上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

清源研究院 QING YUAN RESEARCH INSTITUTE

SJTU

自动驾驶决策规划

蔡盼盼
上海交通大学

2024年5月10日

关于我

蔡盼盼
上海交通大学
清源研究院

教育背景

2007 - 2011	浙江大学	学士
2011 – 2016	新加坡南洋理工大学	博士

工作经历

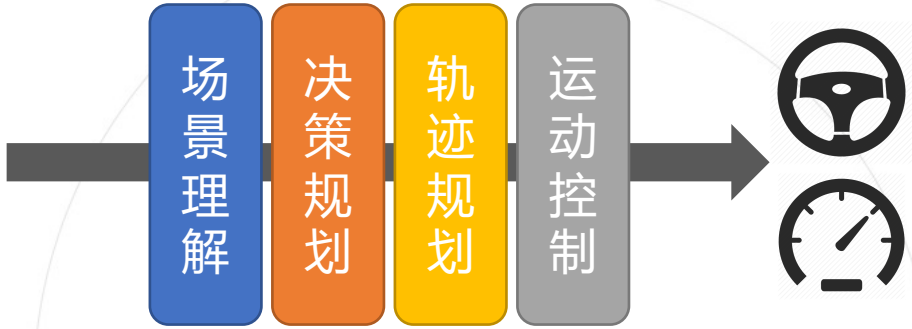
2017 - 2020	新加坡国立大学	博士后
2021 - 2022	新加坡国立大学	高级博士后
2022 至今	上海交通大学	副教授、博导

2022 上海市领军人才（海外）
2023 国家海外优青
2024-2027 机器人顶刊T-RO编委

自动驾驶能力堆栈

- **场景理解**：场景里有什么元素？是什么实体？具有什么行为？将如何与自车交互？
- **决策规划**：自车应当采取什么行为？
- **轨迹规划**：自车应当采取什么运动轨迹？
- **运动控制**：应当如何控制车辆？

模块化系统（RobotTaxi）



模块化系统（RobotTaxi）

场景理解 | 决策规划 | 轨迹规划 | 运动控制

基于对城市环境的结构化模型
进行显式的推理、规划与决策

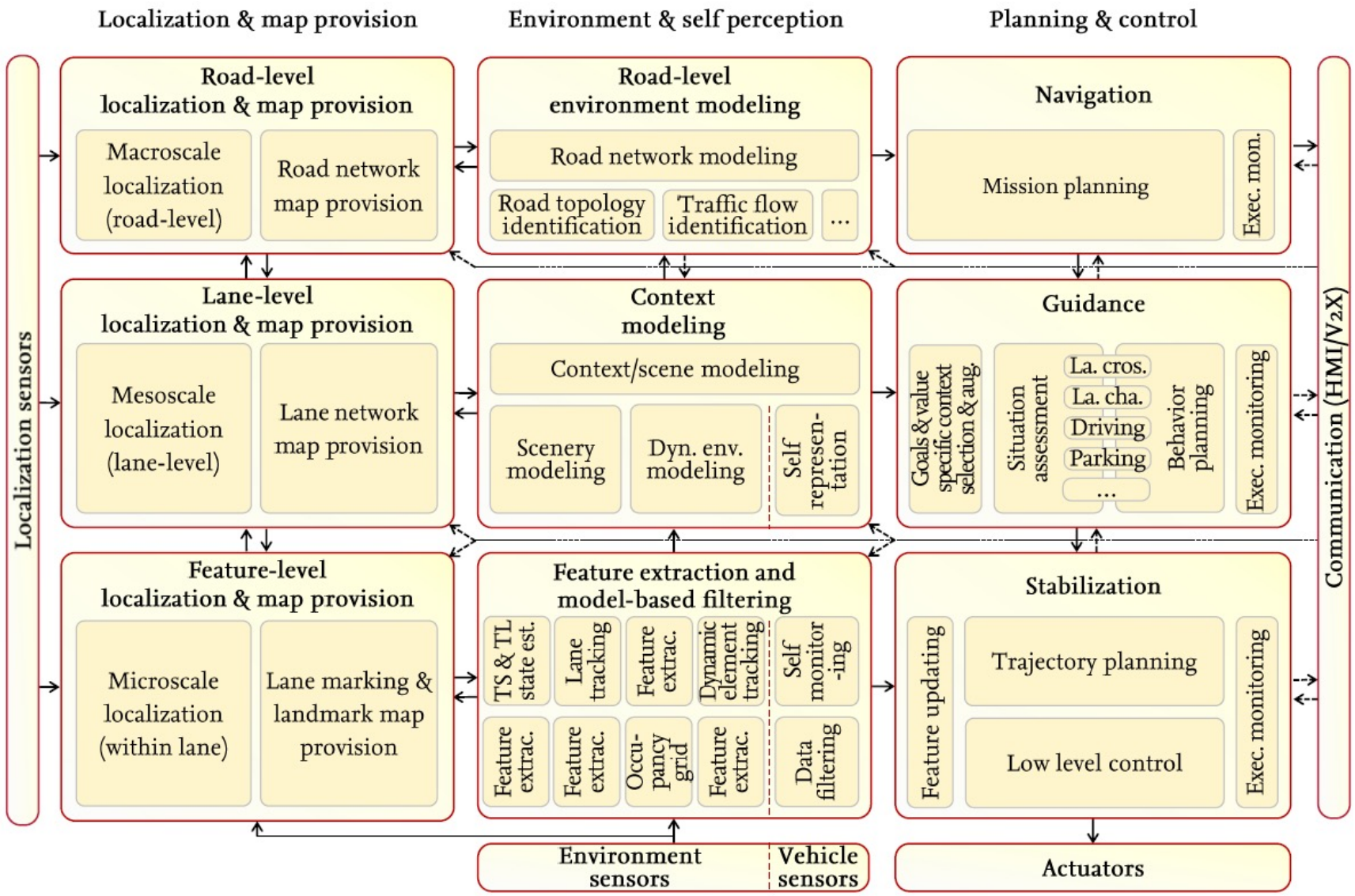
模块化系统（RobotTaxi）

场景理解

决策规划

轨迹规划

运动控制



模块化系统（RobotTaxi）

场景理解

决策规划

轨迹规划

运动控制

模块化系统面临的挑战

非结构化场景：难以对环境结构进行提前建模或现场理解的场景

显式场景理解
具有长尾问题

端到端系统

通过从数据中直接学习，解决显式场景理解带来的系统瓶颈

端到端系统面临的挑战

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

安全性、可靠性：神经网络的决策无法确保安全性和一致性，缺乏可解释性

ID	Method	Ego Status in BEV	Ego Status in Planer	L2 (m) ↓ 1s	2s	3s	Avg.	Collision (%) ↓ 1s	2s	3s	Avg.	Intersection (%) ↓ 1s	2s	3s	Avg.	ckpt. source
0	ST-P3	✗	✗	1.59†	2.64†	3.73†	2.65†	0.69†	3.62†	8.39†	4.23†	2.53†	8.17†	14.4†	8.37†	Official
1	UniAD	✗	✗	0.59	1.01	1.48	1.03	0.16	0.51	1.64	0.77	0.35	1.46	3.99	1.93	Reproduce
2	UniAD	✓	✗	0.35	0.63	0.99	0.66	0.16	0.43	1.27	0.62	0.21	**1.32**	3.63	1.72	Official
3	UniAD	✓	✓	0.20	0.42	0.75	0.46	0.02	**0.25**	0.84	0.37	**0.20**	1.33	**3.24**	**1.59**	Reproduce
4	VAD-Base	✗	✗	0.69	1.22	1.83	1.25	0.06	0.68	2.52	1.09	1.02	3.44	7.00	3.82	Reproduce
5	VAD-Base	✓	✗	0.41	0.70	1.06	0.72	0.04	0.43	1.15	0.54	0.60	2.38	5.18	2.72	Official
6	VAD-Base	✓	✓	0.17	0.34	0.60	0.37	0.04	0.27	**0.67**	**0.33**	0.21	2.13	5.06	2.47	Official
7	GoStright	-	✓	0.38	0.79	1.33	0.83	0.15	0.60	2.50	1.08	2.07	8.09	15.7	8.62	-
8	Ego-MLP	-	✓	**0.15**	**0.32**	0.59	**0.35**	**0.00**	0.27	0.85	0.37	0.27	2.52	6.60	2.93	
9	BEV-Planner*	✗	✗	0.27	0.54	0.90	0.57	0.04	0.35	1.80	0.73	0.63	3.38	7.93	3.98	-
10	BEV-Planner	✗	✗	0.30	0.52	0.83	0.55	0.10	0.37	1.30	0.59	0.78	3.79	8.22	4.26	-
11	BEV-Planner+	✓	✗	0.28	0.42	0.68	0.46	0.04	0.37	1.07	0.49	0.70	3.77	8.15	4.21	-
12	BEV-Planner++	✓	✓	0.16	0.32	**0.57**	**0.35**	**0.00**	0.29	0.73	0.34	0.35	2.62	6.51	3.16	-

可扩展性挑战！

上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

大规模交互场景：与大量交通参与者的实时交互

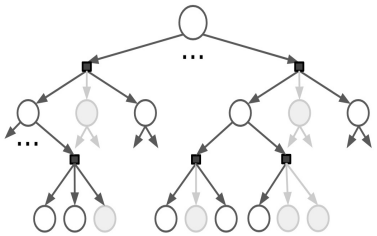
过度激进：在未充分考虑各种风险的情况下，做出草率的行为
过度保守：在各类风险下自车不敢作为，导致交通拥堵与混乱

