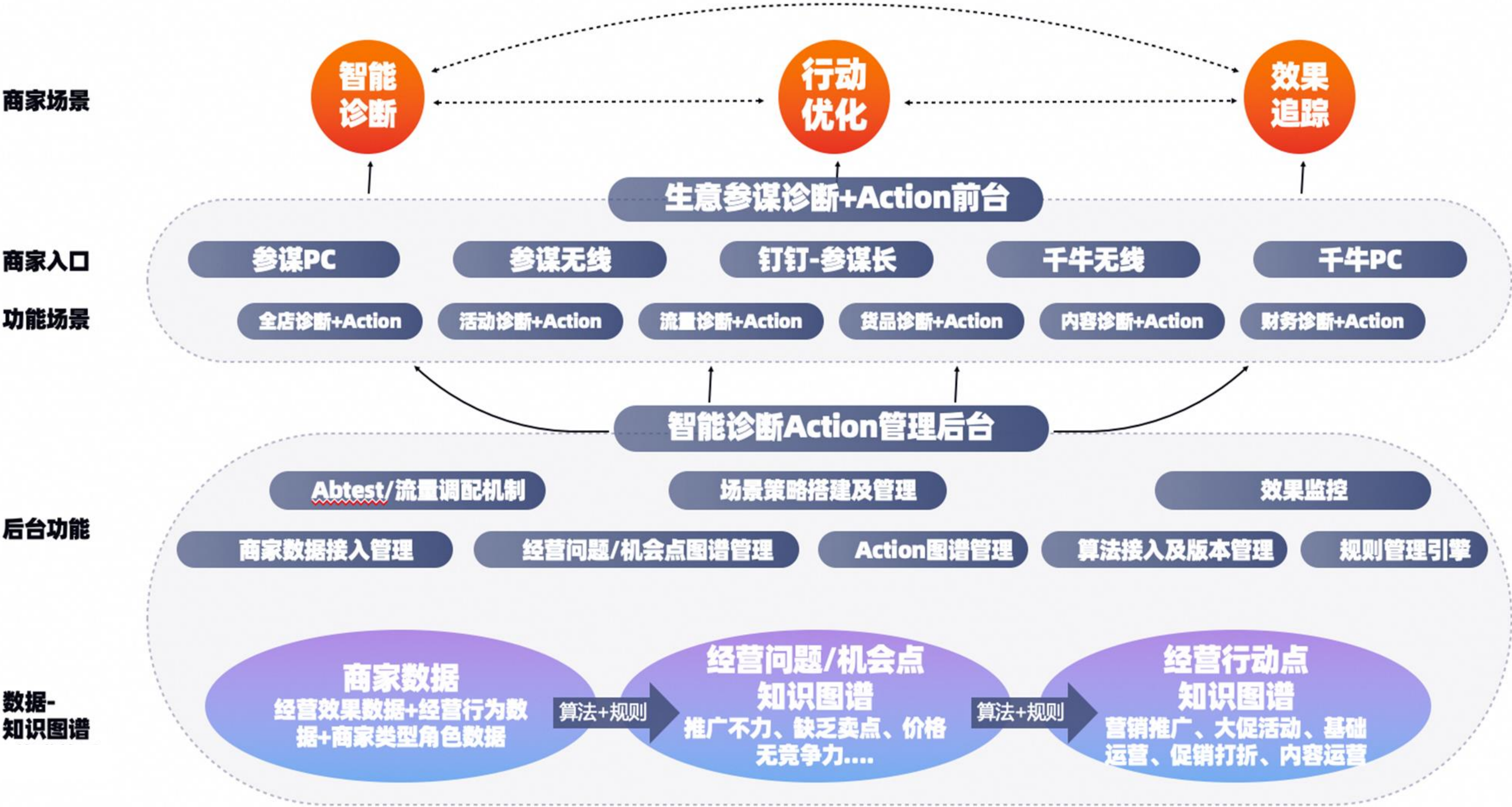
02

生意参谋
商家经营智能诊断

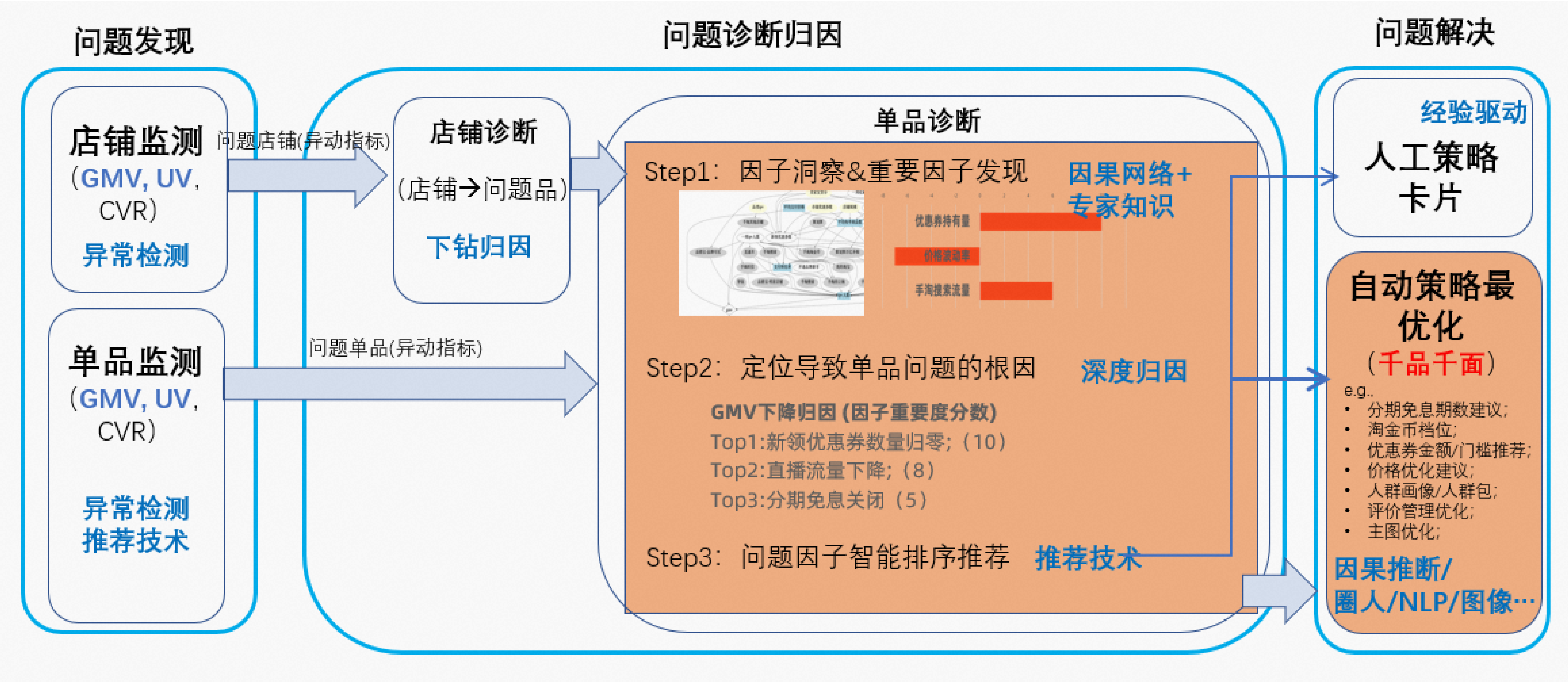
生意参谋智能诊断项目（1/3）

- 围绕**电商经营全链路**，帮助千万淘系商家完成**一站式分析、决策和行动优化**





支持商家经营状态监测（问题发现）、问题诊断归因、问题解决建议的全链路全自动智能服务。



03

智能诊断关键技术节选

● 主要技术挑战

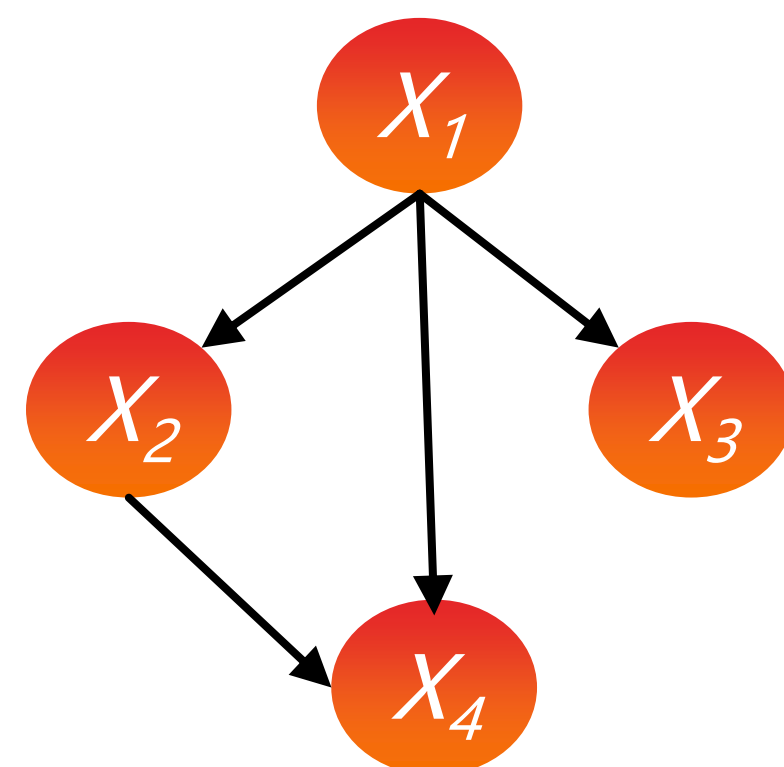
- **混合数据类型**：现实场景数据多为混合类型数据；既有方法一般假设数据类型单一，少数方法考虑混合变量上因果关系发现，但做线性假设or规模化受限；
- **大规模CD**：变量维度高，样本规模大；



● 提出混合因果网络新模型

Assumption: Causal Markov

Identifiability Proved



DAG结构

Node: 每个 X_i 代表一个观察变量，可以是categorical类型 $X_i \in \{c_1, \dots, c_i\}$ ，也可以是连续类型 $X_i \in R$ ；

