深蓝学院
shenlanxueyuan.com

Magic Lab

不确定性感知的决策过程

丁文超

大纲

- 决策过程的概念
- Alpha Go & Alpha Zero中的决策
- 自动驾驶的决策：Safe RL
- MPDM：简化的决策模型

- POMDP基础
- EPSILON高效分支
- MARC多模态决策规划

L5 决策过程

L6 不确定性感知的决策过程

MDP马尔科夫决策过程

马尔可夫决策过程（MDP）回顾

- 定义：

  S：状态集

  A：动作集

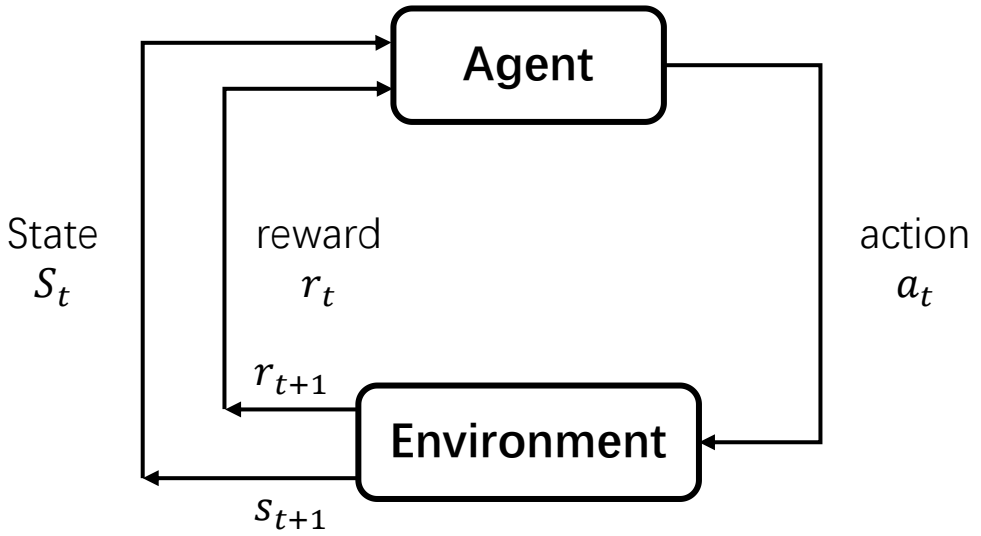
  H：智能体行动空间

  T：$S \times A \times S \times \{0,1,\ldots,H\} \rightarrow [0,1]$,

  $T_t(s,a,s') = P(s_{t+1} = s' | s_t = s, a_t = a)$

  R：$S \times A \times S \times \{0,1,\ldots,H\} \rightarrow \mathbb{R}$,

  $R_t(s,a,s') =$ reward for $(s_{t+1} = s', s_t = s, a_t = a)$



- 目标：

  找到策略π：$S \times \{0,1,\ldots,H\} \rightarrow A$，最大化奖励之和的期望，即

$$\pi^* = \arg\max_{\pi} E\left[\sum_{t=0}^{H} R_t(S_t, A_t, S_{t+1}) | \pi\right]$$