大规模交互场景与可扩展性挑战

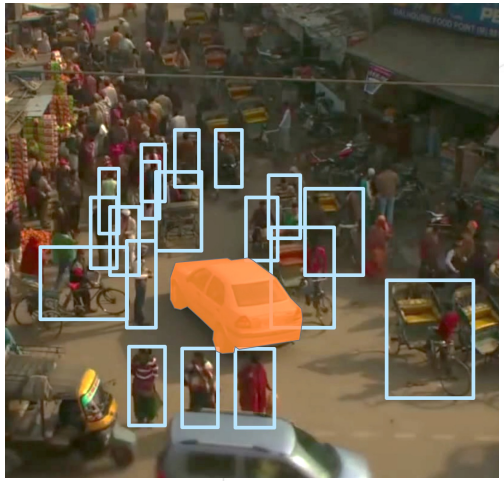
模块化系统

- 场景理解
- 决策规划
- 轨迹规划
- 运动控制

规划

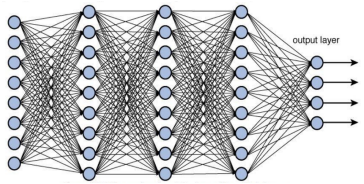


计算量！



数据量！

学习



端到端系统

- 场景理解
- 决策规划
- 轨迹规划
- 运动控制

理想的自动驾驶系统？

安全性、舒适性、解释性

模块化通路

- 场景理解
- 决策规划
- 轨迹规划
- 运动控制

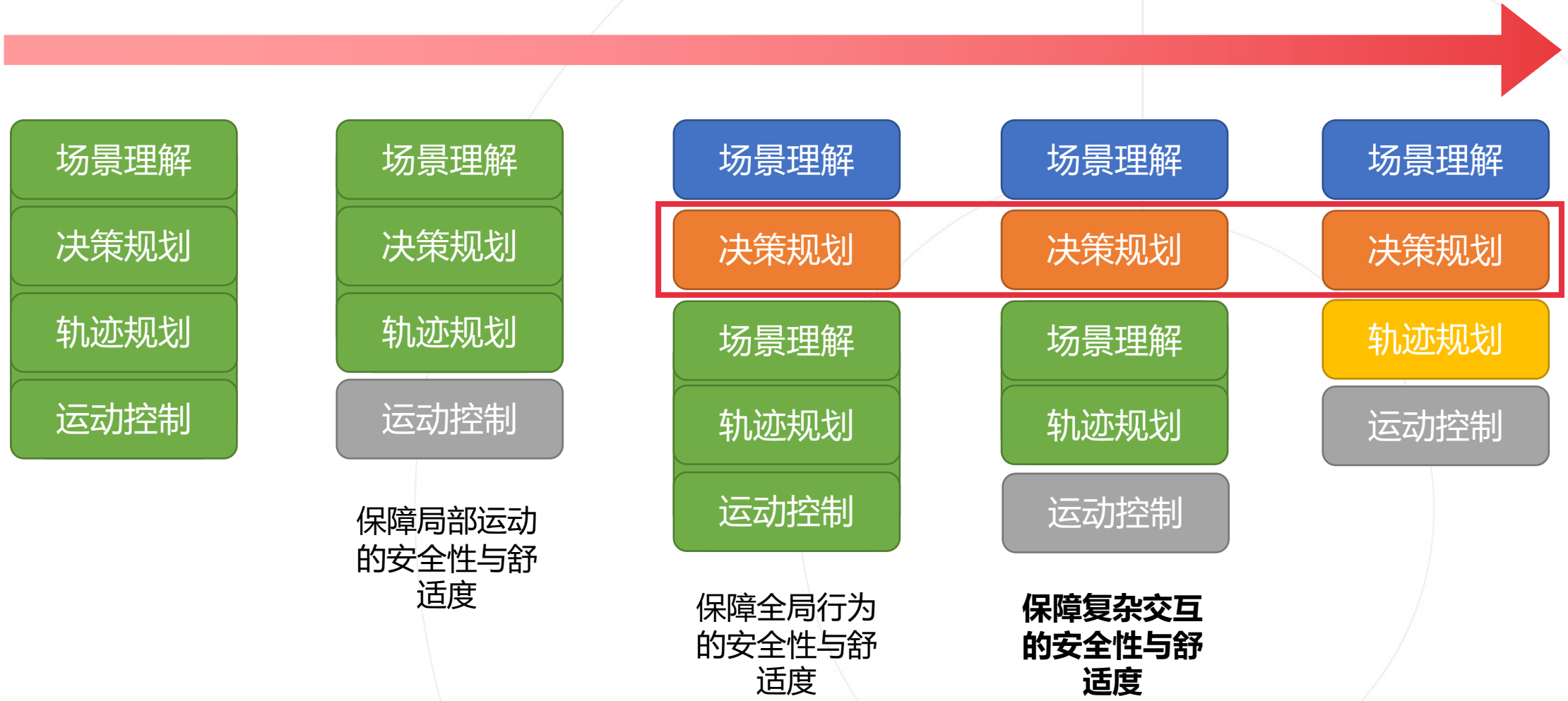
模块化与端到端的融合

端到端通路

- 场景理解
- 决策规划
- 轨迹规划
- 运动控制

解决场景理解长尾问题

模块化系统与端到端系统的融合



Step 1：分析问题结构

Step 2：设计规划算法

Step 3：实用算法优化

Step 1：分析问题结构

Step 2：设计规划算法

Step 3：实用算法优化

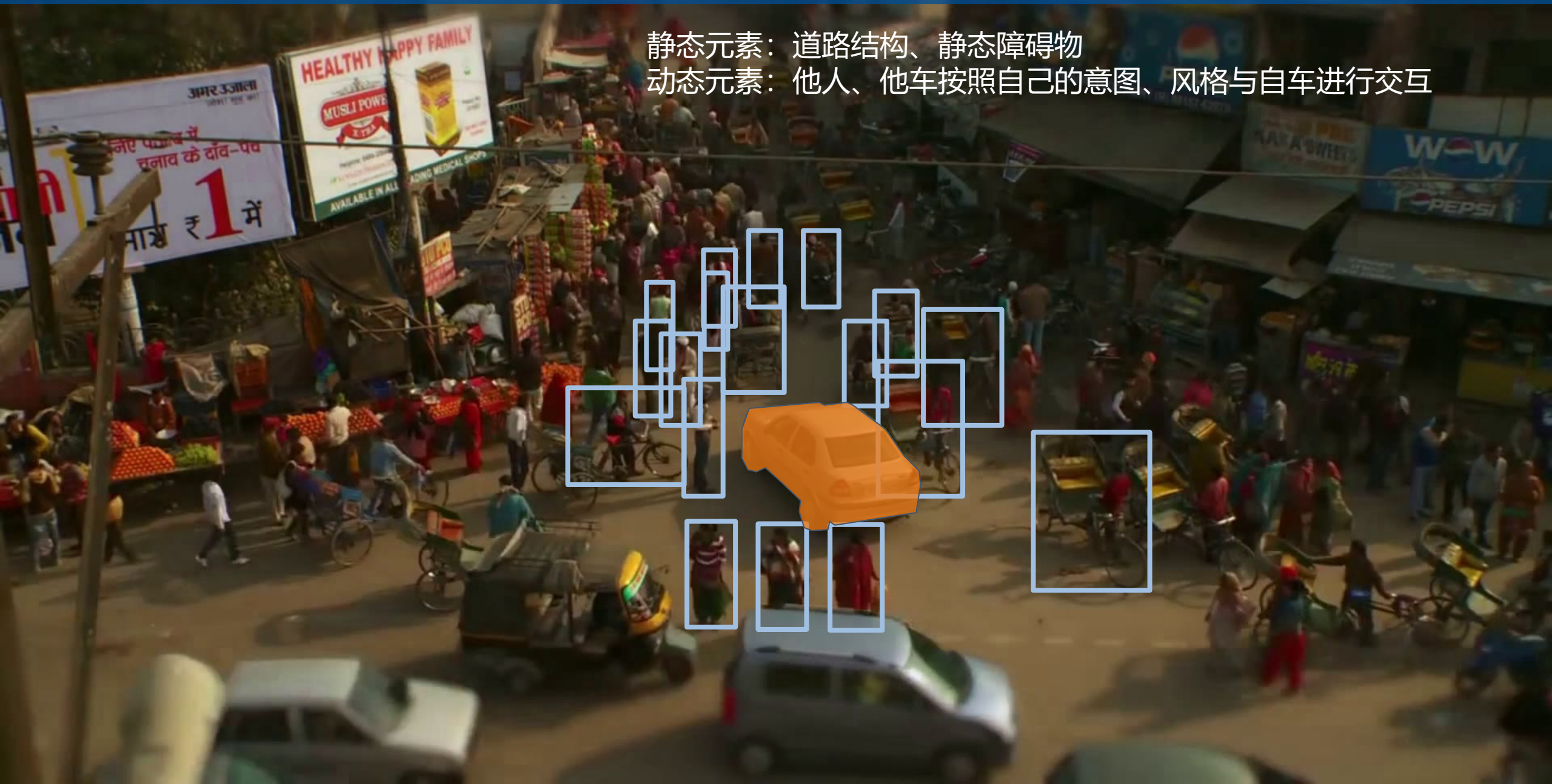

HEALTHY HAPPY FAMILY
MUSLI POWE
AVAILABLE IN ALL
PEPSI
W-W
PEPSI
₹1 में

自车

目标是在避免与他车碰撞的同时，尽快穿过这个混乱的十字路口。

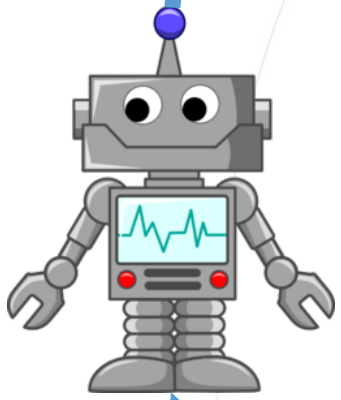
环境

静态元素：道路结构、静态障碍物
动态元素：他人、他车按照自己的意图、风格与自车进行交互



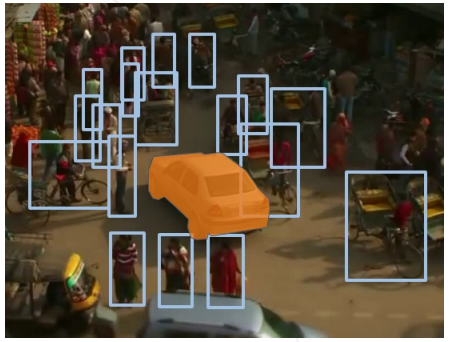
问题的抽象建模（上帝视角）

自动驾驶系统
Agent



动作
Action

物理世界
World

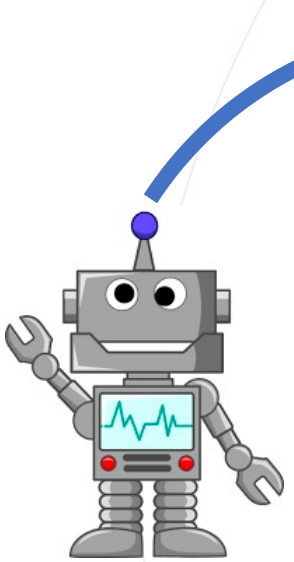


感知
Perception

马尔科夫决策过程（Markov Decision Process）

MDP 模型具有5个元素: $< S, A, T, R, \gamma >$

状态空间、动作空间



动作 $a \in A$

动作 a 表示自车的行为
底层行为：方向盘角度、加速度
高层行为：跟车、变道、避让，...

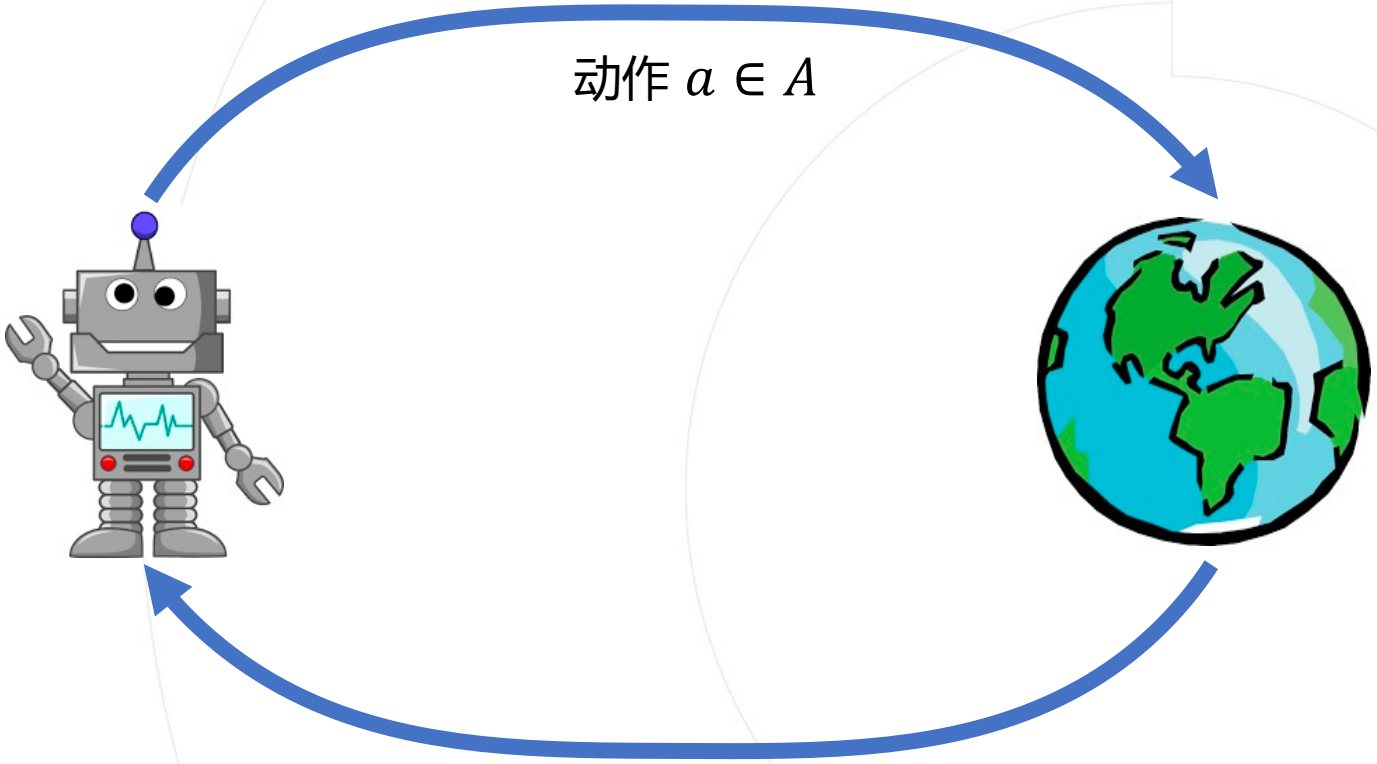
状态 $s \in S$

状态 s 表示自车与他人/车的几何、运动学、行为状态

马尔科夫决策过程（Markov Decision Process）

MDP 模型具有5个元素: $< S, A, T, R, \gamma >$

状态转移、奖励函数　折扣因子



若当前世界处在状态 s, 自车执行动作 a, 世界下一步转移到状态 s′ 的概率是多少?

状态 $s \in S$

$T(s, a, s') = p(s' \mid s, a)$

下一步状态 $s' \in S$

指定机器人任务的方式：若机器人在世界状态 s 执行动作 a, 所获得的即时奖励 r 是多少?

