



廈門大學
XIAMEN UNIVERSITY

自 / 强 / 不 / 息 止 / 于 / 至 / 善

01

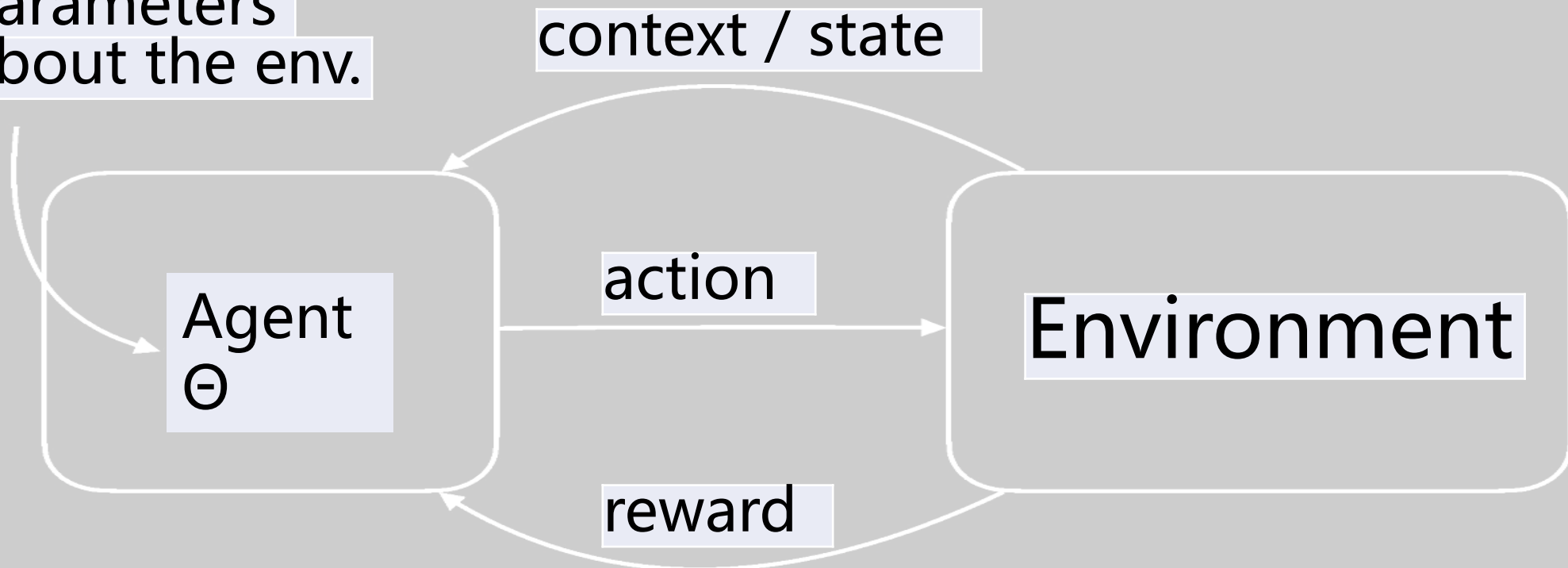
Causal Reinforcement Learning

Designed by LDJ

	两个例子了解全貌	特点
ALL	21点纸牌游戏的场景运用	1.由于整个环境描述较为简单， 我们可以凭借经验知识轻易获得因果关系
		2.目的是优化强化学习的速度和精度
	改进广告布置的加权	1.蕴含隐藏变量，这里抽离为未知的用户信息
		2.利用设置的因果结构进行一个重要性采样的加权改变，同时 匹配置信度问题