POMDP - Partially Observable MDP

定义

- 定义：

  S：状态集

  A：动作集

  T：一个集合，它指定给定前一个状态和动作的下一个状态的条件概率

  R：$S \times A \to \mathbb{R}$，奖励函数

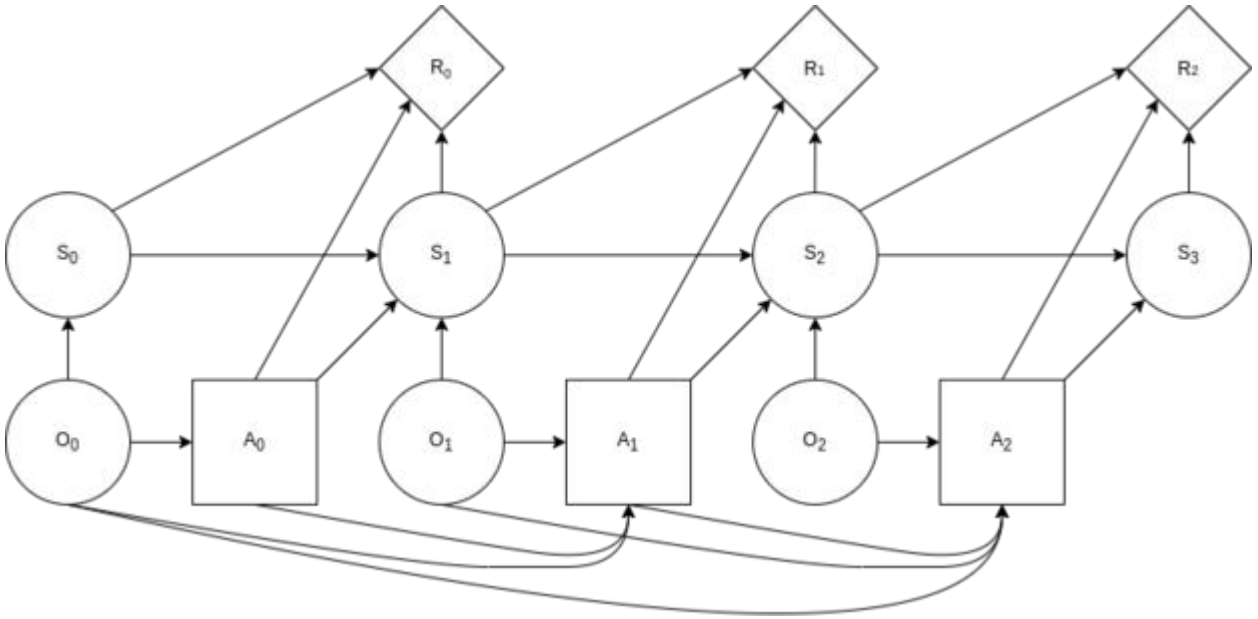
  Ω：观测值的集合

  O：观测概率的集合

  γ：$\gamma \in [0,1]$，折扣因子

- 目标：

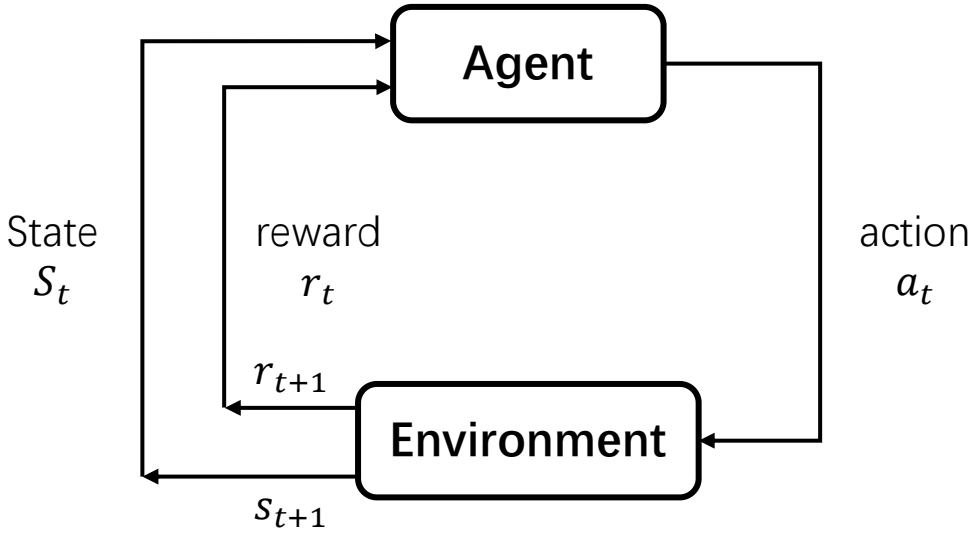
  最大化奖励之和的期望，即 $E\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$

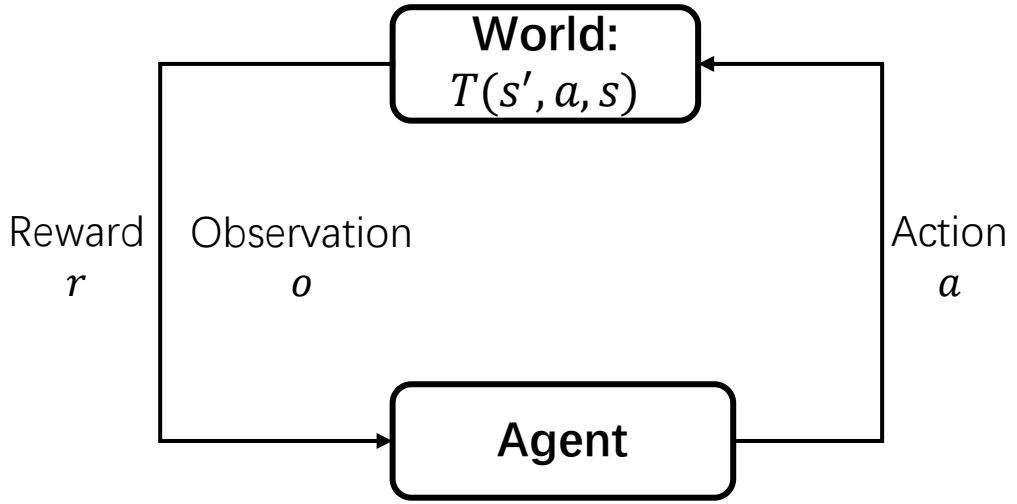


POMDP - Partially Observable MDP

和MDP的联系及区别

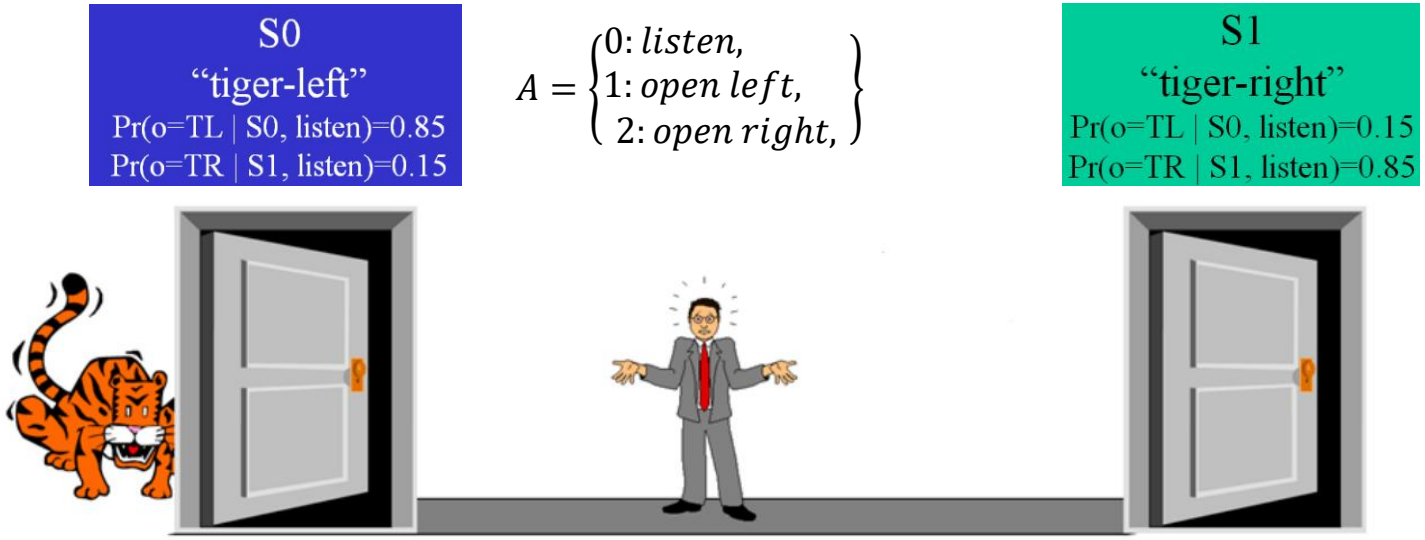
- POMDP 状态无法直接观测到，而是来源于传感器数据
- MDP是一种特殊的POMDP（观测=状态）
- 问题变为：给定当前概率分布，而不是当前状态，应该采取什么行为。
- 需要：
  1. 一种估计状态的滤波；
  2. 根据状态分布的策略。





POMDP

例子 – 从哪扇门逃出房间?