MDP 决策规划

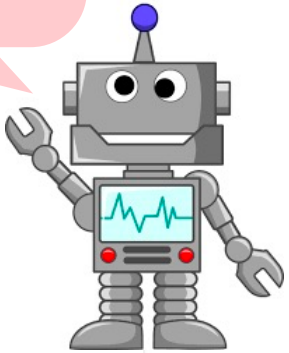
定义：利用 MDP 模型，求解机器人的最优闭环策略

状态 $s \in S$
世界的真实状态

策略 $\pi: S \to A$
为每个状态 s 指定一个动作 a

选择 $a = \pi(s)$

动作 $a \in A$



$< S, A, T, R, \gamma >$

状态 $s \in S$

$T(s, a, s') = p(s' \mid s, a)$

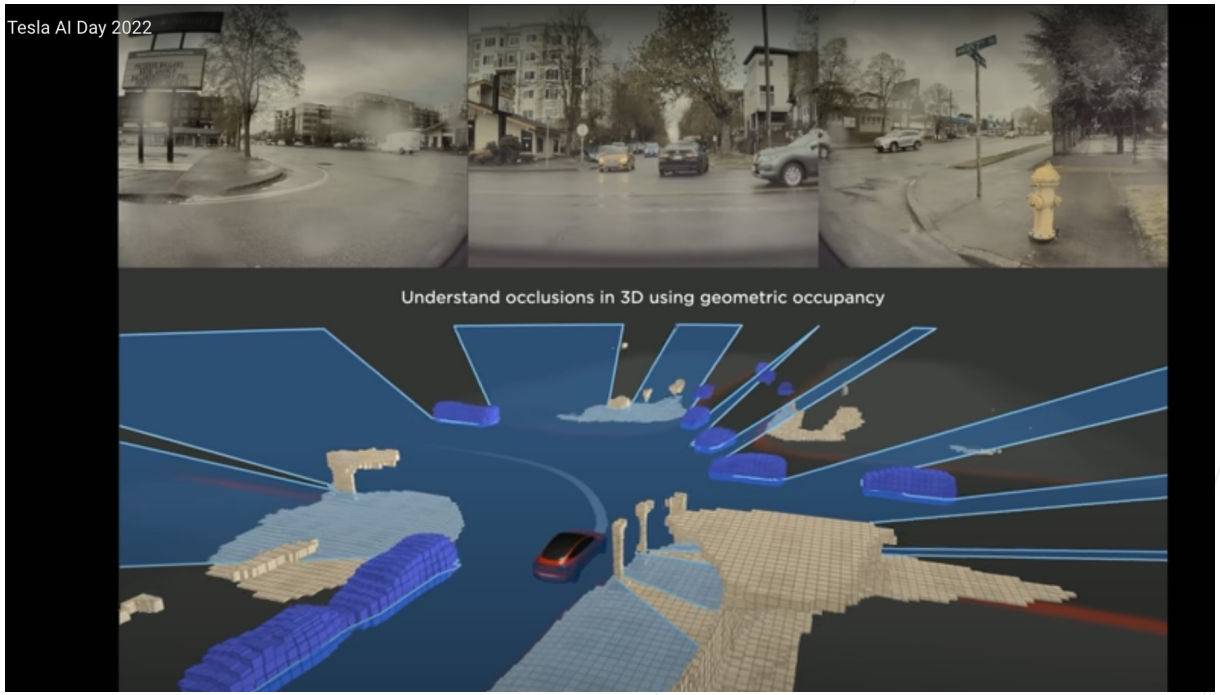
下一步状态 $s' \in S$

感知 s'

奖励 r

$r = R(s, a)$

例：Tesla 的 MDP 模型（2022）

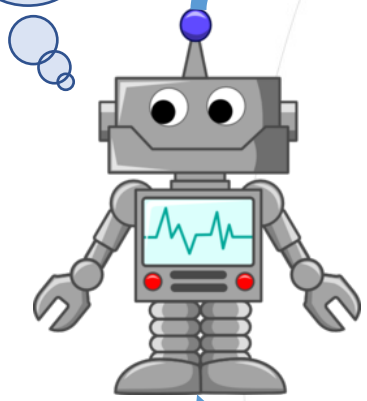


例： Tesla 的 MDP 模型（2022）

动作 a：多层动作

第一层：局部车道图上的不同目标位置
第二层：是否对个体一进行避让
第三层：是否对个体二进行避让
…

状态 s：
自车与他人/车的几何与运动学状态



$< S, A, T, R, \gamma >$

状态转移T：联合轨迹优化

给定当前状态、目标位置与交互方式，由神经网络生成所有参与者的初始轨迹，由最优化方法生成所有参与者的最终轨迹

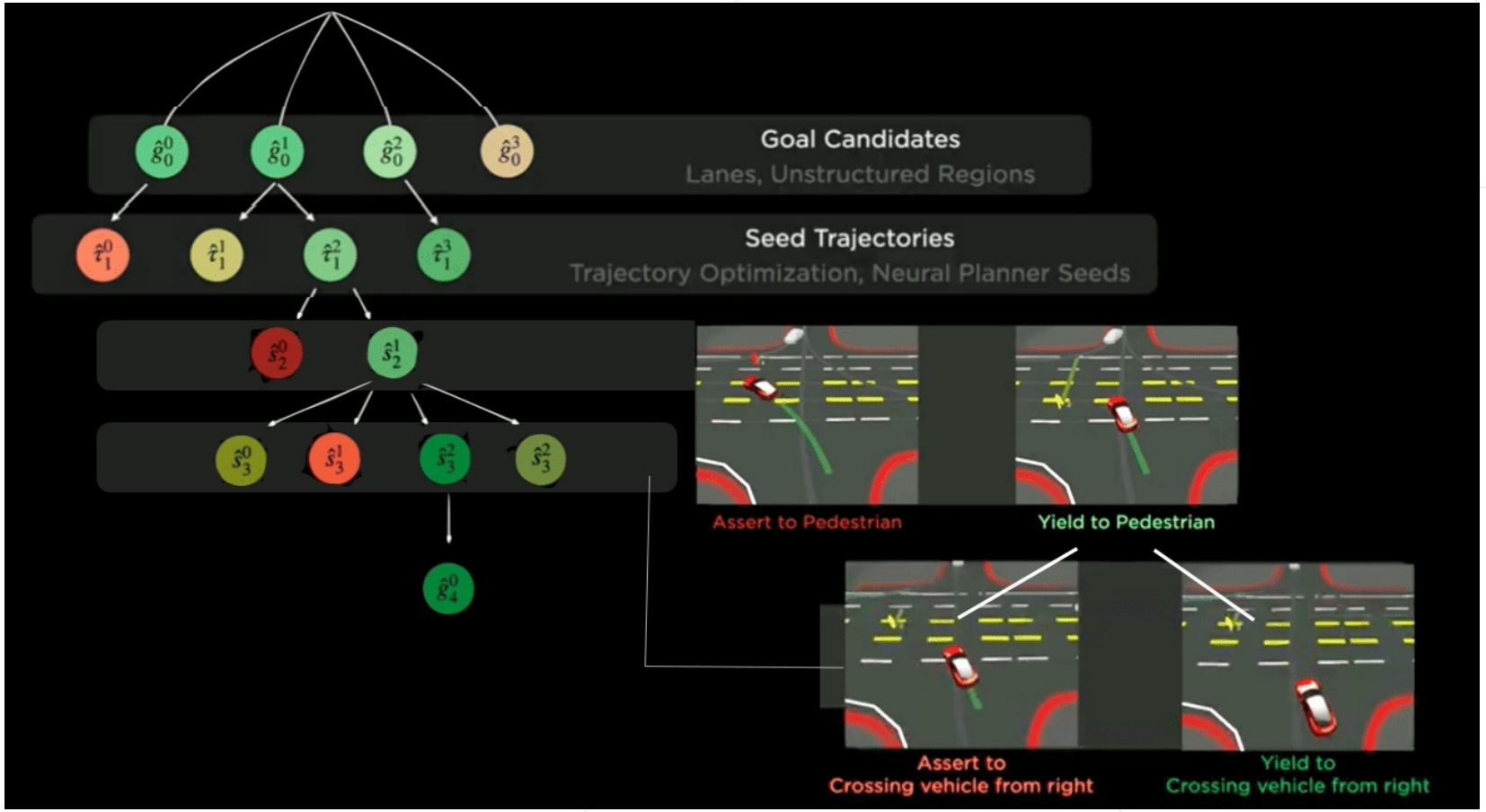
奖励函数R：轨迹打分

显式打分：碰撞检测、舒适度分析
学习打分：接管可能性、行为的类人程度

完美感知，直接获取世界状态 s

蒙特卡洛树搜索 (MCTS)

算法：通过推演未来可能发生的情况，从而计算自车的最优策略



例： Tesla 的 MDP 模型（2022）

动作 a：分层决策
第一层：局部车道图上的不同目标位置
第二层：是否对个体一进行避让
第三层：是否对个体二进行避让
…

状态 s：
自车与他人/车的几何与运动学状态

① 认为可以准确知道他人的运动学状态

② 对他人/车运动轨迹进行了确定性假设

状态转移：对自车与他车的联合轨迹优化
给定当前状态、目标位置与交互决策，由神经网络生成初始轨迹，由优化方法生成最终轨迹

奖励函数：轨迹打分
显式打分：碰撞检测、舒适度分析
深度学习打分：接管可能性、行为的类人程度

完美感知，直接获取世界状态 s

例： Tesla 的 MDP 模型（2022）

动作 a：分层决策
第一层：局部车道图上的不同目标位置
第二层：是否对个体一进行避让
第三层：是否对个体二进行避让
…

状态 s：
自车与他人/车的几何与运动学状态

① 认为可以准确知道他人的运动学状态

② 对他人/车运动轨迹进行了确定性假设

状态转移：对自车与他车的联合轨迹优化
给定当前状态、目标位置与交互决策，由神经网络生成初始轨迹，由优化方法生成最终轨迹

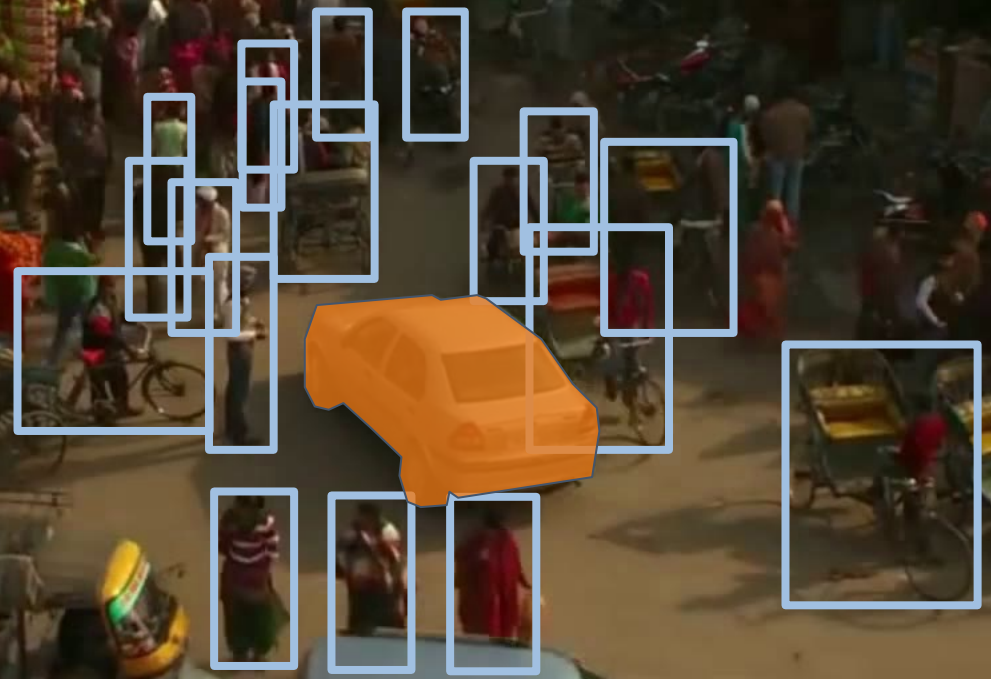
奖励函数：轨迹打分
显式打分：碰撞检测、舒适度分析
深度学习打分：接管可能性、行为的类人程度

完美感知，直接获取世界状态 s

感知的不确定性

对他人、他车状态的感知只能通过传感器获得
感知模块的输出常常带有噪音和错误

他人、他车的几何与运动学状态是隐藏状态，只能通过观察到的信息进行概率推断



例： Tesla 的 MDP 模型（2022）

动作 a：分层决策
第一层：局部车道图上的不同目标位置
第二层：是否对个体一进行避让
第三层：是否对个体二进行避让
…

状态 s：
自车与他人/车的几何与运动学状态

① 认为可以准确知道他人的运动学状态

② 对他人/车运动轨迹进行了确定性假设

状态转移：对自车与他车的联合轨迹优化
给定当前状态、目标位置与交互决策，由神经网络生成初始轨迹，由优化方法生成最终轨迹

奖励函数：轨迹打分
显式打分：碰撞检测、舒适度分析
深度学习打分：接管可能性、行为的类人程度

完美感知，直接获取世界状态 s

人类行为的不确定性


依据历史轨迹无法确定性地预测未来的运动
HEALTHY HAPPY FAMILY
MUSLI POWER
AVAILABLE IN ALL LEADING MEDICAL SHOPS
अमर उजाला
चुनाव के दाँव-पेच
₹1 में
PEPSI
W-W
PEPSI

降低人类行为的不确定性？

意图：目标路径、目的地、目标行为，....

将他人的意图纳入世界状态，可以更好地预测他人的行为
[RAL'18, ICRA'20, RAL' 22]

人类意图的不确定性

他人可能有多种意图，但我们无法直接观察！

意图是隐藏状态，只能透过观察进行概率推断

贝叶斯推断 [RAL'18, RAL' 22]

信念更新公式：

$$b'(s') = \eta O(o|s', a) \sum_{s \in S} T(s'|s, a)\, b(s)$$

历史轨迹　　转移概率：GAMMA

关于意图的概率分布

信念（Belief）

对所有个体进行推断，获得关于场景内所有个体意图的联合分布。

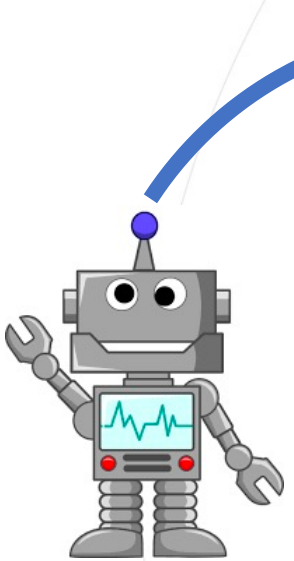
部分可观马尔科夫决策过程 (POMDP)

Partial observability: 真实世界的状态只能被间接地、部分地、有噪音地感知

部分可观马尔科夫决策过程 (POMDP)

POMDP 模型具有7个元素: $< S, A, Z, T, O, R, \gamma >$

状态、动作、观察空间



动作 $a \in A$

动作 a 表示自车的行为
底层行为：方向盘角度、加速度
高层行为：跟车、变道、避让，…

状态 $s \in S$

表示自车与他人/车的几何、运动学与行为状态

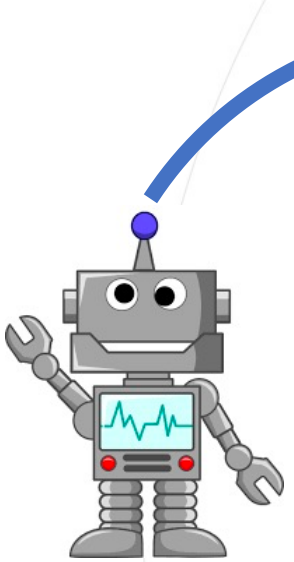
图像数据、
点云数据、
几何与运动学信息，
…

观察 $z \in Z$

部分可观马尔科夫决策过程 (POMDP)

POMDP 模型具有7个元素: $< S, A, Z, T, O, R, \gamma >$

状态转移、观察、奖励函数



动作 $a \in A$

状态 $s \in S$

$T(s, a, s') = p(s' \mid s, a)$

下一步状态 $s' \in S$

对感知机制的建模:
若自车执行动作 a 使世界状态转移到 s′，获得观察 z 的概率是多少?