Edge: $X_i \rightarrow X_j$ ，表示 X_i 是 X_j 的直接原因；

当 X_j 为连续变量时，
$$X_j = f_j(\dot{X}_{\mathcal{P}_{(j)}^g}) + E_j \quad (1)$$

$$E_j \perp\!\!\!\perp X_{\mathcal{P}_{(j)}^g}$$

当 X_j 为离散变量时，

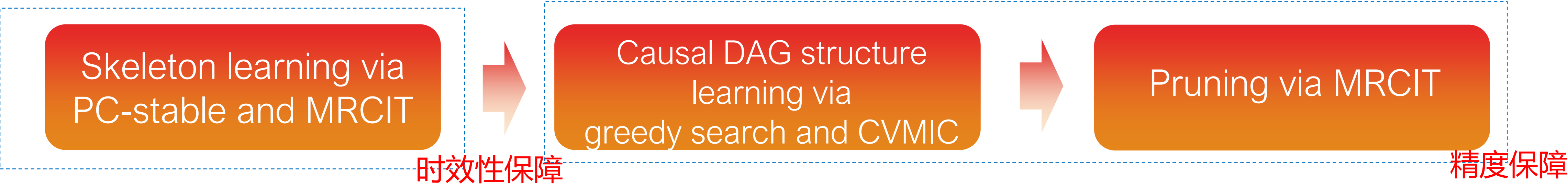
$$X_j = \arg \max_{k \in \{1, \dots, c_j\}} f_{j,k} \left(X_{\mathcal{P}_{(j)}^g} \right) + E_{j,k}, \quad (3)$$

Gumbel distribution

$$\begin{aligned} \Pr(x_{i,j} | X_{\mathcal{P}_{(j)}^g} = \mathbf{x}_{i,\mathcal{P}_{(j)}^g}, f_j) &= \Pr(E_j = x_{i,j} - f_j(\mathbf{x}_{i,\mathcal{P}_{(j)}^g}) | X_{\mathcal{P}_{(j)}^g}) \\ &= \Pr(E_j = x_{i,j} - f_j(\mathbf{x}_{i,\mathcal{P}_{(j)}^g})), \quad (2) \end{aligned}$$

$$\Pr(X_j = l | X_{\mathcal{P}_{(j)}^g} = \mathbf{x}_{i,\mathcal{P}_{(j)}^g}, f_j) = \frac{\exp(f_{j,l}(\mathbf{x}_{i,\mathcal{P}_{(j)}^g}))}{\sum_{c=1}^{c_j} \exp(f_{j,c}(\mathbf{x}_{i,\mathcal{P}_{(j)}^g}))}. \quad (4)$$

● 混合推理算法HCM



提出（条件）独立性检验新方法MRCIT：
支持混合变量、非线性

提出混合信息范式CVMIC，指导
DAG发现

Characterization of Conditional
cross-covariance operator

KCIT: $\Sigma_{\ddot{X}Y|Z} = 0 \iff X \perp\!\!\!\perp Y|Z.$

$$\Sigma_{YX|Z} = \Sigma_{YX} - \Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}.$$

$$S(\mathcal{G}; \mathbf{X}) = \sum_{j=1}^n S_j(\mathcal{P}_{(j)}^{\mathcal{G}}; X_j)$$

$$S_j(\mathcal{P}_{(j)}^{\mathcal{G}}; X_j) = \sum_{q=1}^k L(X_j = \mathbf{x}_{\mathcal{D}_q, j} | \mathcal{P}_{(j)}^{\mathcal{G}}; \hat{f}_j^{(q)}) - \frac{\log(m)}{2} |\mathcal{P}_{(j)}^{\mathcal{G}}|,$$

MRCIT:

基于随机傅里叶
特征优化KCIT，
提升时效性

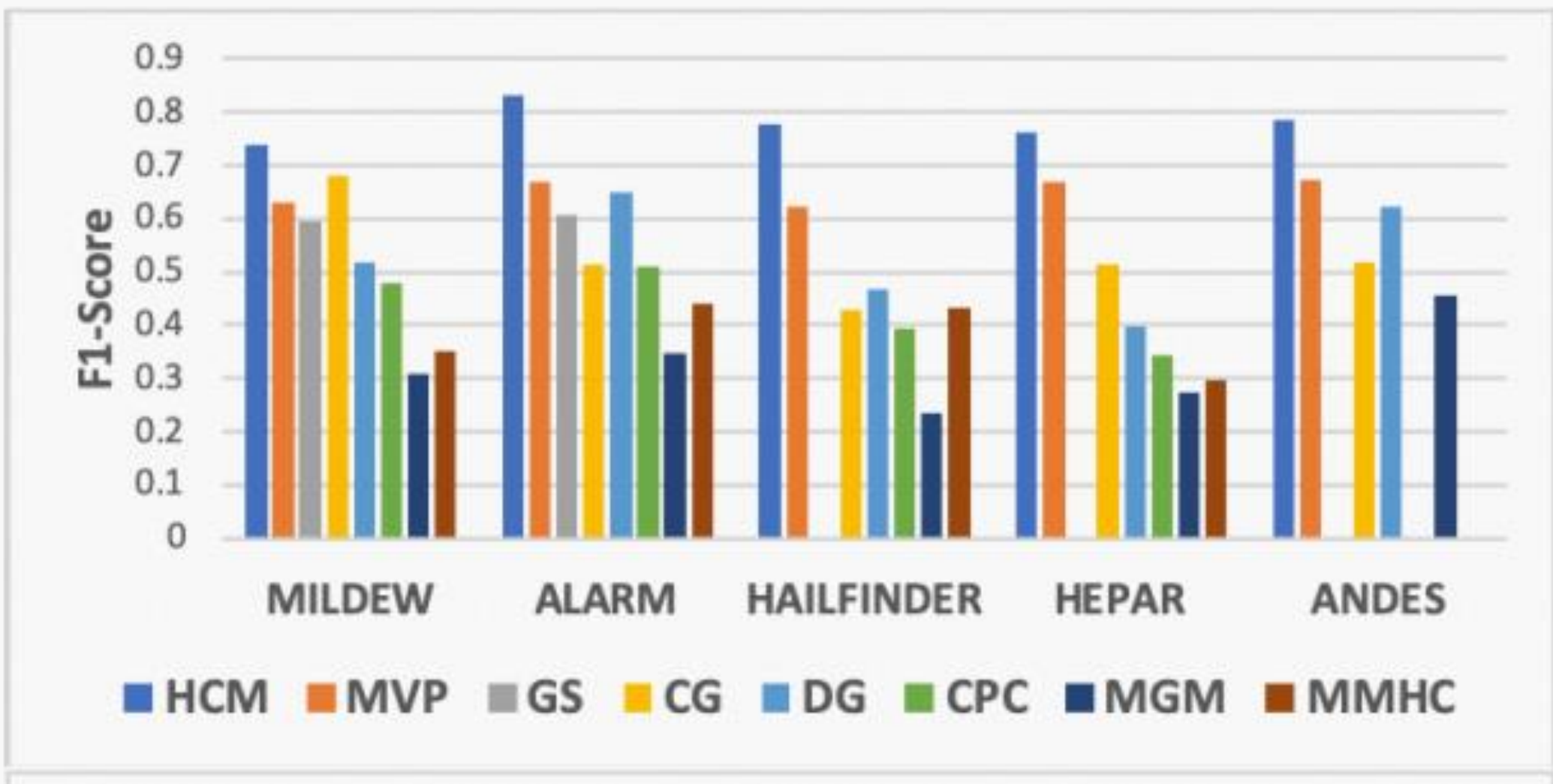
$$\begin{aligned} k(\mathbf{x}, \mathbf{y}) &= \int e^{i\omega_{ct}^T \Delta_{ct}} dF_{\omega_{ct}} + \int e^{i\omega_{ca}^T \Delta_{ca}} dF_{\omega_{ca}} \approx \frac{1}{S} \text{Tr}(\varphi(\mathbf{x})\varphi(\mathbf{y})^*) \\ &= \sum_{j=1}^n k(\mathbf{x}_j, \mathbf{x}) \alpha_j \\ &= \sum_{j=1}^n \frac{1}{S} \sum_s \left[\varphi_{ct}(\mathbf{x}_{jct})^* \varphi_{ct}(\mathbf{x}_{ct}) + \varphi_{ca}(\mathbf{x}_{jca})^* \varphi_{ca}(\mathbf{x}_{ca}) \right] \alpha_j \\ &= \varphi_{ct}(\mathbf{x}_{ct})^* \mathbf{w}_{ct} + \varphi_{ca}(\mathbf{x}_{ca})^* \mathbf{w}_{ca}, \end{aligned} \tag{8}$$

K-fold 交叉loss Penalty

● 实验结果

仿真数据

name	MILDEW	ALARM	HAILFINDER	HEPAR	ANDES
# of nodes	35	37	56	70	223
# of edges	46	46	66	123	338