奖励函数:

开错门: -100

开对门: $+10$

listen行为的代价: -1

观测:

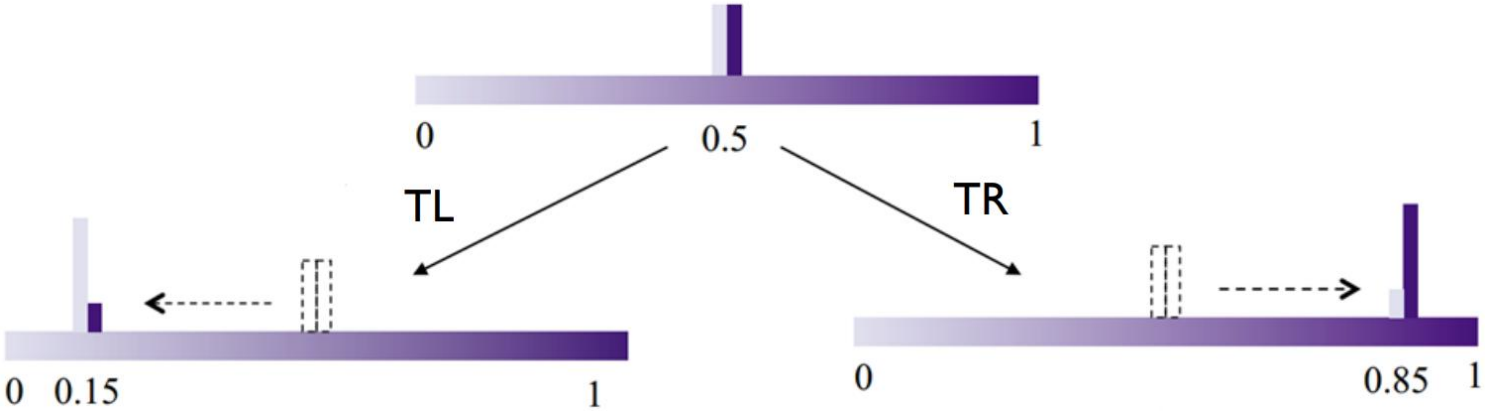
听到老虎在左侧 (TL)

听到老虎在右侧 (TR)

POMDP

置信状态

- $S0, S1$作为真实状态的可能性，其初始置信值$P(S1) = P(S0) = 0.5$;
- 一旦采取了listen（观测），置信状态相应进行更新;
- POMDP的流程：行动→分叉→观测→分叉→置信更新。



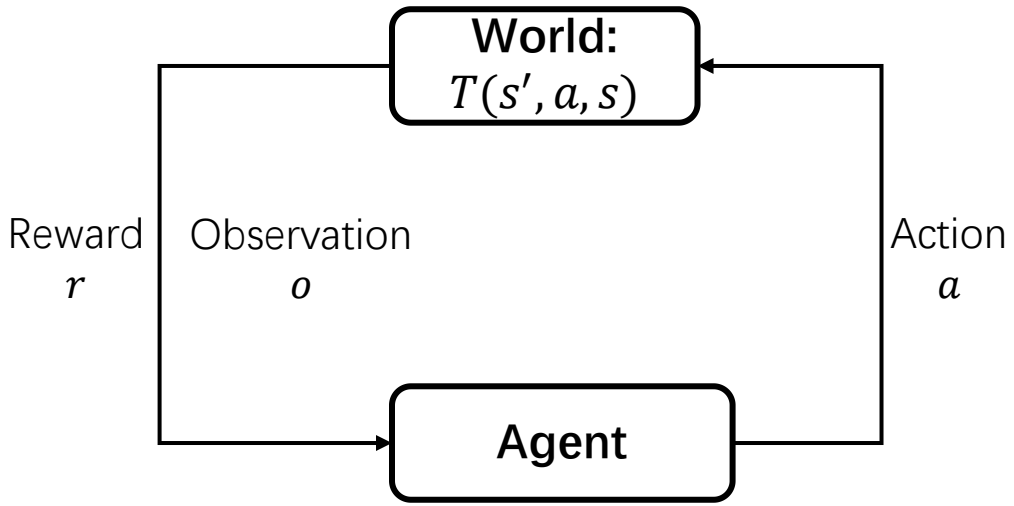
POMDP求解

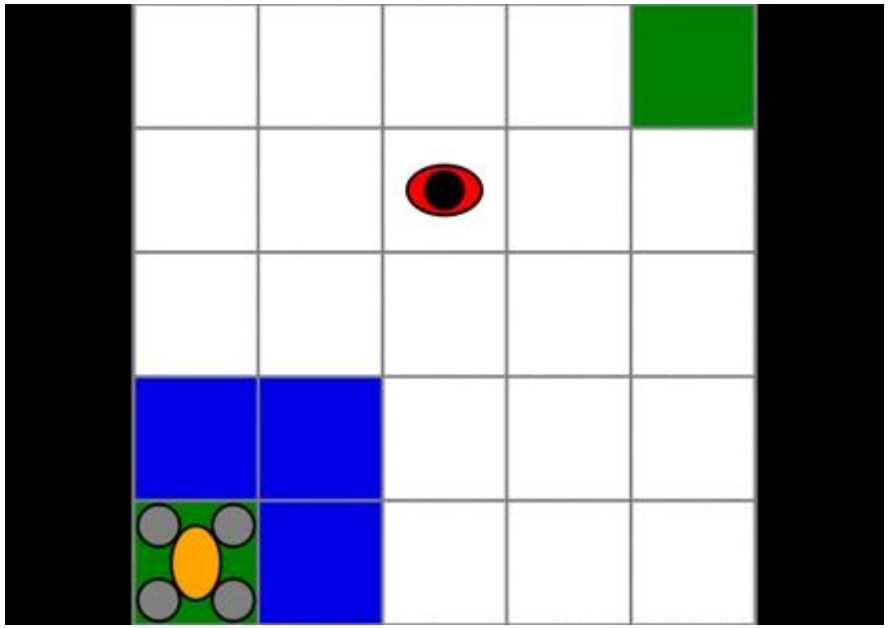
置信更新

- 由先验置信b，最后一个动作a，最后o描述的观察决定：$b' = \tau(b, a, o)$
- 新状态为s'时，智能体以$O(o|s', a)$的概率观测到$o \in \Omega$
- $b(s)$表示环境处于状态s的概率，已知$b(s)$，动作a和观测o，有：

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

其中$\eta = \frac{1}{\Pr(o|b,a)}$，$\Pr(o|b, a) = \sum_{s' \in S} O(o|s', a) \sum_{s \in S} T(s'|s, a) b(s)$





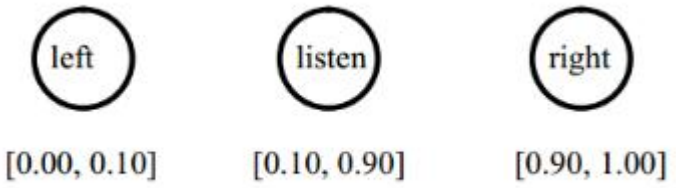
https://wiki.ubc.ca/Course:CPSC522/Partially_Observable_Markov_Decision_Processes#/