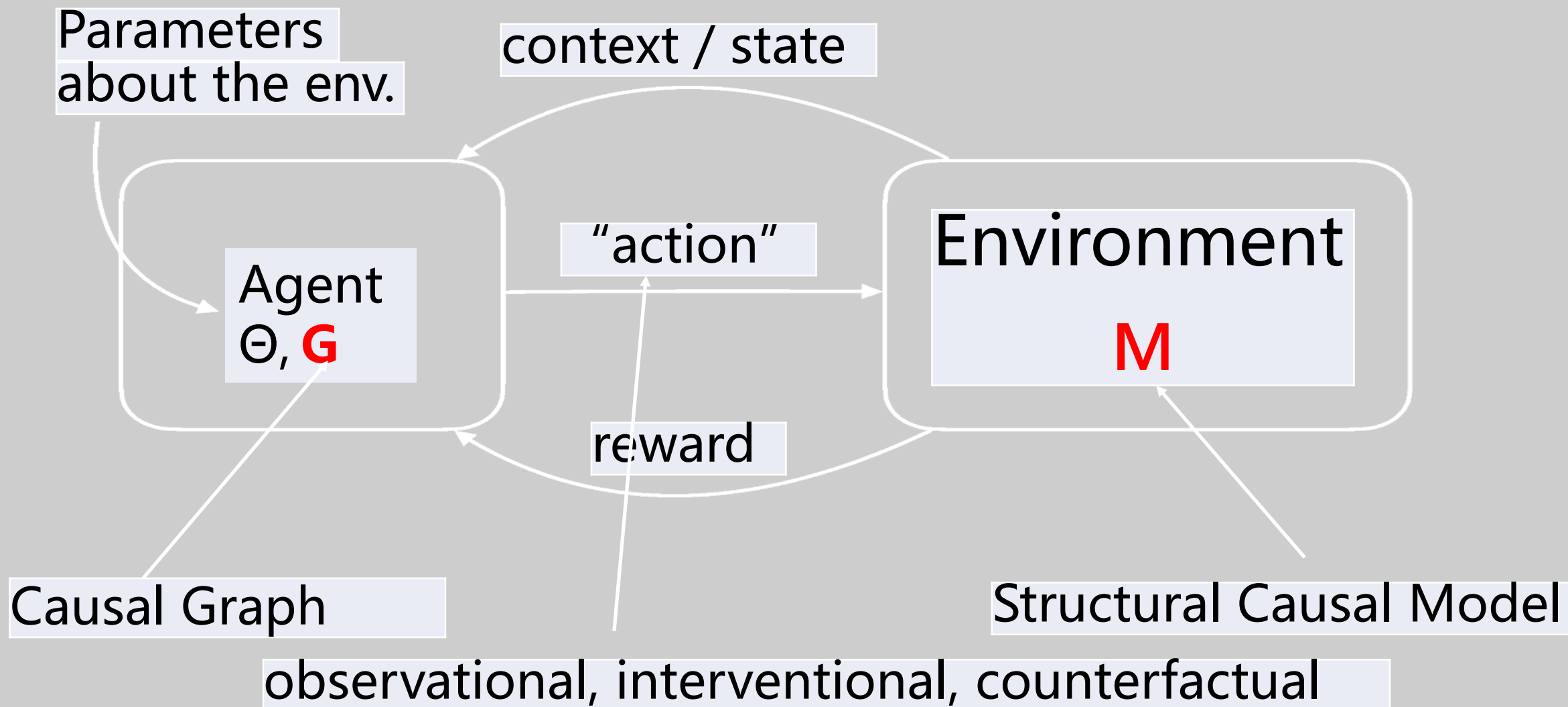
Causal RL = CI + RL

Parameters
about the env.

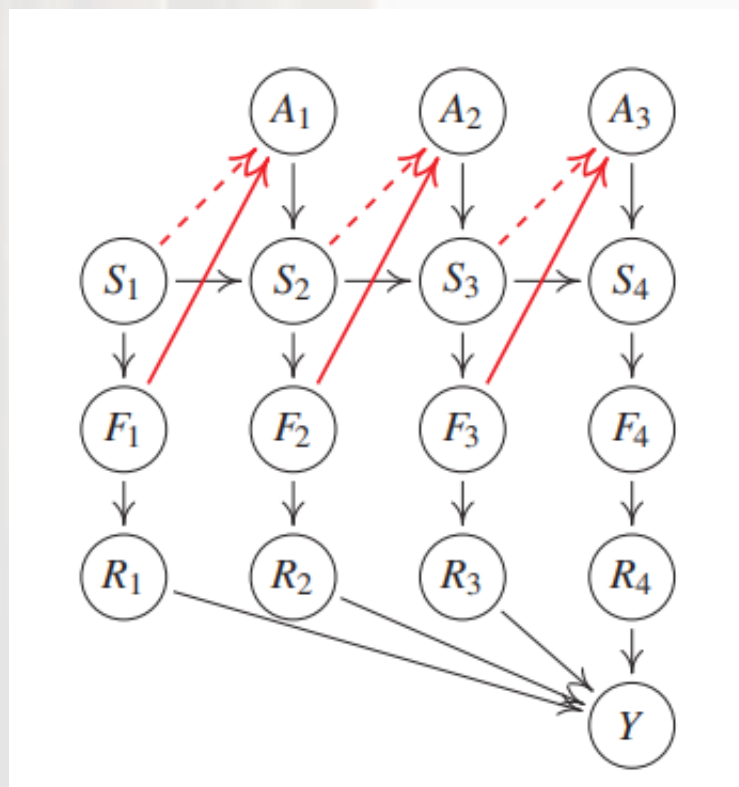


- Receive feedback in the form of rewards.
- Agent' s utility is defined by the reward function.
- Must (learn to) act so as to maximize expected rewards.