$O(z, a, s') = p(z|a, s')$

观察 $z \in Z$

奖励 r

$r = R(s, a)$

POMDP 决策规划

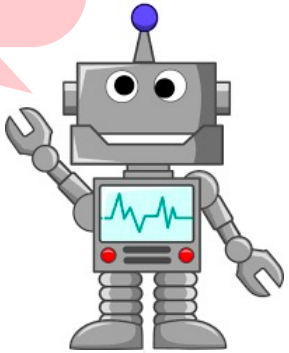
定义：利用 POMDP 模型，求解机器人的最优闭环策略

信念 $b \in \mathcal{B}$
关于世界状态的概率分布

策略 π: $\mathcal{B} \rightarrow A$
为每个信念 b 指定一个动作 a

选择 $a = \pi(b)$

动作 $a \in A$



$< S, A, Z, T, O, R, \gamma >$

状态 $s \in S$

$T(s, a, s') = p(s' \mid s, a)$

下一步状态 $s' \in S$

$O(z, a, s') = p(z|a, s')$

观察 $z \in Z$

奖励 r

$r = R(s, a)$

混乱路况自动驾驶POMDP模型 [RSS'19, ICRA'20, CoRL' 22, T-RO'22]

Belief b

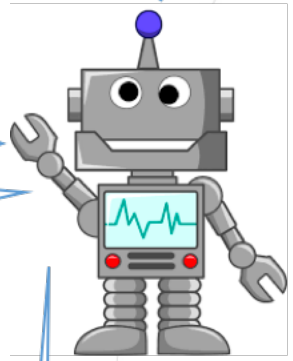
Distribution over states

Transition function T

GAMMA

Observation function O

Gaussian noise



Action a

Observation z

State s

Pos, vel, headings,

human behavioral states

Reward r

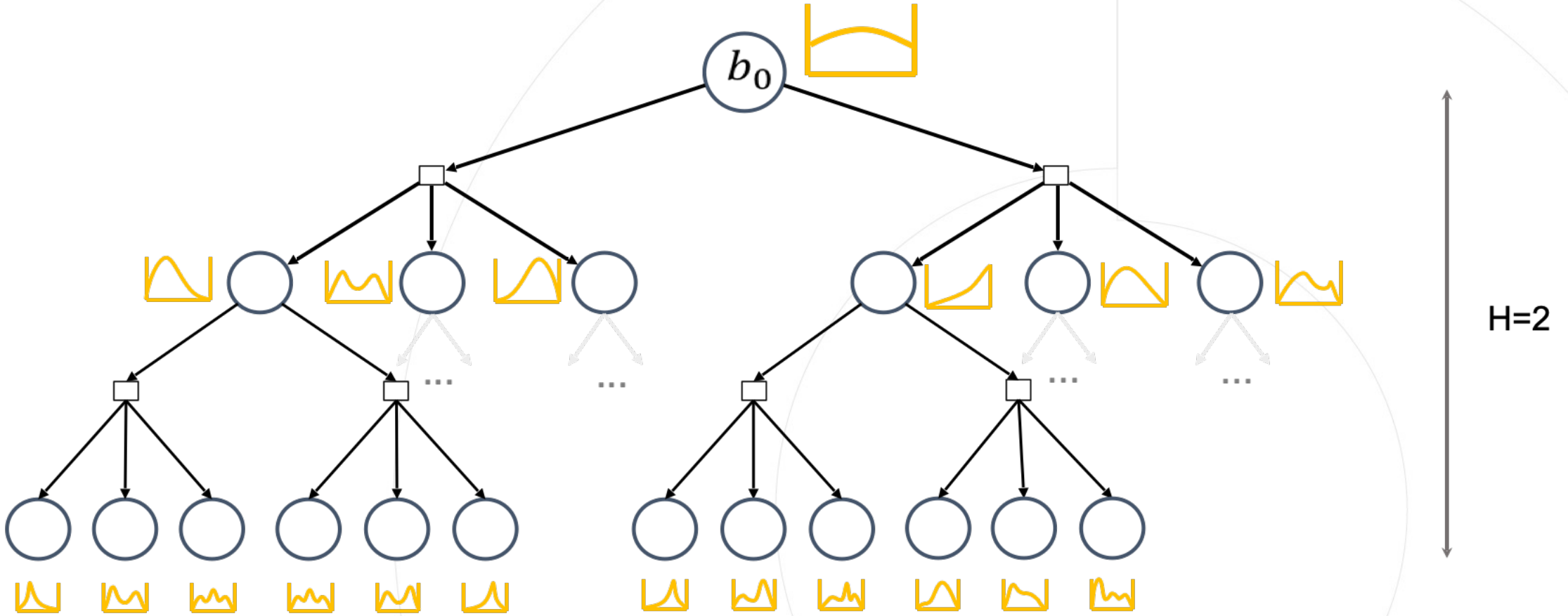
Safety, efficiency, smoothness

Step 1：分析问题结构

Step 2：设计规划算法

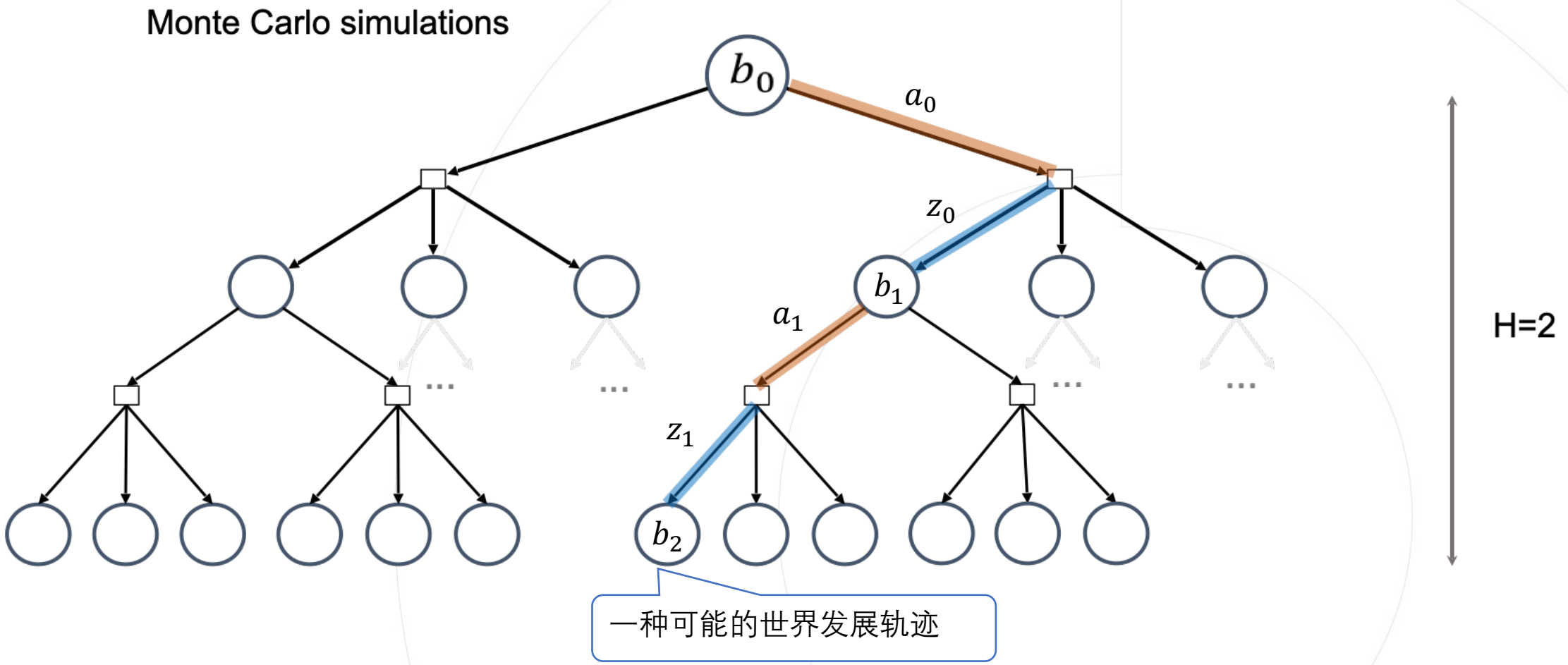
Step 3：实用算法优化

信念树（Belief Tree）

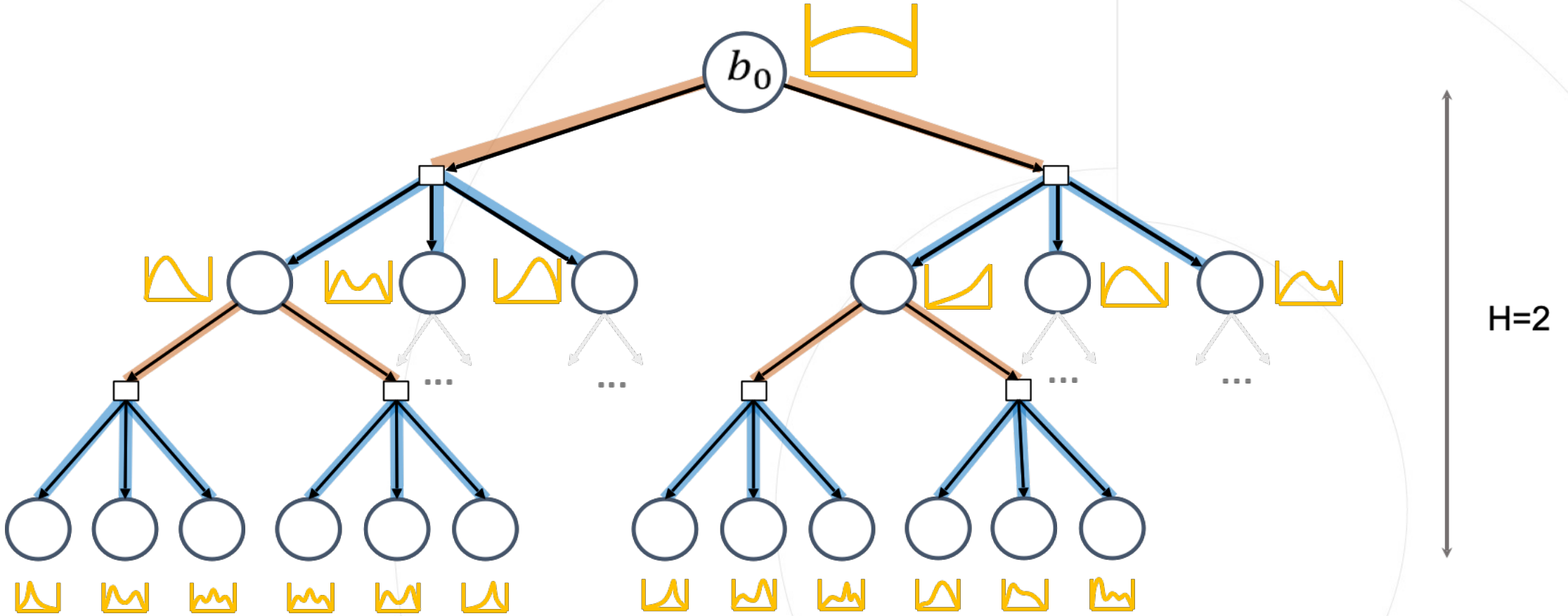


Belief tree with 2 actions and 3 observations