POMDP

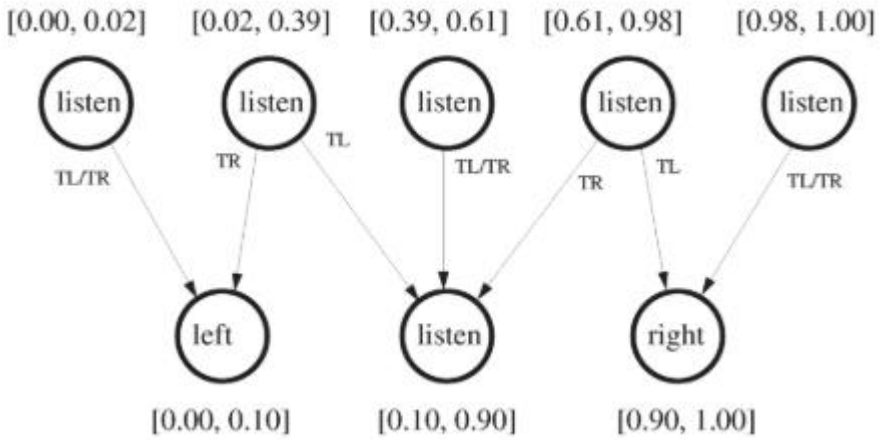
策略

- 策略π是从[0,1] ->{listen, open-left, open-right}的映射
- 应该怎么制定策略?
  - ✓ 大致上一直listen直到确定，然后开门
- 改变策略的临界？90%确定时选择开门

只有一次动作机会



两次动作机会：先listen



POMDP求解

常规方法1：连续状态的“置信MDP”

- 计算价值迭代，但状态空间为概率分布空间

  →根据已有的状态分布，采取动作，获得所有可能观测的概率，每个观测的分支计算后验概率

  →相当于构建了从当前状态到未来可能状态集合的动力学模型

  →为所有可能的状态确定价值和最优动作（MDP价值迭代）

  →由于收集更多的信息会带来更多的长期奖励，上述过程将自动权衡信息收集与改变状态的行动代价

- 无法进行精确的价值迭代，因为即便上述只有两种状态，有无数种置信（上述例子中，置信状态为[0,1]中的任意值）

  →需要近似方法

POMDP求解

问题求解

- 已知初始置信b_0，策略π的期望奖励为

$$V^{\pi}(b_0) = \sum_{t=0}^{\infty} \gamma^t r(b_t, a_t) = \sum_{t=0}^{\infty} \gamma^t E[R(s_t, a_t)|b_0, \pi]$$

- 通过优化奖励得到最优策略$\pi^* = \underset{\pi}{\operatorname{argmax}} V^{\pi}(b_0)$，则

$$V^*(b) = \max_{a \in A} \left[r(b,a) + \gamma \sum_{o \in \Omega} Pr(o|b,a) V^*(\tau(b,a,o)) \right]$$

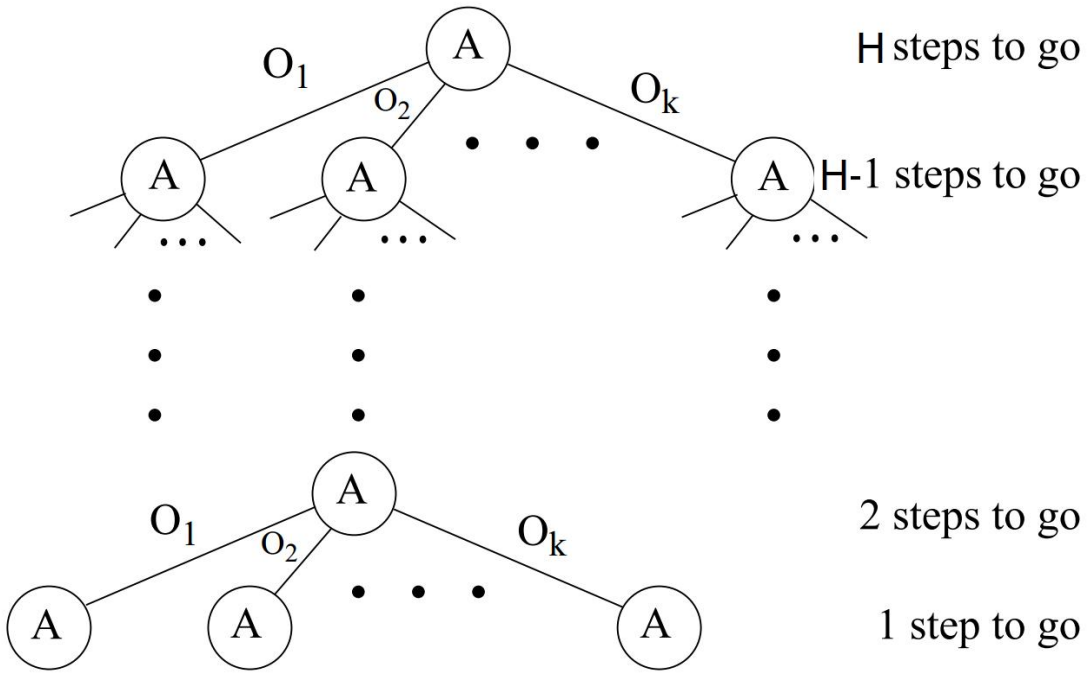
价值迭代：给每个置信状态赋值价值？

维度爆炸！！！

POMDP求解

常规方法2

- 以有限的前瞻搜索动作序列（类似MPC）
- 对动作和观测进行剪枝，采样更大概率出现的状态



有限区间：节点数$|A|\frac{|O|^H-1}{|O|-1}$

POMDP求解

常规方法2

- 一些专门的近似求解方法：当$|S|$在百万级以内可求解
  - ✓ 基于点的α-向量技巧[1]，价值函数会以信仰空间内的超平面形式出现
  - ✓ 蒙特卡洛树搜索[2]（从当前状态搜索状态/观测序列）
  - ✓ ……

[1] Shani G, Pineau J, Kaplow R. A survey of point-based POMDP solvers[J]. Autonomous Agents and Multi-Agent Systems, 2013, 27: 1-51.
[2] Silver D, Veness J. Monte-Carlo planning in large POMDPs[J]. Advances in neural information processing systems, 2010, 23.