Causal RL = CI + RL

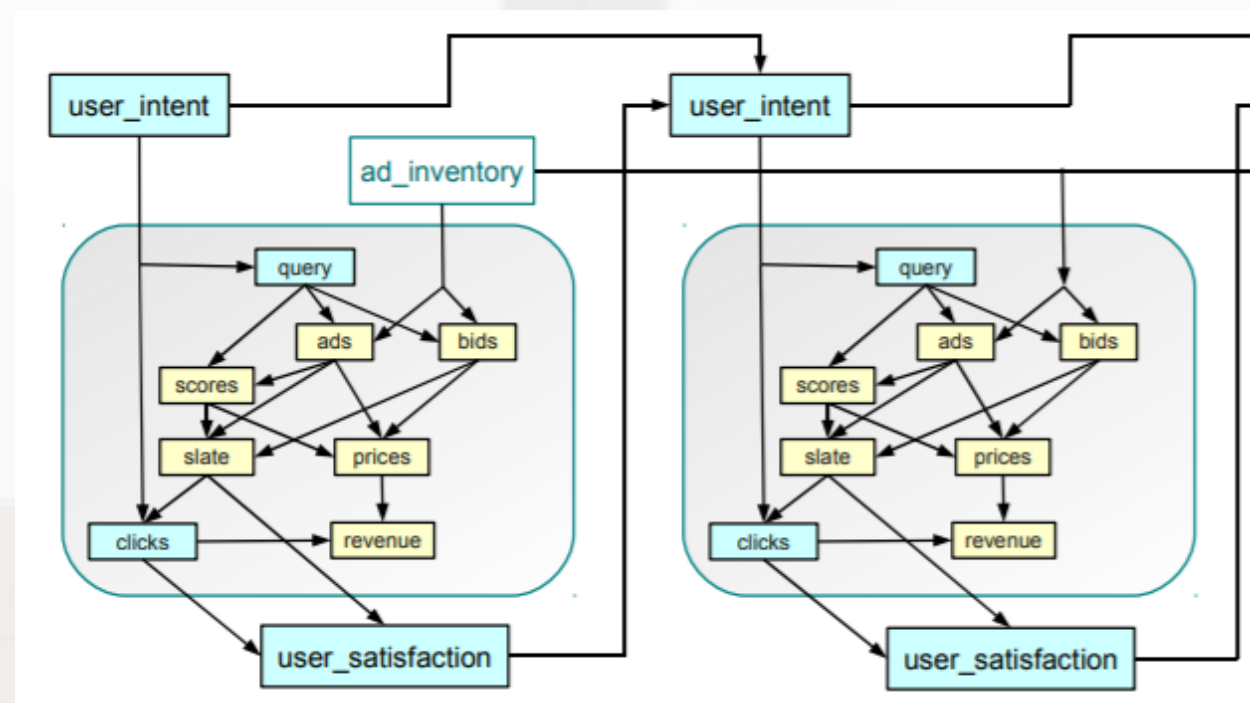


21点纸牌场景运用



Causal
Graph

改进广告布置的加权



SCM

Causal RL 基本可以分为三种任务：

- ✓ 我们能不能借助观测数据来加速强化学习的过程呢？
- ✓ 我们应该如何去干预？
- ✓ 通过反事实来进行决策

前两个例子可以说明

通过Causal Graph
找到最佳干预点

Counterfactual Policy Evaluation for Decision-Making in Autonomous Driving

CRL的四个主题：

- ✓ 去除强化学习算法里的混杂效应 (Regarding confounding)
- ✓ 在强化学习中应用反事实框架 (Regarding counterfactuals)
- ✓ 因果表示学习 (Regarding causal representation learning)
- ✓ 使用强化学习的方法进行因果发现 (Regarding causal structure learning)