信念树（Belief Tree）

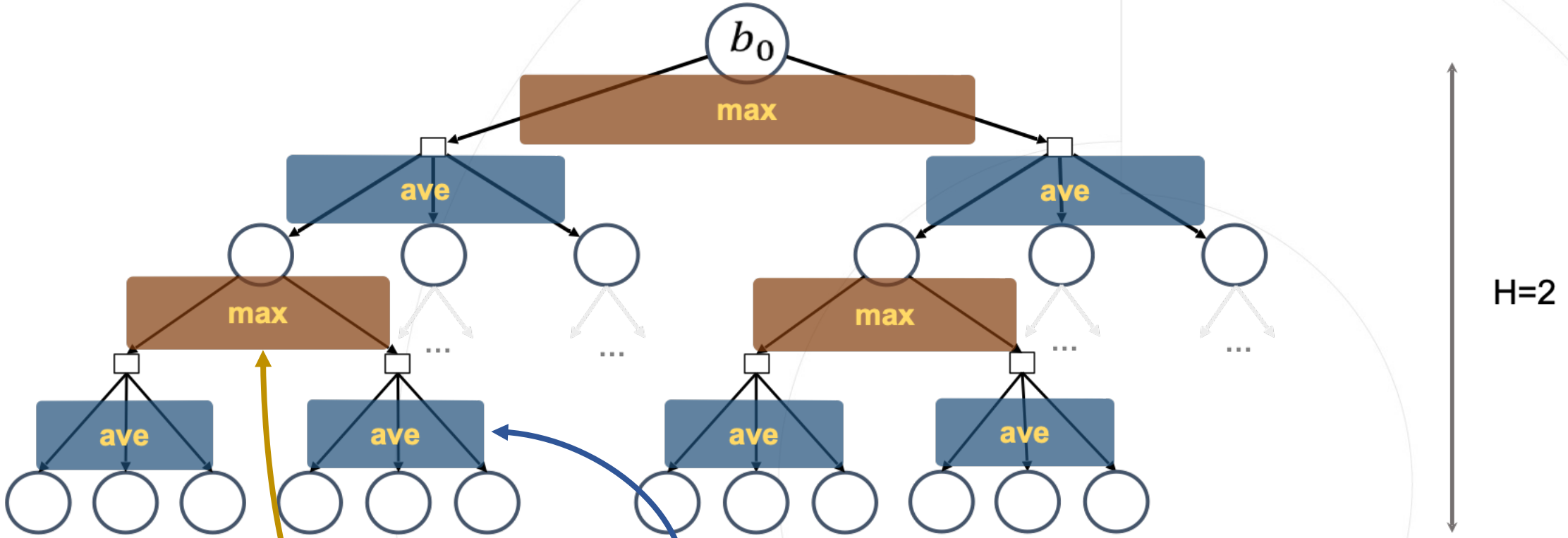


信念树（Belief Tree）



Belief tree with 2 actions and 3 observations

信念树搜索（Belief Tree Search）

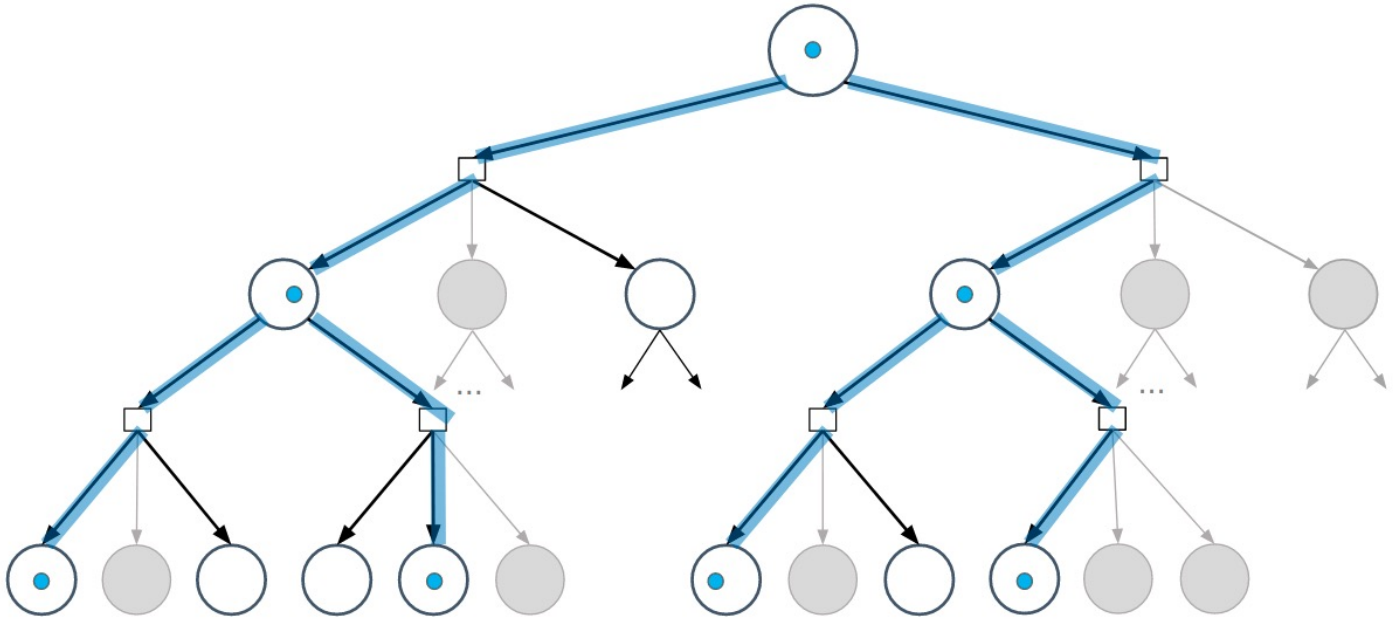


$$V^*(b) = \max_a R(b,a) + \gamma \sum_{z \in Z} p(z|b,a) V^*(b') \quad \text{(Bellman等式)}$$

计算复杂度：$O(|A|^H |Z|^H)$

DESPOT 算法

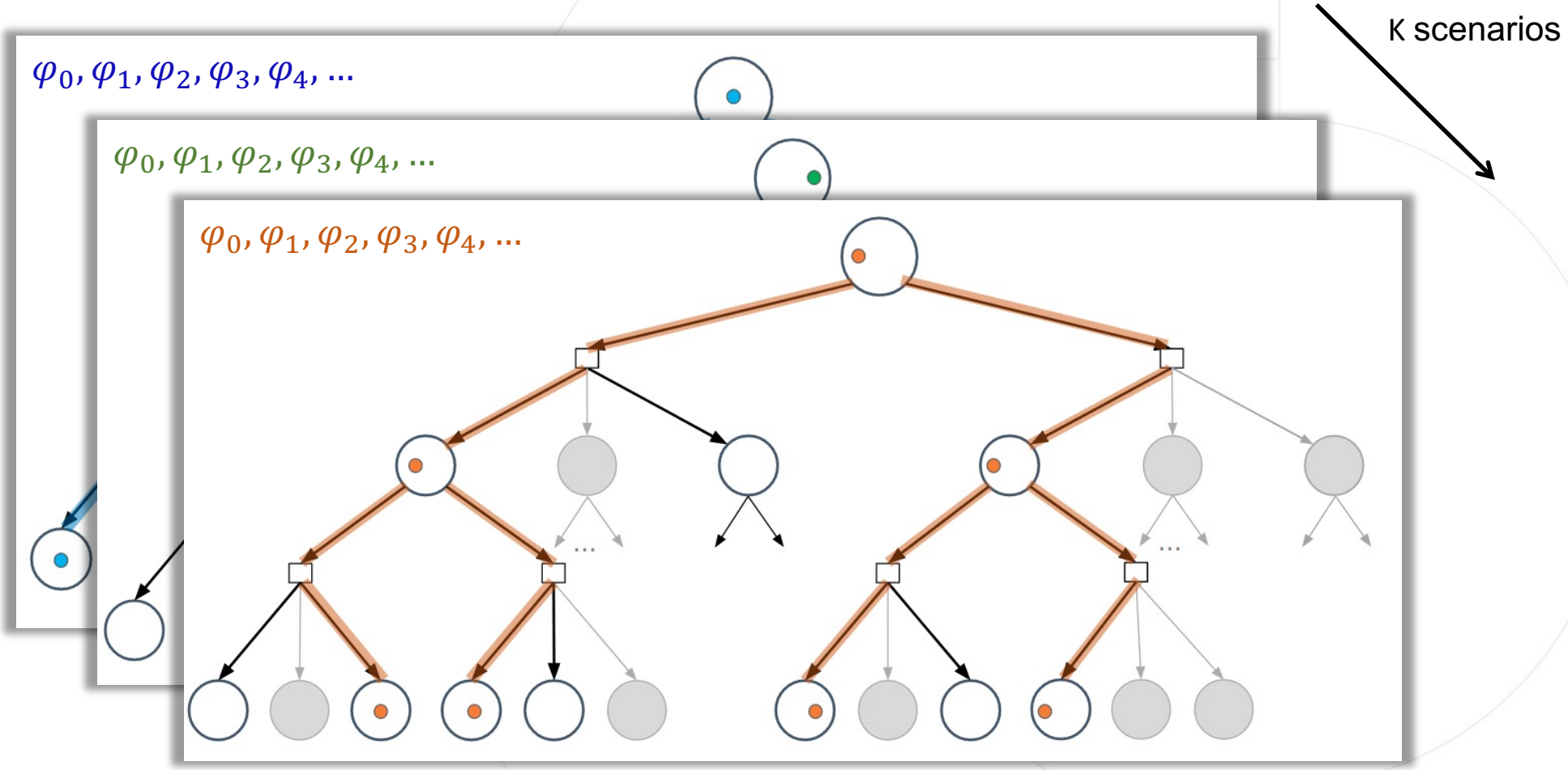
- 只考虑有限个情形（scenario），构造稀疏信念树（sparse belief tree），进行近似最优的决策
  - **Scenario**: 使用固定的 random seed $\{\varphi_0, \varphi_1, \varphi_2, \varphi_3, \varphi_4, \dots\}$ 进行蒙特卡洛模拟



复杂度： $O(|A|^H)$

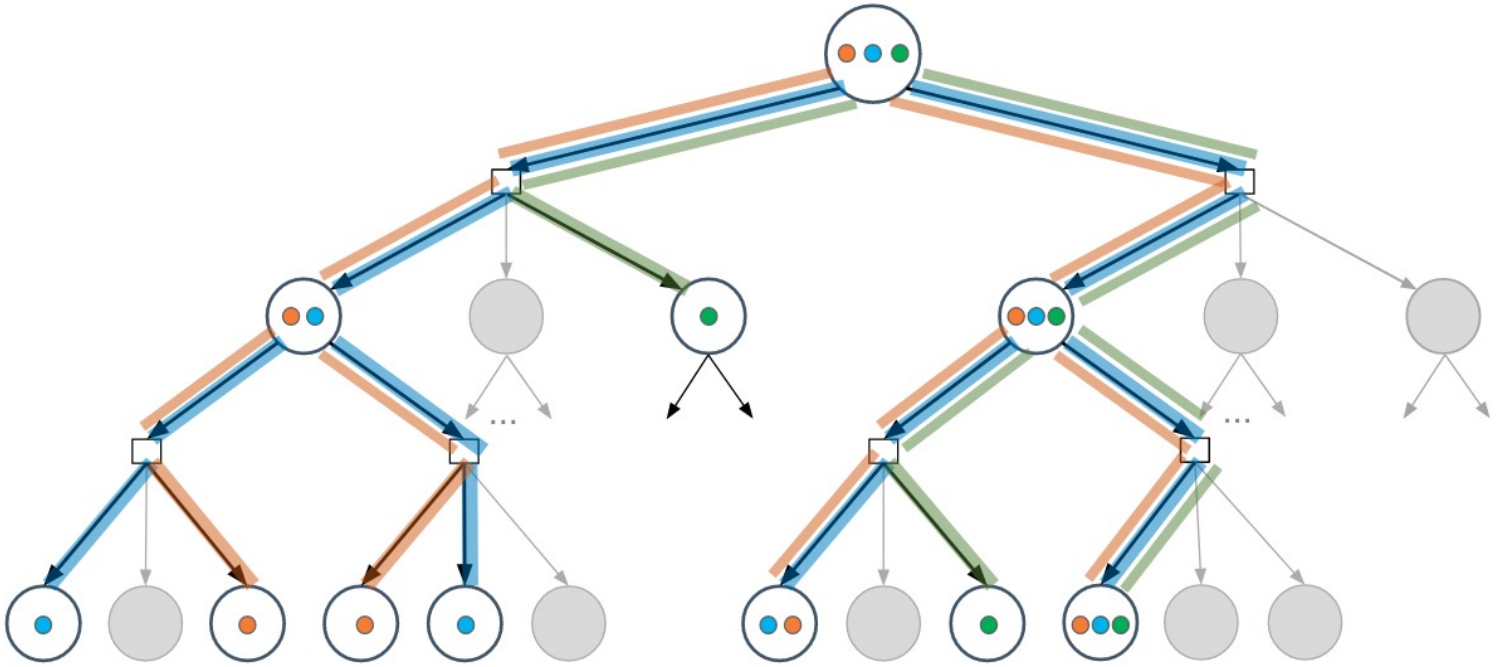
DESPOT 算法

- 只考虑有限个情形（scenario），构造稀疏信念树（sparse belief tree），进行近似最优的决策



DESPOT 算法

- 只考虑有限个情形（scenario），构造稀疏信念树（sparse belief tree），进行近似最优的决策

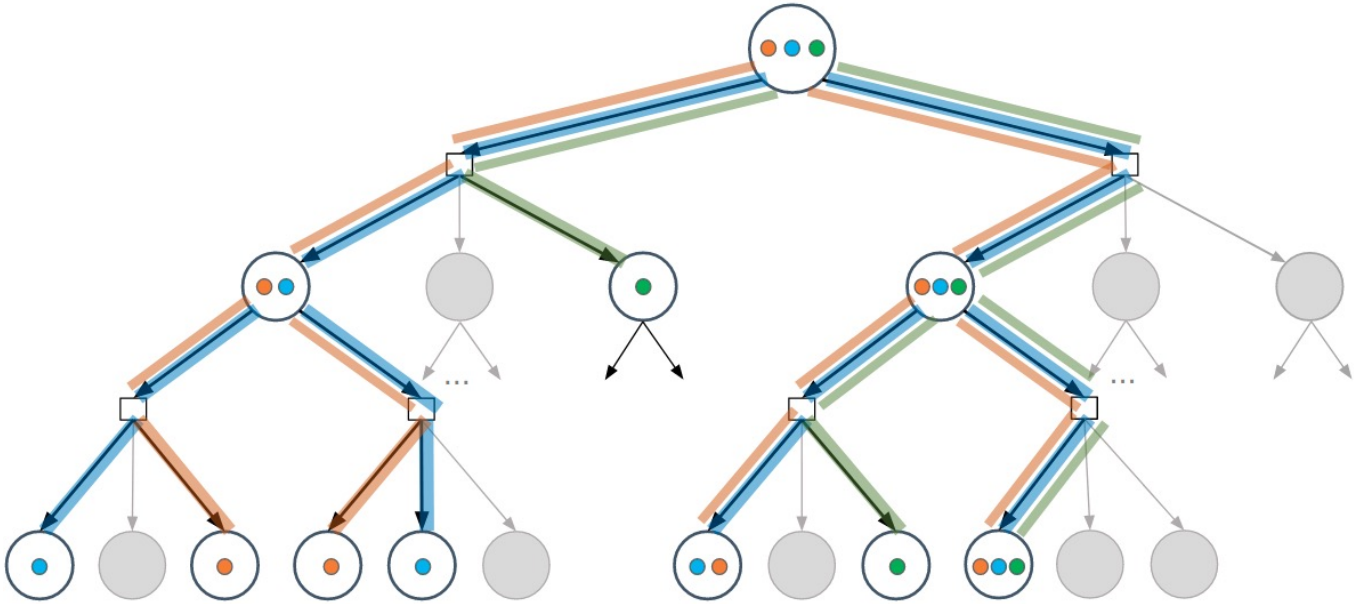


DEterminized Sparse Partially Observable Tree

复杂度：$O(|A|^H K)$

自动驾驶 POMDP 规划

DESPOT[1]: O(|*A*|^{H}|*Z*|^{H}) -> O(|*A*|^{H}K)



|A| = 9, H = 20, K = 100

DESPOT树大小: O(100*9^{20})