POMDP求解

常规方法3 在MDP中规划

- 假设概率最大的置信状态即实际状态
- 通过概率推断（滤波）跟踪概率分布
- 根据当前最可能的状态选择最优的行动

尽管这在计算上比较高效，但无法主动地收集信息，当具备较高不确定性时容易失败。

大纲

- POMDP基础
- **EPSILON高效分支**
- MARC多模态决策规划

EPSILON 解读

- An Efficient Planning System for Automated Vehicles in Highly Interactive Environments

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

香港科技大學-
大疆創新科技聯合實驗室
HKUST-DJI JOINT INNOVATION LABORATORY

Magic Lab

EPSILON: An Efficient Planning System for Automated Vehicles in Highly Interactive Environments

Supplementary Materials

Wenchao Ding*, Lu Zhang*, Jing Chen, Shaojie Shen

** W. Ding and L. Zhang contribute equally to this paper.*

⋆ https://github.com/HKUST-Aerial-Robotics/EPSILON.git ⋆

EPSILON 解读

自动驾驶决策规划现存问题

- 没有考虑不确定性和多模态，应该从概率的角度对问题建模
  - 车端处理的数据具有高度不确定性且不确定性来源广泛（感知、检测、遮挡等噪声）
- 没有考虑多智能体之间的交互
  - 其他交通参与者的行为具有随机性和交互性
- 现有方法大多通过模拟仿真或数据集进行验证
  - 难以模拟真实数据的不确定性和交通参与者行为的随机性
  - 难以从仿真场景迁移到真实环境

EPSILON 解读

该系统主要贡献

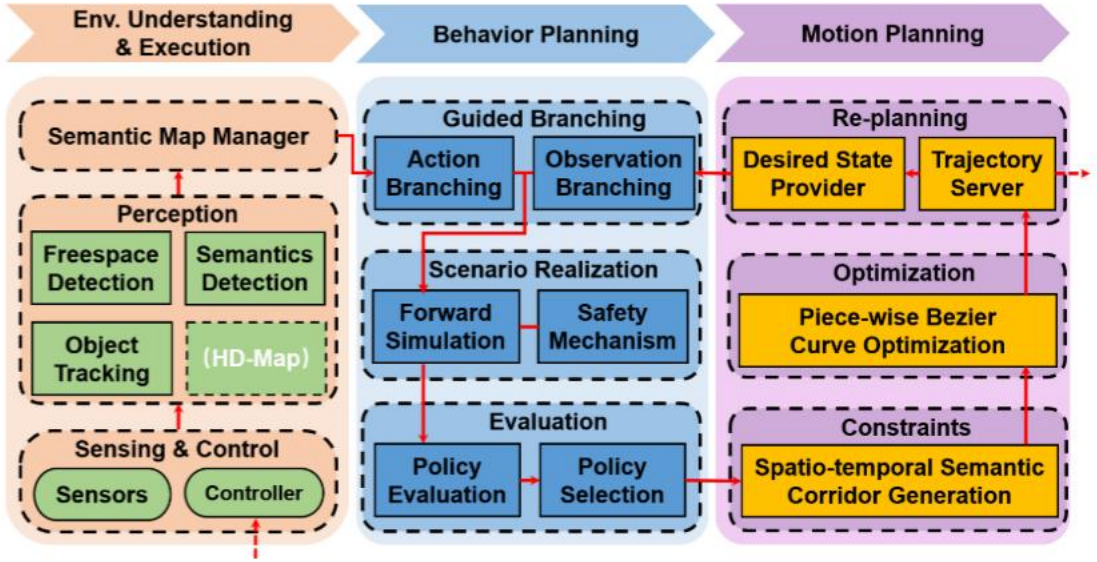
- 提出了一个有效的规划系统，可以解决包括高效交互、不确定性处理、执行效率等问题
- 引入一个新的前向仿真模型——具有安全机制，以克服由于先验知识的潜在不完全性所带来的风险，增强了行为规划层的鲁棒性，即便其他交通参与者行为不合理，也能保证良好的安全性
- 不仅在仿真器和数据集上进行测试，而且在没有HD map的情况下仅使用车端传感器进行验证
- 代码已经开源，欢迎使用⭐ https://github.com/HKUST-Aerial-Robotics/EPSILON.git ⭐

EPSILON 解读

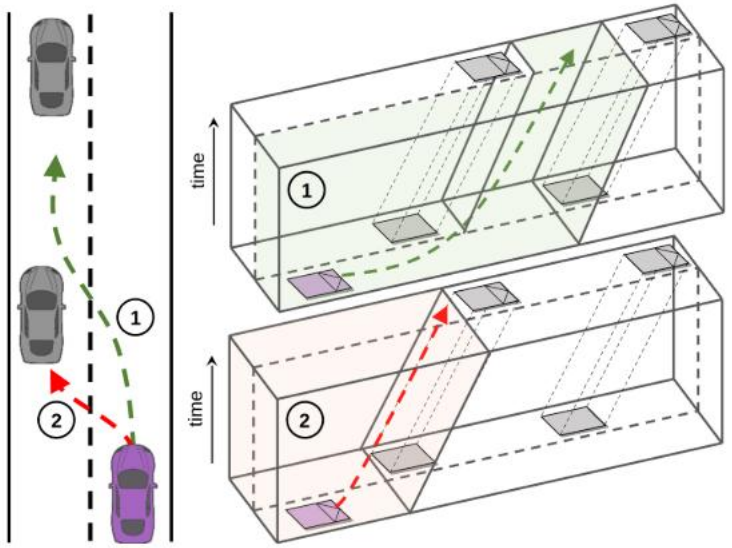
系统概述

EPSILON由行为规划层和运动规划层的分层结构组成：

- 行为规划层采用状态序列表示并覆盖运动规划范围的初步决策；
- 运动规划层利用状态序列生成可用于闭环执行的安全平滑轨迹。



EPSILON模块图



行为规划和运动规划之间的关系说明图

EPSILON 解读

问题描述

符号说明

- E_t: 时刻 t 时，以自车为中心的局部环境，包括道路结构、交通信号和静态障碍物的占用栅格。
- x_t^i: 车辆 i 在时刻 t 时的状态，当 $i = 0$ 时，表示自车状态。为了符号方便，下标缺失表示所有时间步，上标缺失表示所有车辆。
- Z_t: 时刻 t 时，自车接受到的观测量。
- $D_t := [x_{t+1}, x_{t+2}, ..., x_{t+H}]$由一系列离散状态所定义的决策。

行为规划层和运动规划层解耦，行为规划器只需要以相对粗略的分辨率推理未来场景，而运动规划层在给定决策的局部解空间中工作。

- 行为规划层：输入 E_t 和 Z_t，输出 D_t;
- 运动规划层：输入 D_t 和 Z_t，输出用于跟踪控制的轨迹。

EPSILON 解读

问题描述

问题：在现实世界中，规划系统总是受到交通参与者之间未知的交互模式的影响。

解决思路：以 *POMDP* 的形式对不确定性行为规划问题进行建模。

*POMDP*可以定义为元组$\langle X, A, Z, T, O, R\rangle$。状态转移概率模型$T$和观测模型$O$，这两个函数反映了运动模型的随机性和不确定性。