厦门大学
XIAMEN UNIVERSITY

自 / 强 / 不 / 息 止 / 于 / 至 / 善

02

Causal Time Series

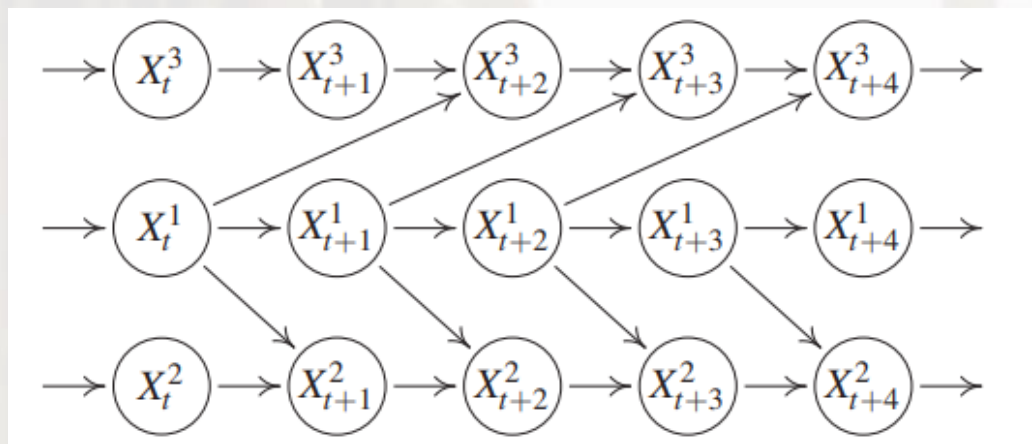
对不同时间变量之间的因果关系进行推理比没有时间结构的因果推理**更容易**。
因果结构必须与时间顺序**一致**。

知道节点的因果顺序，并假设没有隐藏变量，发现因果DAG不需要除马尔科夫条件和最小化外的其他假设(例如faithfulness、restricted function classes是不需要的)。

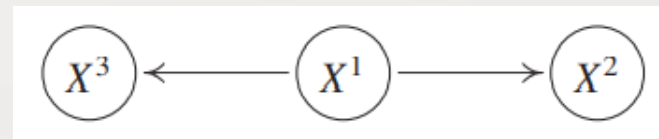
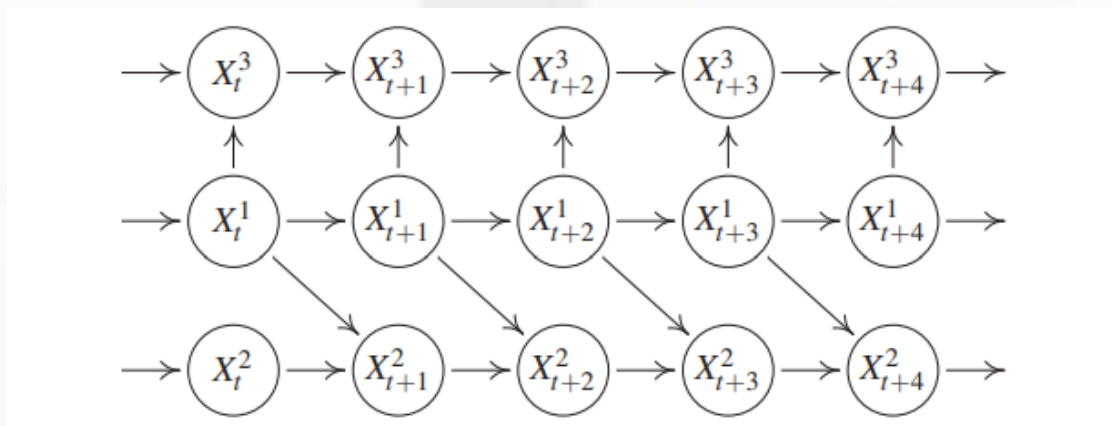
时间序列存在的三个问题：