Step 1：分析问题结构

Step 2：设计规划算法

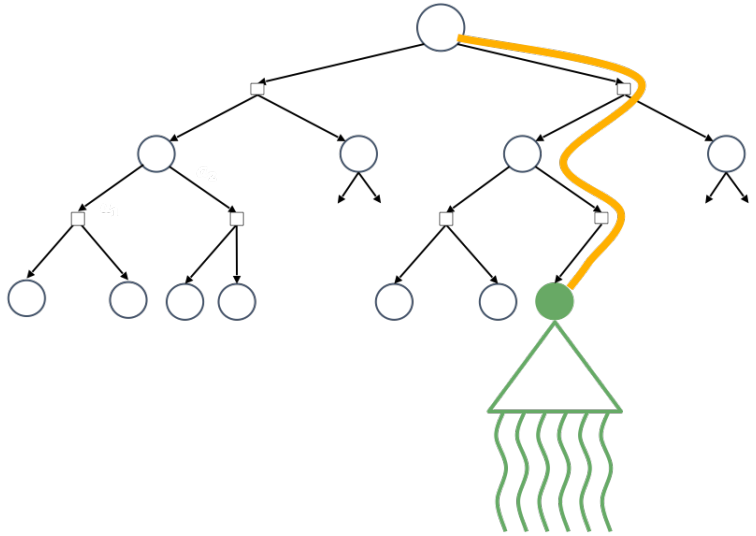
Step 3：实用算法优化

混合并行DESPOT （HyP-DESPOT）

[RSS'18, IJRR'20]

任务拆解与重整合

DESPOT



计算复杂度

$O(|A|^H K)$

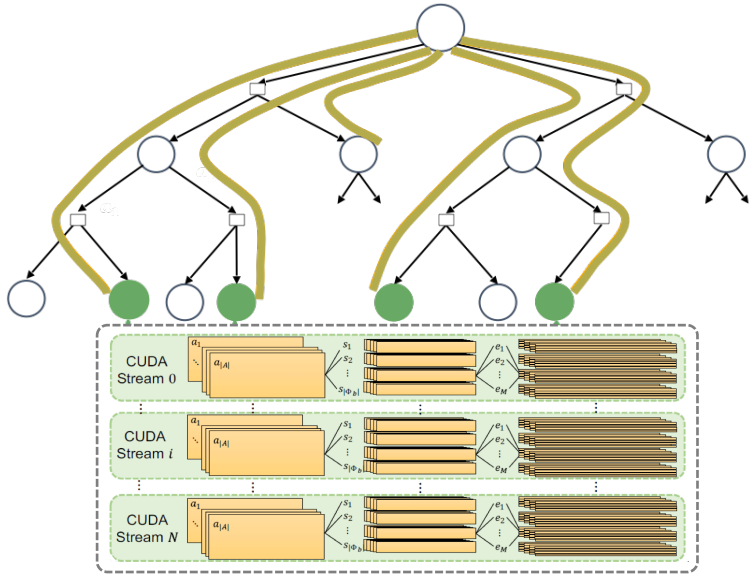
CPU并行
信念树搜索

- 灵活的数据结构
- 频繁的数据共享

GPU并行
蒙特卡洛 Rollout

- 独立的未来情形
- 相似的运算逻辑

HyP-DESPOT



400+倍计算加速

自动驾驶 POMDP 规划

上海交通大学 SHANGHAI JIAO TONG UNIVERSITY

[RSS'18, IJRR'20]

问题解决？

- 指数复杂度 $O(|A|^H K)$ / N:
  - ➔ 只适用于少量动作、短期规划
- 采样不足:
  - ➔ K 个情形对未来可能性的覆盖密度随着 H 指数级下降
- 模型误差累积:
  - ➔ 随着时域 H 的增长，预测变得越来越不准确

问题关键

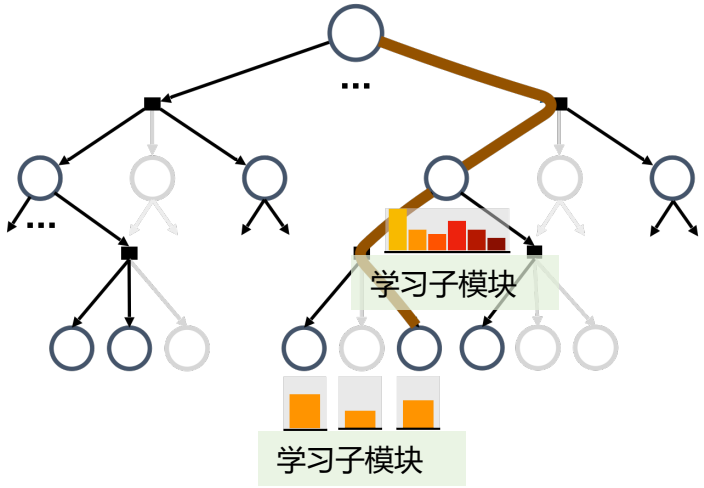
- 指数复杂度 $O(|A|^H K)$ / N:
  - ➔ 只适用于少量动作、短期规划
- 采样不足:
  - ➔ K 个情形对未来可能性的覆盖密度随着 H 指数级下降
- 模型误差累积:
  - ➔ 随着时域 H 的增长，预测变得越来越不准确

长期规划，但避免深度搜索?

融合规划与学习 [RSS'19, RSS'20, T-RO'22, CoRL'22最佳论文题名]

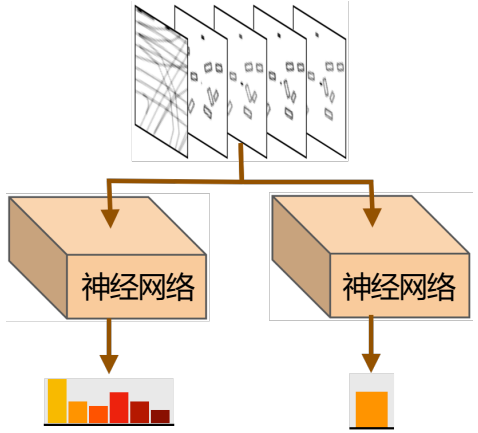
提供高质量数据

规划全局策略



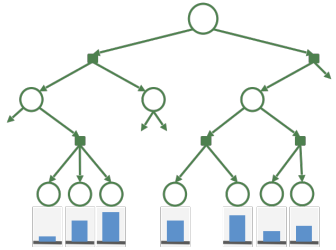
闭环
共同收敛

学习关键模块

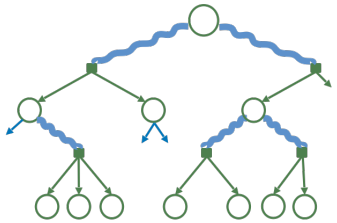


提供高质量模块

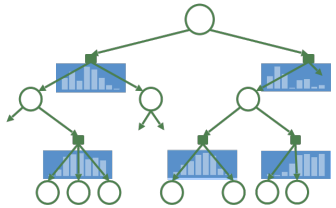
学习**搜索启发**



学习**机器人宏行为**

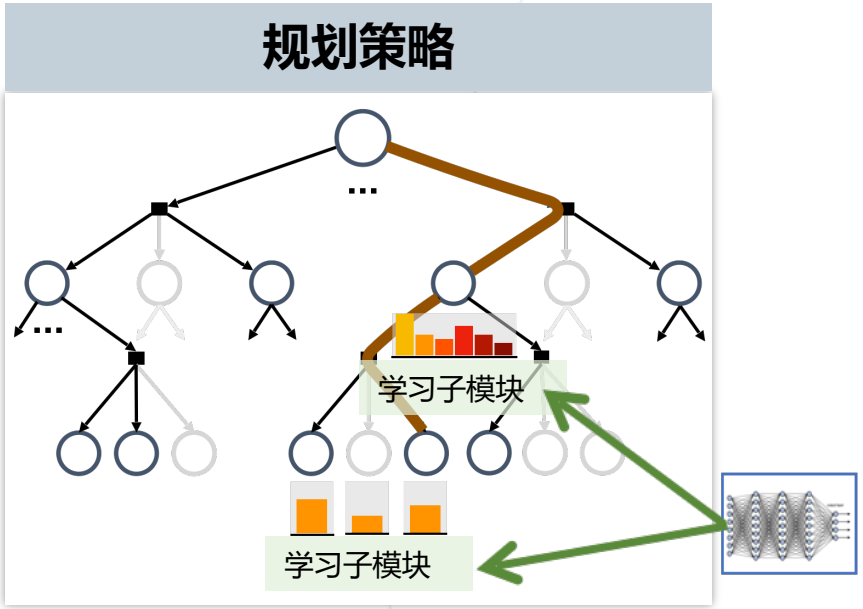


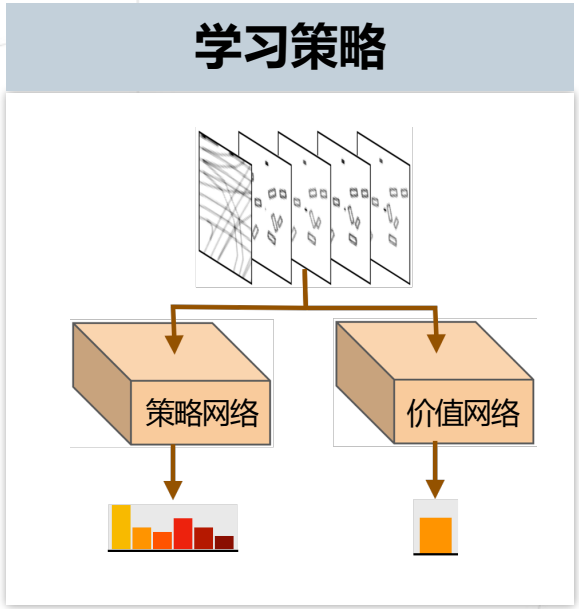
学习**重要性采样**



LeTS-Drive: Driving in a Crowd by Learning from Tree Search

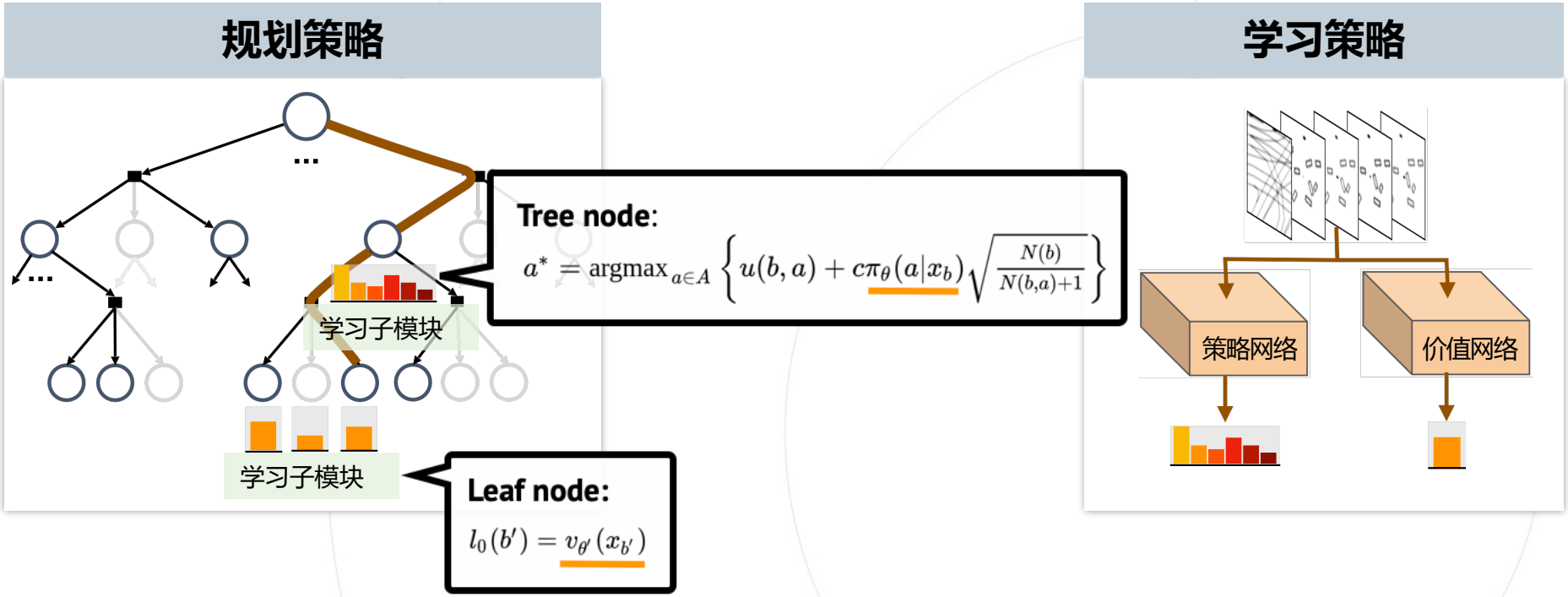
[RSS'19, T-RO'22]