真实数据

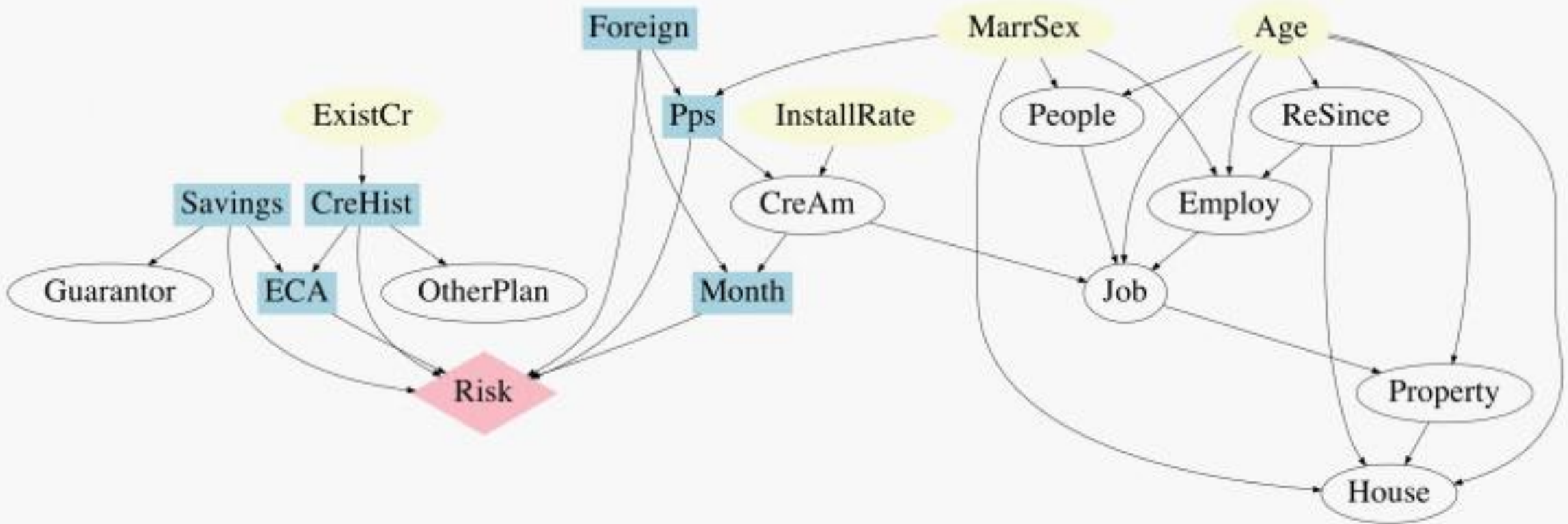
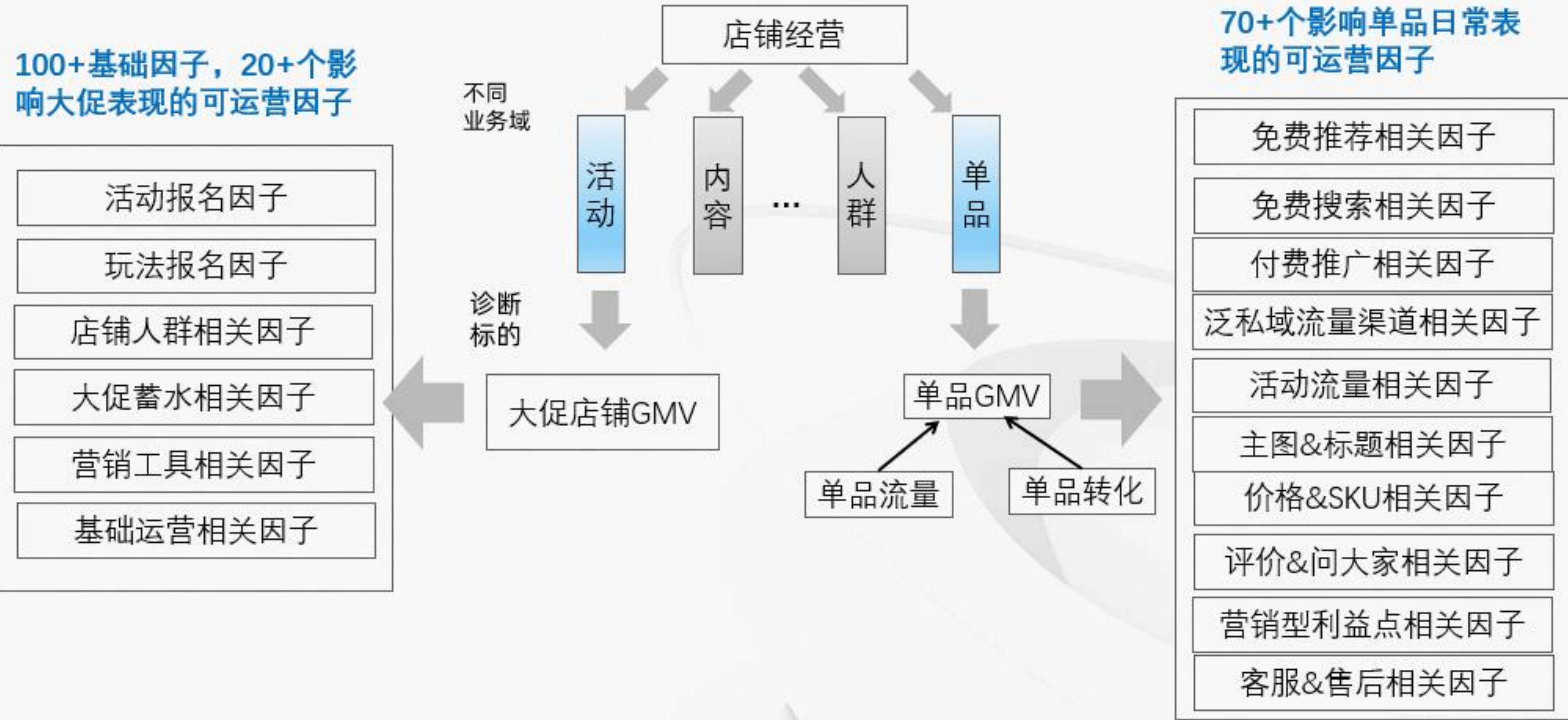


Figure 3: Learned causal graph from credit data set. The red diamond denotes the target, i.e., credit risks. Blue rectangles are direct causes of credit risks.