- 所考虑的变量集合不再是因果充分的
- 存在相同时间的变量，它们不是一个因果有序的先验。
- 在实践中，只有一次的时间序列过程。

没有瞬时影响的时间序列



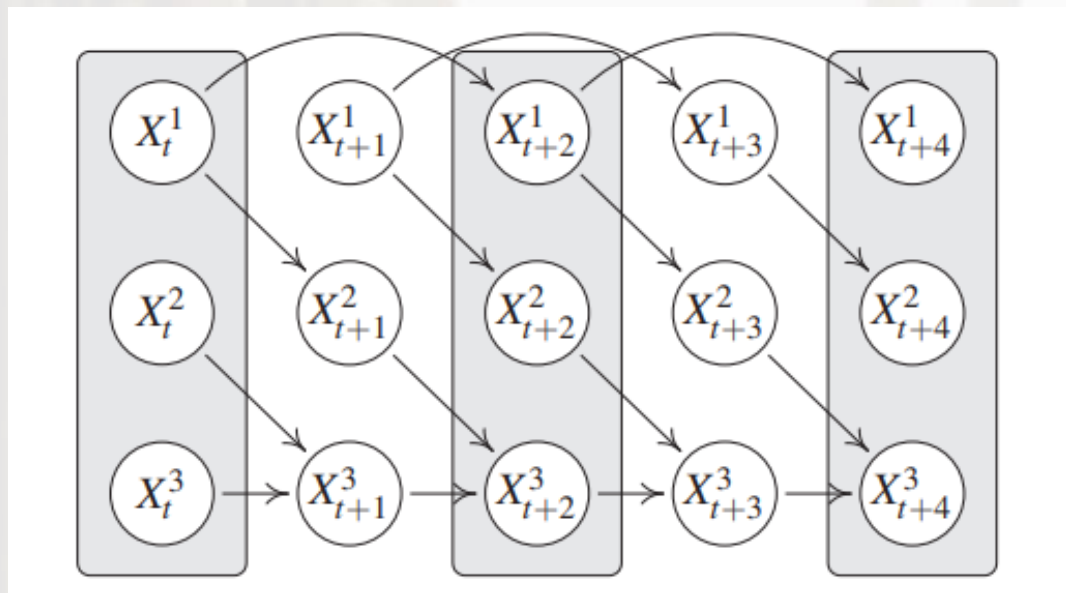
具有瞬时影响的时间序列



全时图 (full time graph)
摘要图 (summary graph)
jointly independent noise terms
干预 (Interventions)

SCM

下采样时间序列：仅有阴影区域中的变量被观测

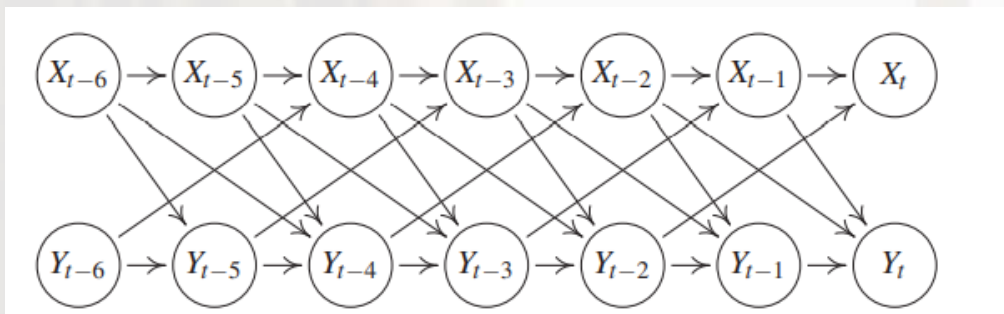


目的：下采样的手段是会存在问题的，并不能够捕获正确的信息

Granger Causality

$$X \text{ Granger-causes } Y \quad :\Longleftrightarrow \quad Y_t \not\perp\!\!\!\perp X_{\text{past}(t)} \mid Y_{\text{past}(t)}.$$