LeTS-Drive: Driving in a Crowd by Learning from Tree Search

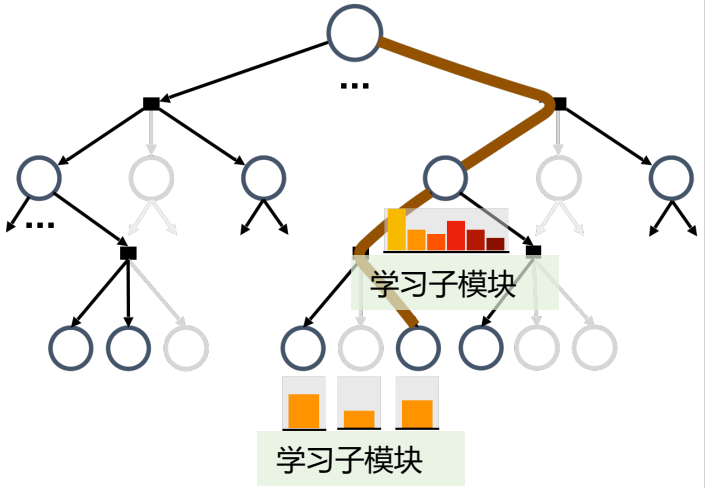
[RSS'19, T-RO'22]



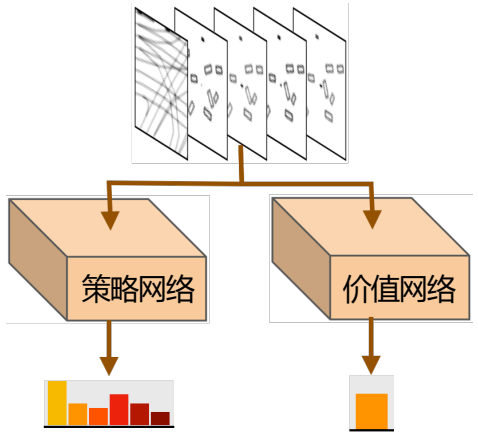
针对实时问题进一步优化！

训练数据？

规划策略



学习策略



提供搜索启发

自监督学习

提供策略与价值标签

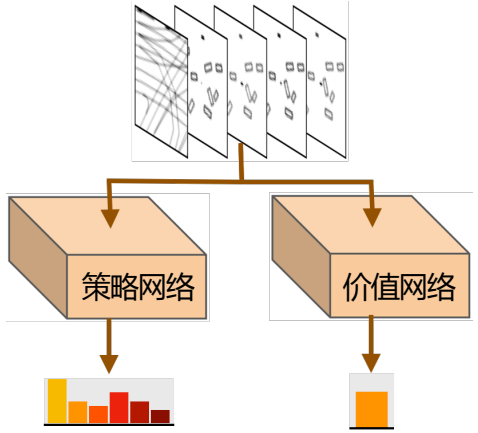
规划策略

专家策略

学习子模块

学习子模块

学习策略



提供搜索启发


强化学习
动作
奖励
规划策略
探索策略
学习子模块
学习子模块
学习策略
策略网络
价值网络
提供搜索启发

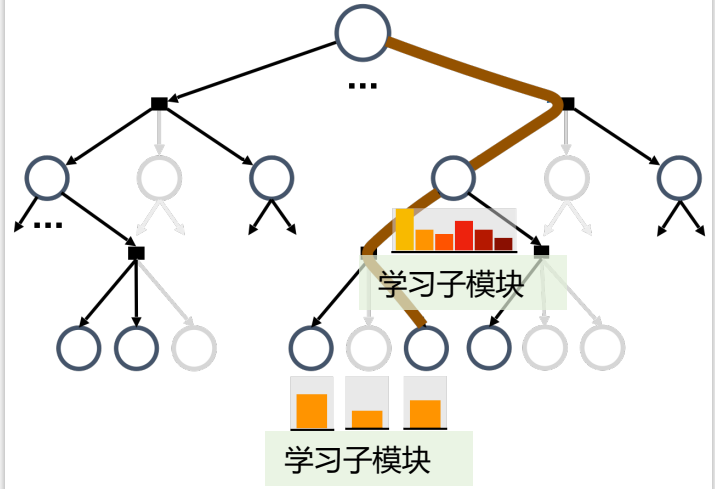
LeTS-Drive: Driving in a Crowd by Learning from Tree Search

高效训练

训练时间**减少约90%**

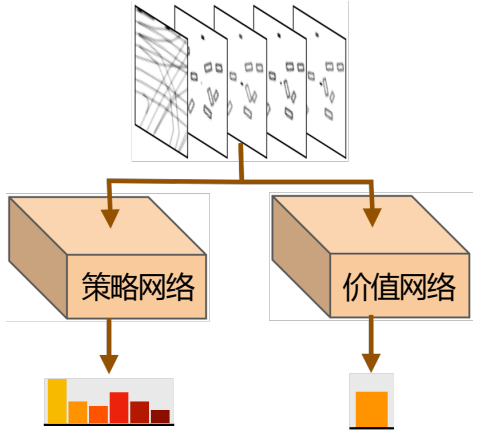
提供自监督学习/强化学习数据

规划策略



闭环
共同收敛

学习策略



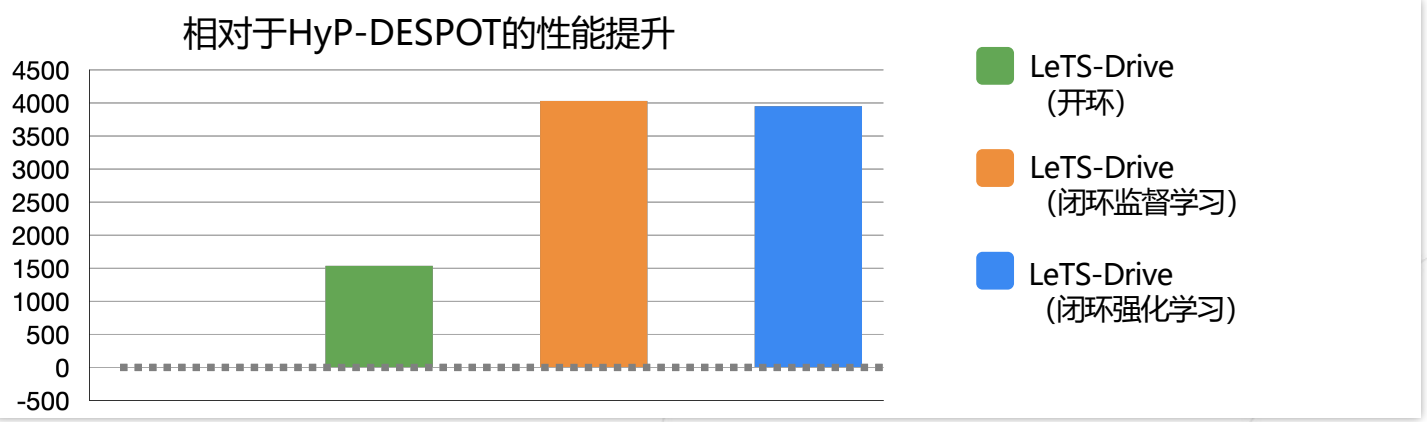
提供搜索启发

解决指数复杂度

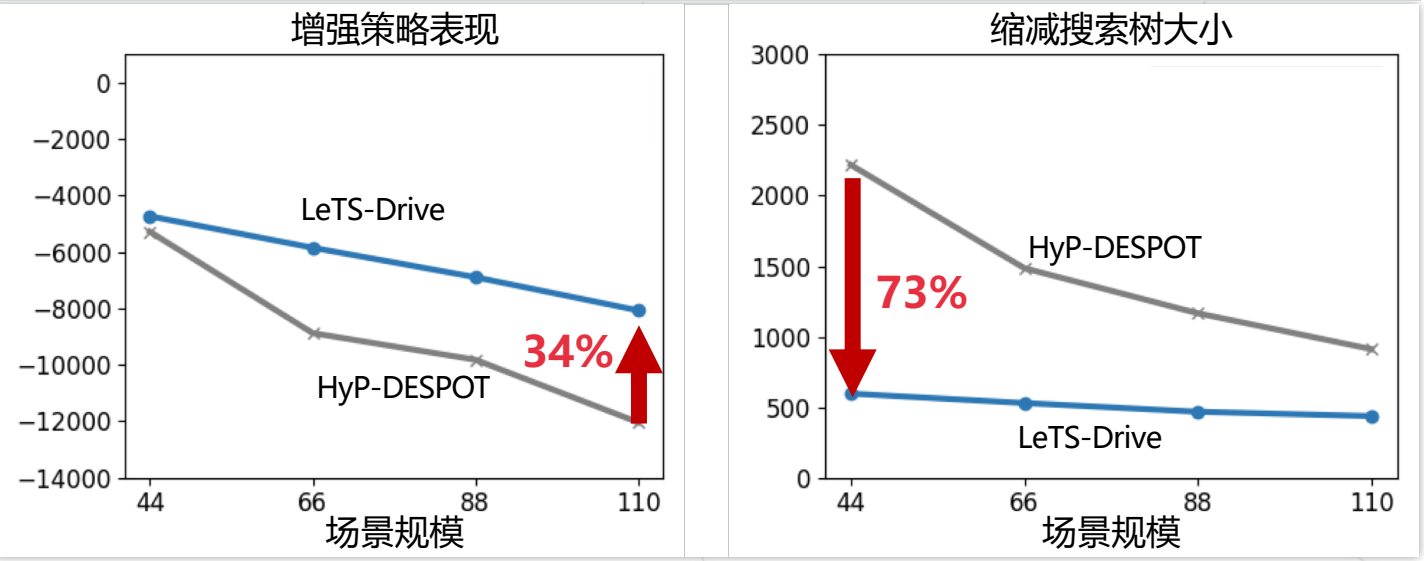
$$O(|A|^H|Z|^H) \rightarrow O(|A|^D|Z|^D)$$
$$D \ll H$$