基于HCM的商家经营知识图谱

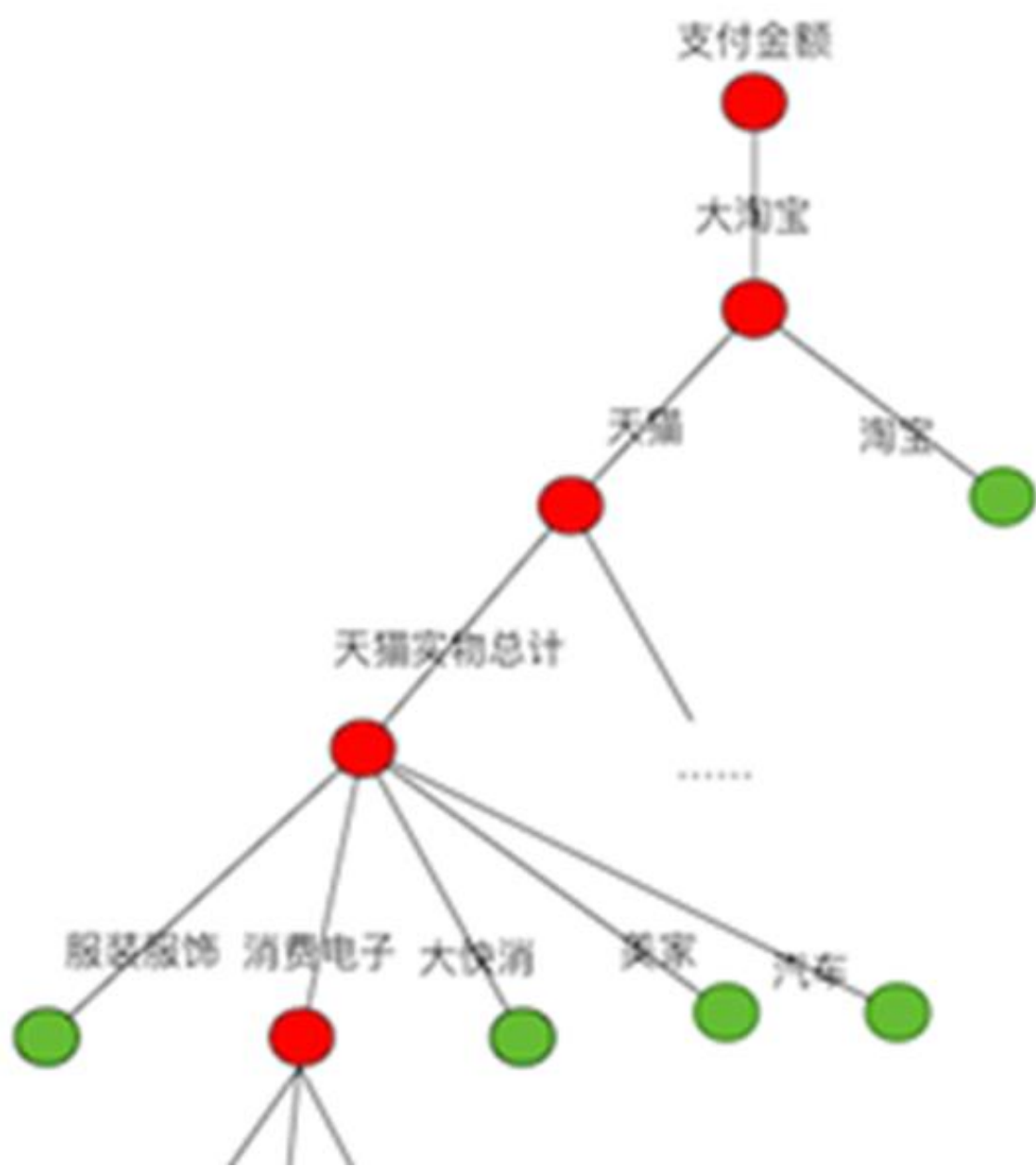


基于因果的深度归因技术

● 常见归因场景

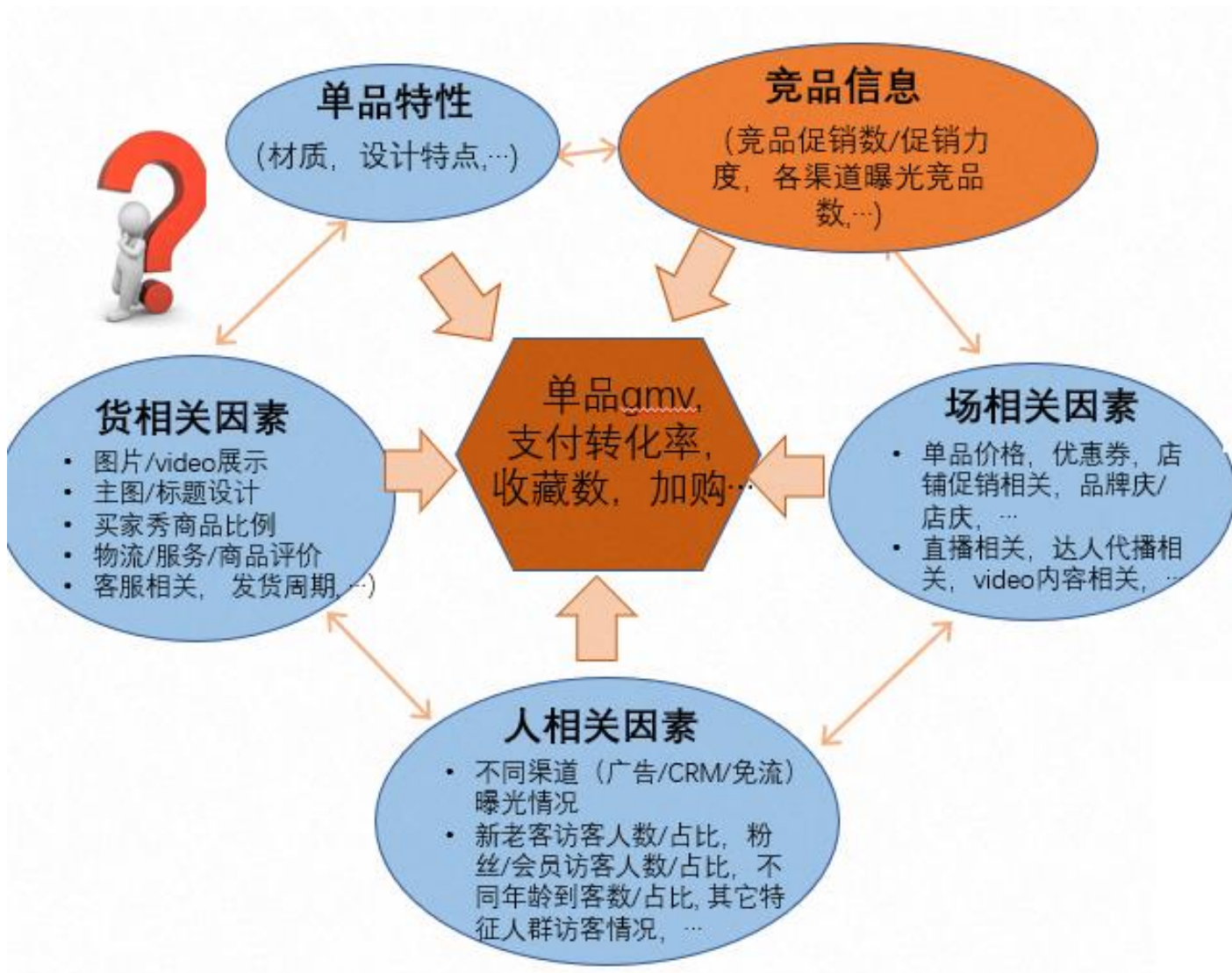
波动归因

拆解归因



每个节点对应的指标都是由样本个体汇总计算而成；整体波动需要定位哪些个体波动造成

深度归因



每个指标都是样本的自身特性，分析特征间的控制特性，比如价格上涨如果影响流量、转化率、...