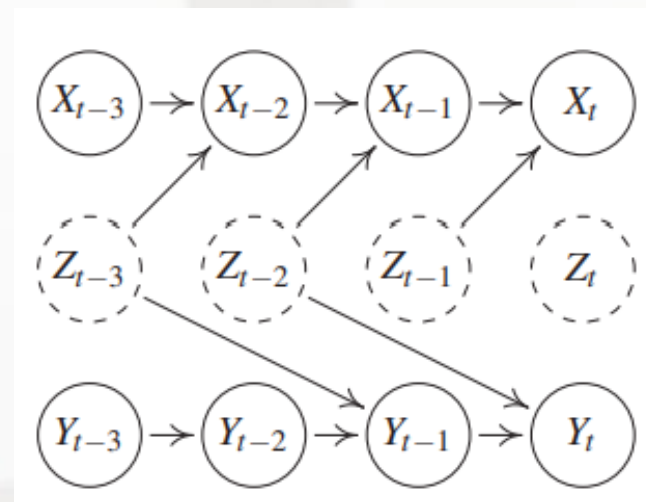
上述表达的意思也就是说：摘要图包含从X到Y的箭头。



Granger因果关系成立的典型，
给定 Y_t 的过去值， Y_t 确实依赖于 X 的过去值。

Limitations

违反因果充分性在因果时间序列中是一个严重问题



如果 X_t 和 Y_t 都受到隐藏时间序列 Z 的影响，如果能观测到 Z ，那么：

$$Y_t \not\perp\!\!\!\perp X_{\text{past}(t)} \mid Y_{\text{past}(t)}$$

但事实上：

$$X_t \perp\!\!\!\perp Y_{\text{past}(t)} \mid X_{\text{past}(t)}.$$



厦门大学
XIAMEN UNIVERSITY

自 / 强 / 不 / 息 止 / 于 / 至 / 善

03

Papers

算法和鲁棒性

现状：已有模型均假设训练数据和测试数据独立同分布 (IID)

事实：训练数据和算法部署场景的数据之间存在一定因果关联

方法

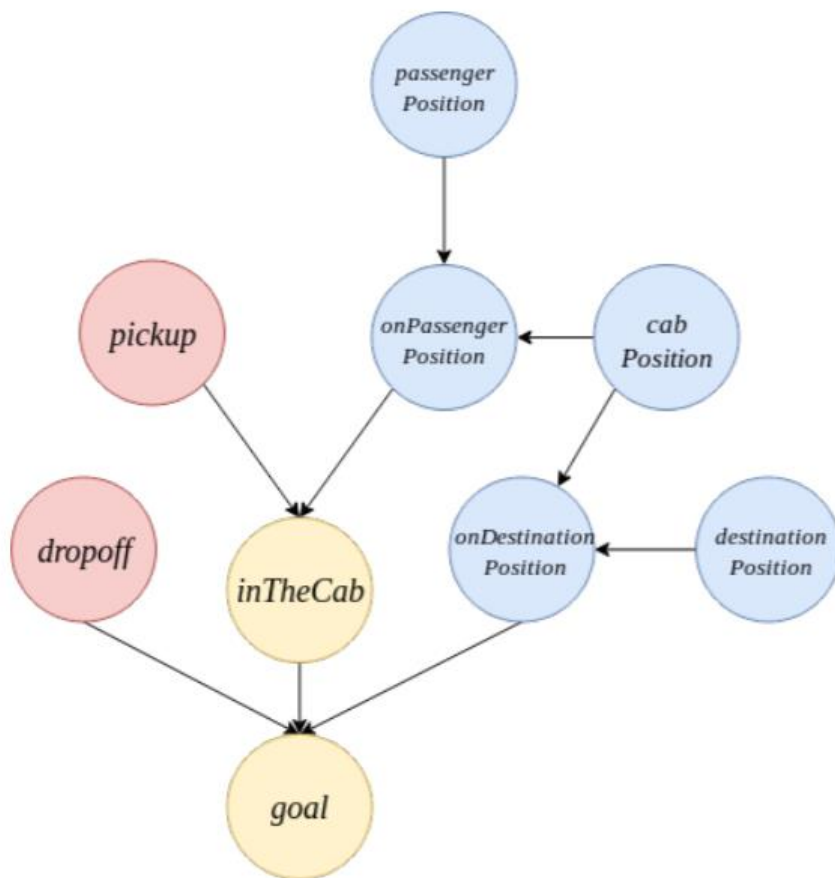
利用因果图信息

利用表示学习, stable learning是这两种思路的融合

因果模型已知	Domain adaptation with conditional transferable components
	Few-shot Domain Adaptation by Causal Mechanism Transfer
因果模型未知	Domain adaptation by using causal inference to predict invariant conditional distributions
	Domain adaptation as a problem of inference on graphical models.

Causal Based on Q-learning

目的：使用因果模型来帮助强化学习



Causal Graph

特点：

他们的假设是因果推断能帮助强化学习更高效地学习价值函数或者策略，也就是通过缩小状态空间或者动作空间的方式。他们提出了Q-learning的一个结合因果模型的方法，对算法中的利用exploitation部分进行了修正。

END



自/强/不/息 止/于/至/善

Thank You