- 状态转移概率模型 $T(x_{t-1}, a_t, x_t) = p(x_t|x_{t-1}, a_t)$
- 观测模型 $O(x_t, z_t) = p(z_t|x_t)$

由于现实世界中的一些状态不能被直接观察到(例如，隐藏的意图或噪声测量)，*POMDP*保持置信度b，这是状态空间X上的概率分布。

$$b_t(x_t) = p(x_t \mid z_t, a_t, b_{t-1}) = \eta O(x_t, z_t) \int_{x_{t-1} \in \mathrm{X}} T(x_{t-1}, a_t, x_t) b_{t-1}(x_{t-1}) dx_{t-1}$$

由上一节*POMDP*的知识可知：

$$\pi^* := \arg\max_\pi \mathrm{E}\left[\sum_{t=t_0}^{t_H} \gamma^{t-t_0} R(x_t, \pi(b_t)) \,\middle|\, b_{t_0}\right]$$

$$V^*(b) = \max_{a \in \mathrm{A}} Q^*(b,a) = \max_{a \in \mathrm{A}} \{\int_{x \in \mathrm{X}} b(x) R(x,a) dx + \gamma \int_{z \in \mathrm{Z}} p(z \mid b, a) V^*\left(\tau(b,a,z)\right) dz\}$$

$$a_t^* = \arg\max_{a_t \in \mathrm{A}} Q^*(b_{t-1}, a_t)$$

$$z_t^* = \arg\max_{z_t \in \mathrm{Z}} p(z_t | \ b_{t-1}, a_t^*)$$

EPSILON 解读

问题描述

将驾驶场景转为POMDP形式：

$$
\begin{aligned}
& p(x_t|x_{t-1}, a_t) \\
\approx & \underbrace{p(x_t^0|x_{t-1}^0, a_t^0)}_{\text{ego transition}} \prod_{i=1}^{N} \int_{\mathcal{A}^i} \underbrace{p(x_t^i|x_{t-1}^i, a_t^i)}_{i\text{-th agent's transition}} \underbrace{p(a_t^i|x_{t-1})}_{\text{driver model}} da_t^i,
\end{aligned}
$$

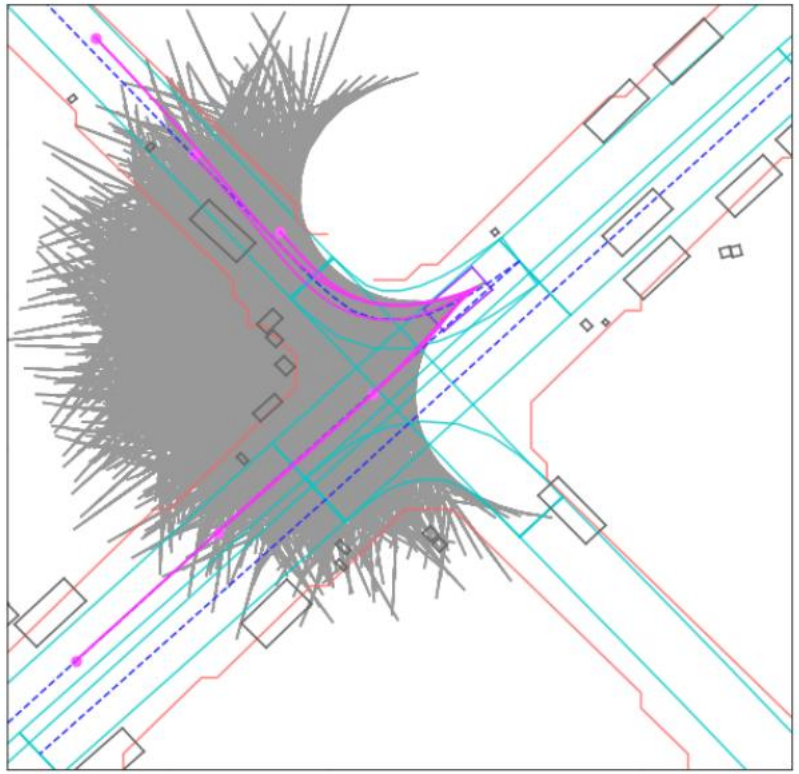
$$
b_t(x_t) = \eta \cdot \overbrace{p(z_t^0|x_t^i) \int_{\mathcal{X}^0} p(x_t^0|x_{t-1}^0, a_t^0) b_{t-1}^0(x_{t-1}^0) dx_{t-1}^0}^{\text{belief update for ego agent}}
$$

$$
\overbrace{\prod_{i=1}^{N} p(z_t^i|x_t^i) \iint_{\mathcal{X}^i \mathcal{A}^i} p(x_t^i|x_{t-1}^i, a_t^i) p(a_t^i|x_{t-1}) b_{t-1}^i(x_{t-1}^i) da_t^i dx_{t-1}^i}^{\text{belief update for other agents}} .
$$

EPSILON 解读

Efficient Behavior Planning -- Motivating Example

- 通过动作空间引导分支，减少原POMDP的决策空间；
- 在策略评估过程中，选取基于自我策略的潜在风险场景，减少计算消耗；
- 提供一种新的具有安全机制的正向仿真模型的实现，可以减少真实交通参与者与假设模型之间的不一致所带来的风险。