异常点归因

典型场景：故障诊断/运维，部分机器出现故障，诊断机器出现故障的原因；

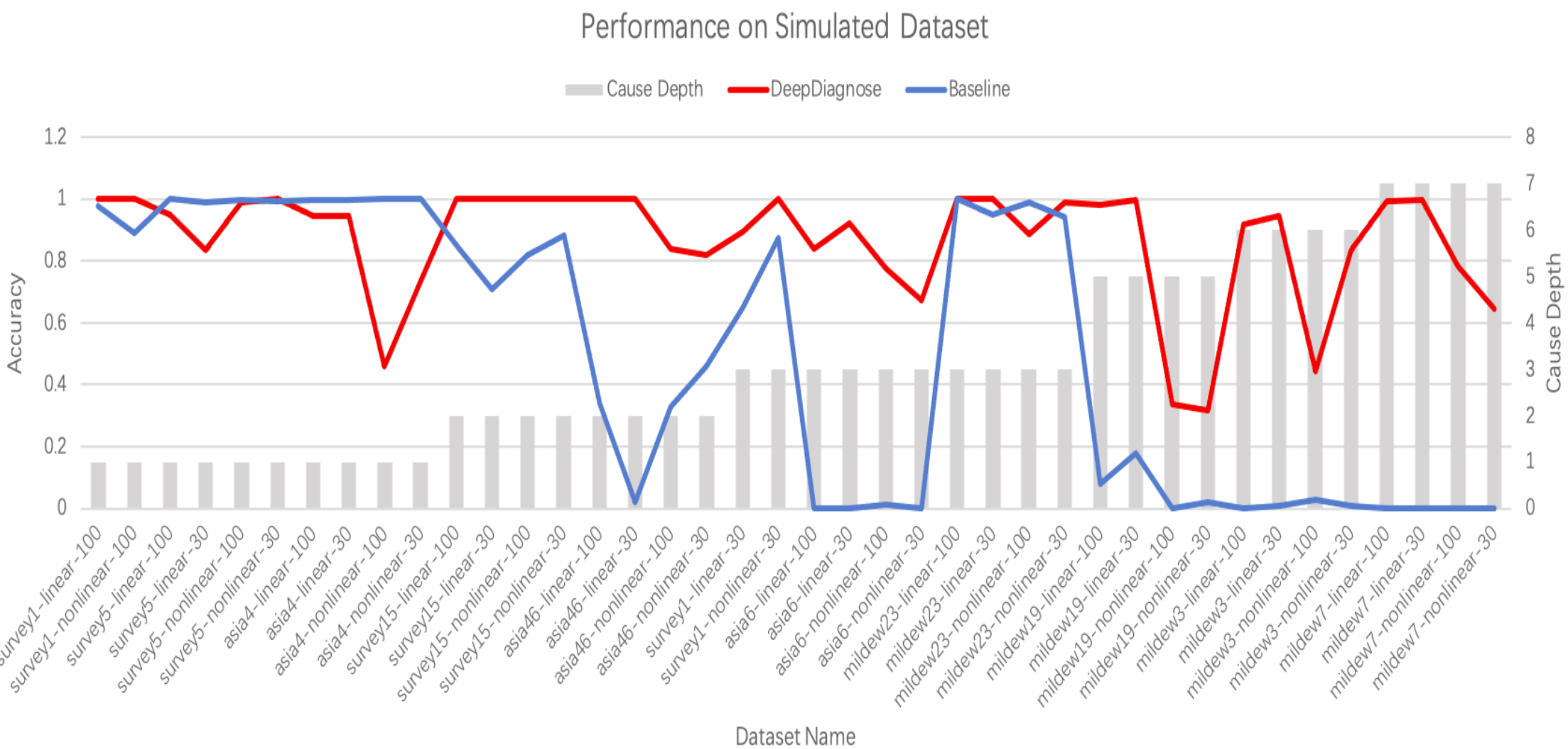
多触点归因

- ✓ 渠道归因（淘宝用户在平台触达多个付费、免费渠道，评价各个渠道对客户转化的贡献；）
- ✓ 广告归因：用户在站内外（线下广告、各平台广告）有多个触点被产品广告曝光，如果评估各触点告对客户转化的贡献？