LeTS-Drive: Driving in a Crowd by Learning from Tree Search

显著提升实时决策质量



提升大规模决策规划的可扩展性



人群中的自动驾驶

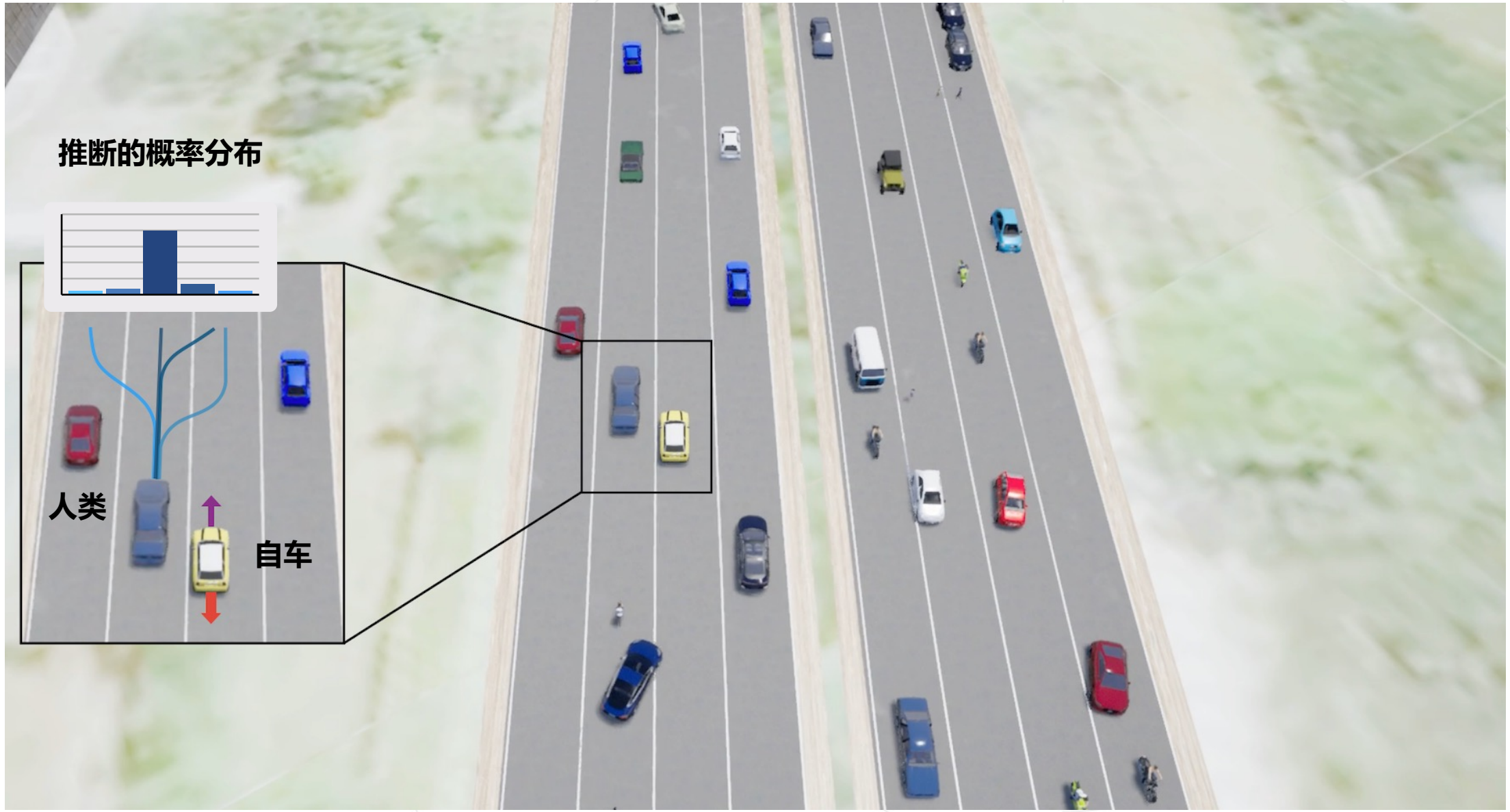
无交通规则路口自动驾驶

更进一步

在同等深度的搜索中，如何进一步提升实时决策能力？

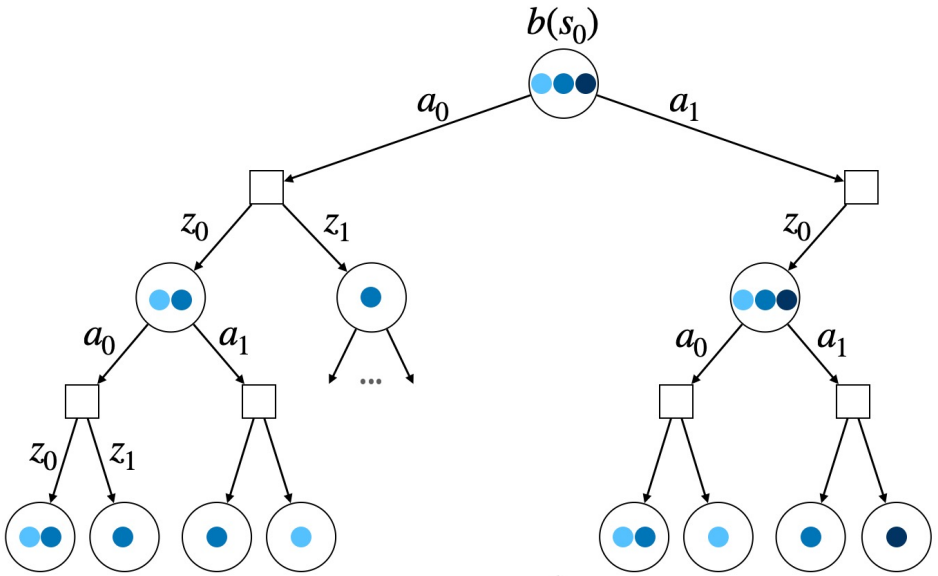
将搜索变得更窄！

如何集中搜索方向？



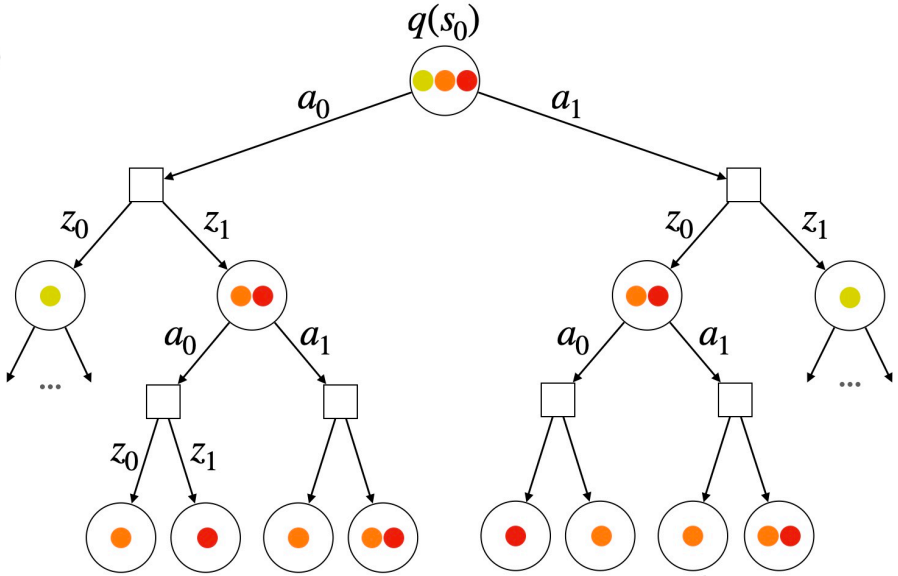
LEADER: Learning Attention over Driving Behaviors

- 学习对他人潜在行为的注意力 [CoRL'22 最佳论文提名]



无注意力机制

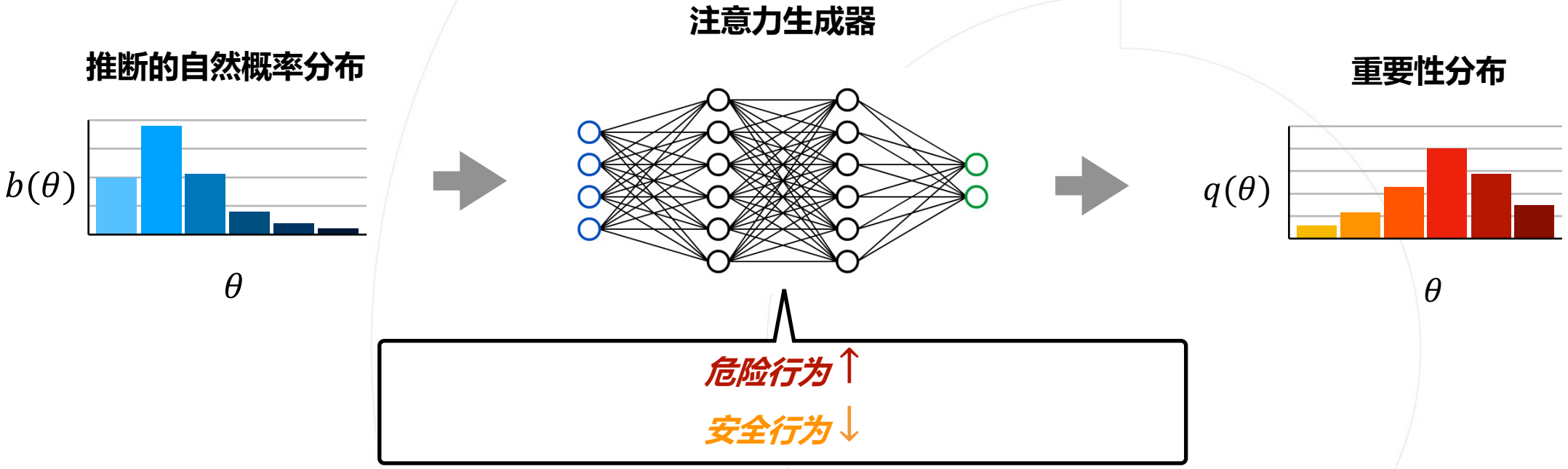
按照自然发生概率进行采样



有注意力机制

有偏向性地采样有风险的行为

LEADER: 从数据中学习行为注意力 [CoRL'22 最佳论文提名]



LEADER: 规划模块与学习模块的最大-最小博弈

[CoRL'22 最佳论文提名]

对风险敏感的规划

$$\max_{\pi \in \Pi} V_\pi(b|q)$$

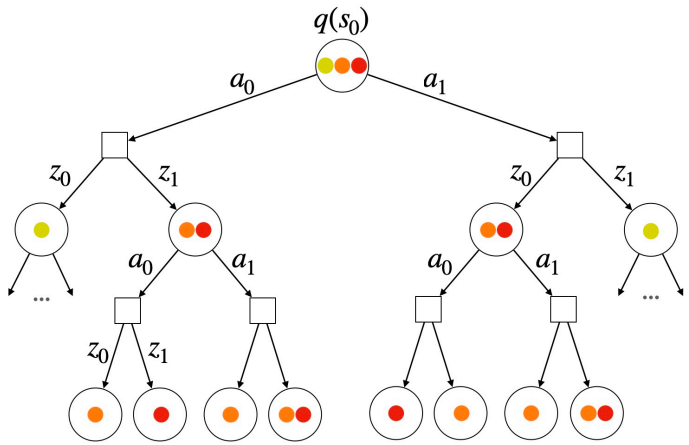
共同指标:
决策性能

$$\min_q \left[\max_{\pi \in \Pi} V_\pi(b|q) \right]$$

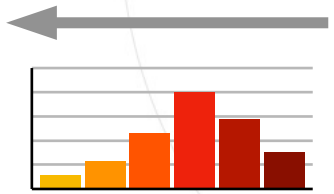
突出高风险行为

$$\min_q \left[\max_{\pi \in \Pi} V_\pi(b|q) \right]$$

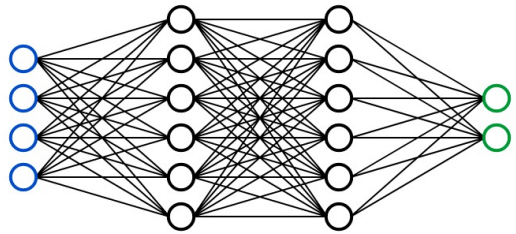
LEADER 规划器 (max)



行为注意力



注意力生成器 (min)



总结：自动驾驶决策规划

Step 1：分析问题结构

世界是一个巨大的 POMDP！

决策规划、强化学习

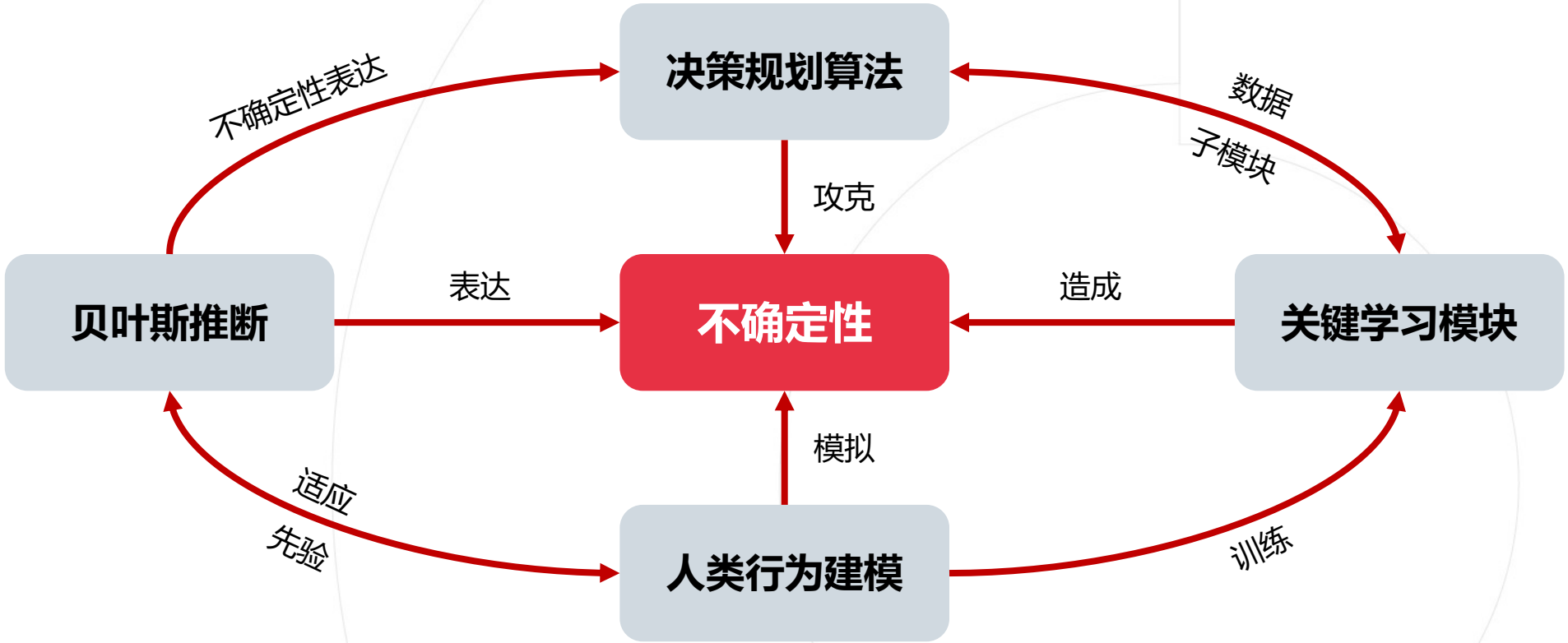
Step 2：设计规划算法

信念树搜索 + 蒙特卡洛采样 + 启发式搜索

Step 3：实用算法优化

并行化 + 融合规划与学习

感谢！


决策规划算法
不确定性表达
数据
子模块
攻克
贝叶斯推断
表达
不确定性
造成
关键学习模块
模拟
适应
先验
人类行为建模
训练

英文主页

中文主页