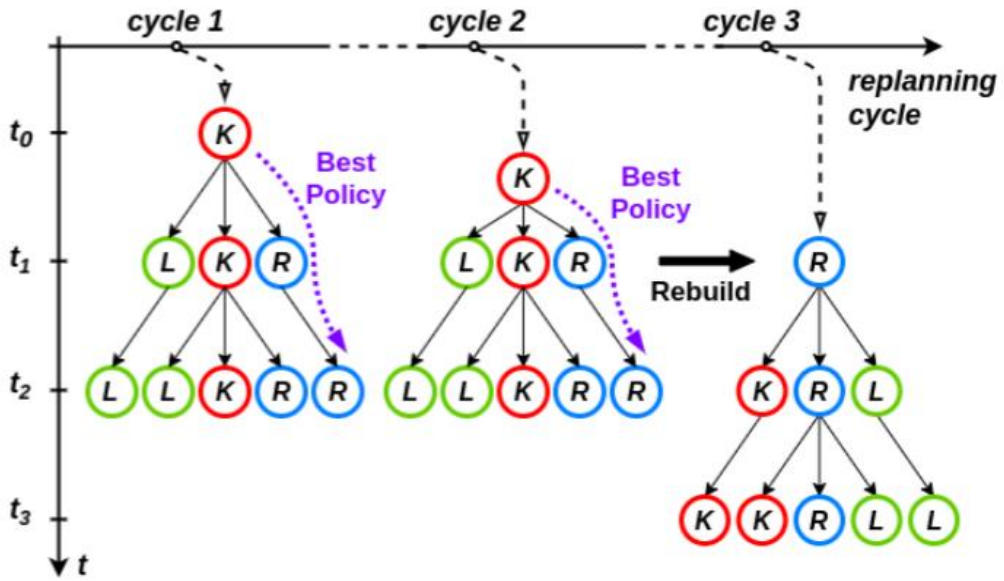
使用语义级动作指导动作空间探索示意图

人类司机倾向于只考虑一些长期的语义层面的行为(例如车道保持，让行，超车等)，通过利用语义级动作，我们可以直接对动作分支的长期高似然轨迹进行采样。

EPSILON 解读

Efficient Behavior Planning -- Guided Action Branching

- 引入可以闭环执行的语义动作（变道，变速等）替代原始动作（转向，油门等）；
- 利用域特定闭环策略树(DCP-Tree)简化决策过程，扩展时域中的决策点；
- 由于每个策略的评估是相互独立的，可以通过并行方式实现规划器。



使用语义级动作指导动作空间探索示意图

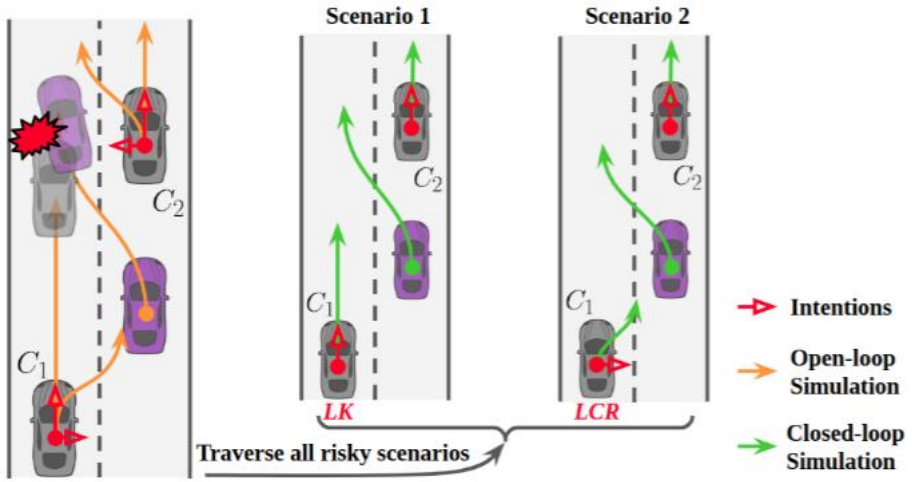
DCP-Tree的节点是预定义的语义动作，树的有向边表示时间上的执行顺序。作为人类驾驶员，我们通常不会在单个决策周期内来回更改驾驶策略。每个策略序列在一个决策周期内最多包含一个行动的变更，而来回的行为是通过重新规划来实现的。

通过遍历DCP-Tree上从根节点到所有叶节点的所有路径来获得候选策略。行为规划被简化为从候选策略集中选择具有最大效用的最佳策略。

EPSILON 解读

Efficient Behavior Planning -- Guided Observation Branching

- 采用黑盒模拟范式，简化置信度更新过程；
- 假设自车状态完全可观察，且观察轨迹无噪声，利用联合概率映射出语义动作，最终转化为置信状态；
- 采用正向模拟，假定其他agent语义动作固定，将预测结果作为初始置信值。



CFB机制示意图

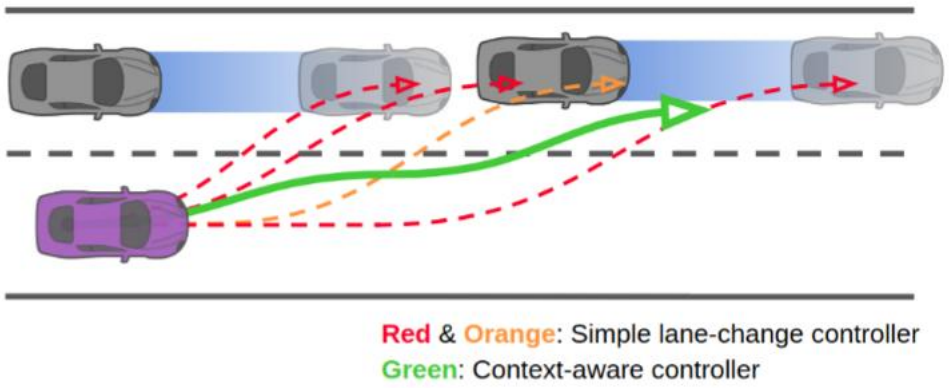
场景的概率通过参与者估计概率的联合分布计算代价太高，可以通过条件聚焦分支(conditional focused branching CFB)用尽可能少的分支发现附近车辆的意图可能导致的风险结果。

其动机来自于人类驾驶员在进行不同机动时对附近车辆的注意力有不同的偏向性。

EPSILON 解读

Efficient Behavior Planning -- Multi-agent Forward Simulation

- 将预测耦合到行为规划中，每个潜在的规划所考虑的场景有它自己相应的未来预期;
- 使用上下文感知控制器来控制自车，使用基于模型的控制器控制其他交通参与者;
- 通过引入新的控制器，扩大求解空间，得到的决策能够更灵活地适应交通配置。



上下文感知控制器示意图

通过上下文感知控制器，所需动作序列的数量大大减少，前向仿真的成功率大大提高，进一步提高了分支的性能和效率。

EPSILON 解读

Efficient Behavior Planning -- Safety Mechanism

- 在环境感知控制器中嵌入安全模块，在一定程度上自动保证控制输出安全或无故障，提高了前向仿真模型的安全性；
- 在责任与安全方面，利用责任敏感安全数学模型(responsibility-sensitive safety, RSS)， RSS可以在各种驾驶情况下快速响应；
- 在策略选择中设置安全准则，增强决策层的鲁棒性。

RSS 根据人类驾驶的常识建立了五个安全原则，适用不同的驾驶场景：

（1）避免碰撞前面的车

（2）保持横向安全距离

（3）合理使用路权

（4）注意视觉盲区

（5）避免事故发生是首要任务

S. Shalev-Shwartz, S. Shammah, and A. Shashua, “On a formal model of safe and scalable self-driving cars,” arXiv preprint arXiv:1708.06374,2017.