基于因果的深度归因技术

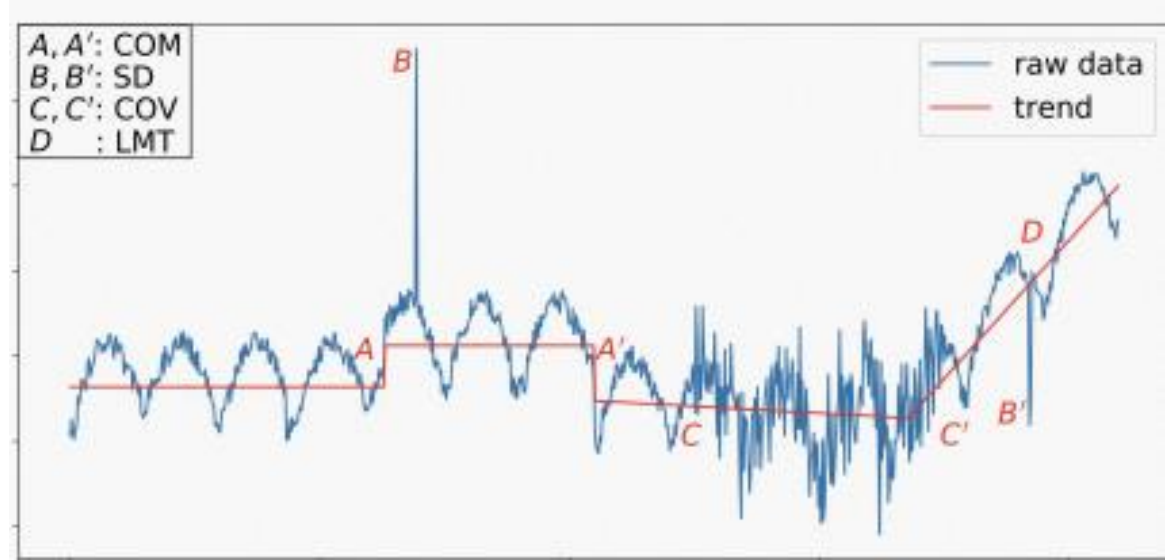
● 现有技术问题

- 基于相关性的归因技术存在把果当因、定位伪因、效果估计偏差导致的只能定位表层原因；
- 近年提出融合causality和shapley的归因技术，能定位到深层次的原因；



● 深度归因新模型

基于RobustSTL等系列异常算法识别目标、因子显著波动



利用因果网络技术识别正确的因果Order

因果order: 变量序列, 排在后面的变量不会是排在前面变量的原因;
(e.g., <到货周期、差评数、老客CVR、新客CVR、CVR、UV、客单、GMV>)

提出新的Multi-ATE估计技术, 并基于因果order, 估计干预变量的波动对目标的影响

$$v(X_i, S_i) = E(Y|do(X_i = x_{curr}, S_i = s_{curr})) - E(Y|do(X_i = x_{ref}, S_i = s_{curr}))$$

基于各因子的Multi-ATE效果估计, 计算其对目标波动的贡献度

03

产品效果

生意参谋智能诊断—产品效果



周度100W+商家活跃，日度10~20W策略采纳



【图中均为测试数据】

生意参谋智能诊断—产品效果

异常预警

连带分析

新品追踪

商品诊断

销量预测

统计时间 2022-11-24 15:12:40

实时

7天

30天

日

周

月

自定义

商品

支付金额

支付件数

支付转化率

商品加购件数

商品访客数

操作

智能诊断机会

自动

8P

机会

您有新品掉出新品礼金孵化池，无法继续获得小黑盒超两千万流量曝光，十亿新品补贴加码等海量资源！

影响值

数据

商品最近7天支付转化率为为1.16%，预估开通新品礼金后，商品支付转化率可以提升6.39%

建议

建议参考优质商品，报名新品礼金

报名新品礼金

有用

无用

智能诊断问题

问题

本商品流量方面近1天购物车访客数下降，导致商品支付金额表现变差

影响值

数据

支付金额对比7日均值下降31.50%；系统诊断购物车访客数对比7日均值从168变化到154，影响转化，间接对商品GMV下降有轻微影响

建议

建议您去电脑端客户运营平台，圈选加购未支付人群并进行短信触达，提升14购物车访客来带动商品支付金额的上升（课程：购物车营销促进客户成交）

前往推送消息

有用

无用

趋势

详情

关注

运营策略

智能诊断建议

调整搜索词

优化品仓部署

设置商品连带

趋势

详情

关注

运营策略

智能诊断建议

提升搜索

优化品仓部署

报名学生价

趋势

详情

关注

运营策略

智能诊断建议

报名新客折扣

补充属性

趋势

详情

关注

运营策略

智能诊断建议

优化品仓部署

设置商品连带

趋势

详情

关注

运营策略

智能诊断建议

设置商品连带

优化品仓部署

【图中均为测试数据】

THNAK YOU