EPSILON 解读

Efficient Behavior Planning -- Policy Selection

Algorithm 1: Process of behavior planning layer

Inputs: Current states of ego and other vehicles x; *Ongoing action* $\hat{\phi}$; Pre-defined semantic action set Φ; Planning horizon H;
$\mathfrak{R} \leftarrow \emptyset$; // set of rewards for each policy;
$\Psi \leftarrow \texttt{UpdateDCPTree}(\Phi, \hat{\phi})$; // DCP-Tree Ψ;
$\hat{\Pi} \leftarrow \texttt{ExtractPolicySequences}(\Psi)$;
foreach $\pi \in \hat{\Pi}$ **do**
  $\Gamma^{\pi} \leftarrow \emptyset$; // set of simulated trajectories;
  $\Omega \leftarrow \texttt{CFB}(x, \pi)$; // set of critical scenarios;
  foreach $\omega \in \Omega$ **do**
    $\Gamma^{\pi} \leftarrow \Gamma^{\pi} \cup \texttt{SimulateForward}(\omega, \pi, H)$;
  end
  $\mathfrak{R} \leftarrow \mathfrak{R} \cup \texttt{EvaluatePolicy}(\pi, \Gamma^{\pi})$;
end
$\pi^*, \hat{\phi} \leftarrow \texttt{SelectPolicy}(\mathfrak{R})$;

通过在动作空间和观测空间进行引导分支，并对状态转移和观测进行近似，行为规划可以简化为有限数量的策略评估问题。

对于每个策略，通过评估规划行为和模拟轨迹来计算每个场景的奖励加权和。

奖励功能由效率、安全、导航三个部分组成。

EPSILON 解读

基于时空语义走廊的轨迹生成

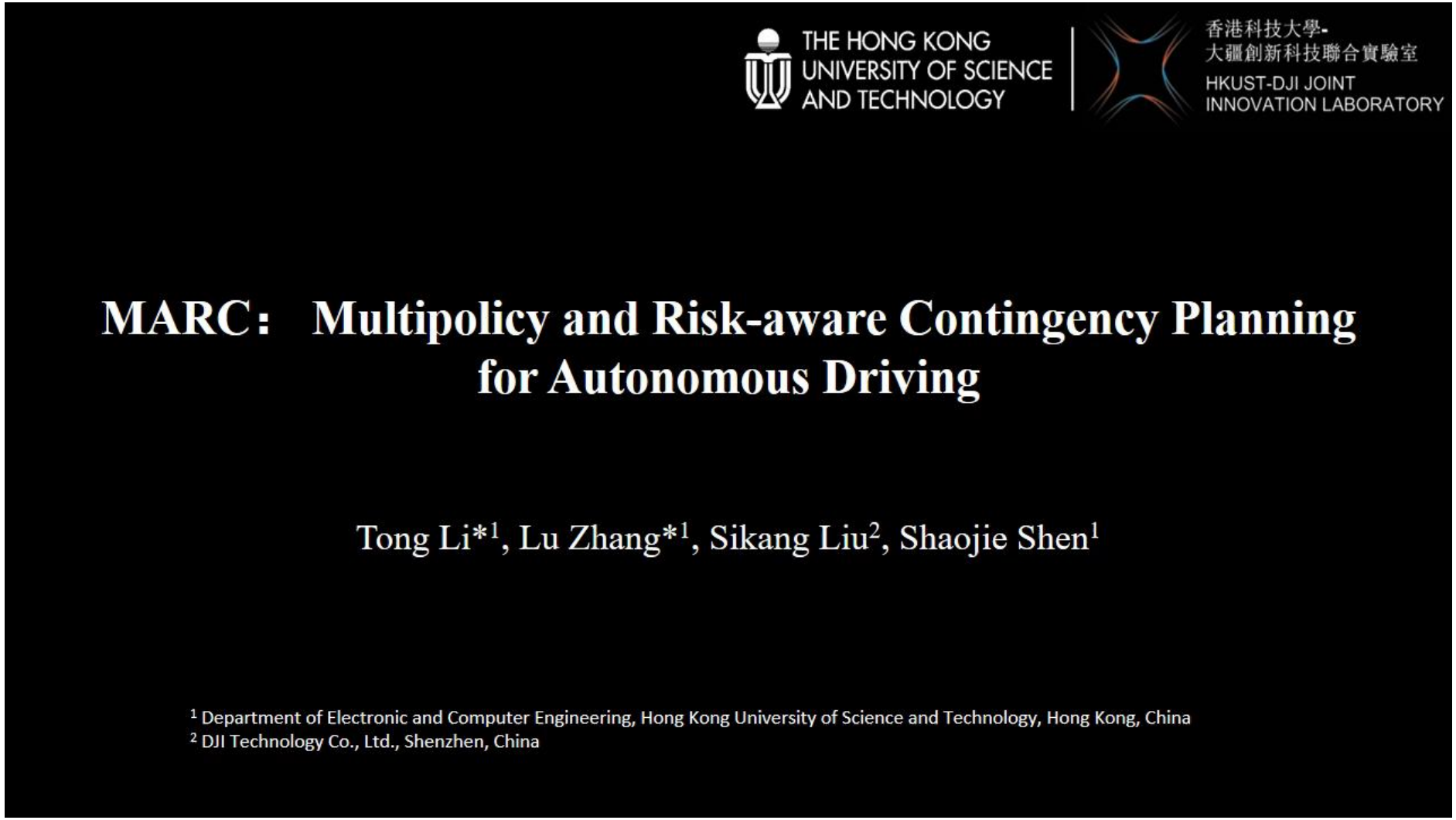
- 与第4章中的时空规划方法类似，这里不做详细阐述。

$$J_j^\sigma = w_s^\sigma \cdot \int_{t_{j-1}}^{t_j} \left(\frac{d^3 f_j^\sigma(t)}{dt^3} \right)^2 dt + w_f^\sigma \cdot \frac{1}{n_j} \sum_{k=0}^{n_j} \left(f_j^\sigma(t_k) - r_{jk}^\sigma \right)^2$$

$$J_j^\sigma = \mathbf{p}_j^{\mathrm{T}} (w_s^\sigma \mathbf{Q}_s + w_f^\sigma \mathbf{Q}_f) \mathbf{p}_j + w_f^\sigma \mathbf{c}^{\mathrm{T}} \mathbf{p}_j = \frac{1}{2} \mathbf{p}_j^{\mathrm{T}} \hat{\mathbf{Q}} \mathbf{p}_j + \hat{\mathbf{c}}^{\mathrm{T}} \mathbf{p}_j$$

MARC 解读

- Multipolicy and Risk-aware Contingency Planning for Autonomous Driving



MARC 解读

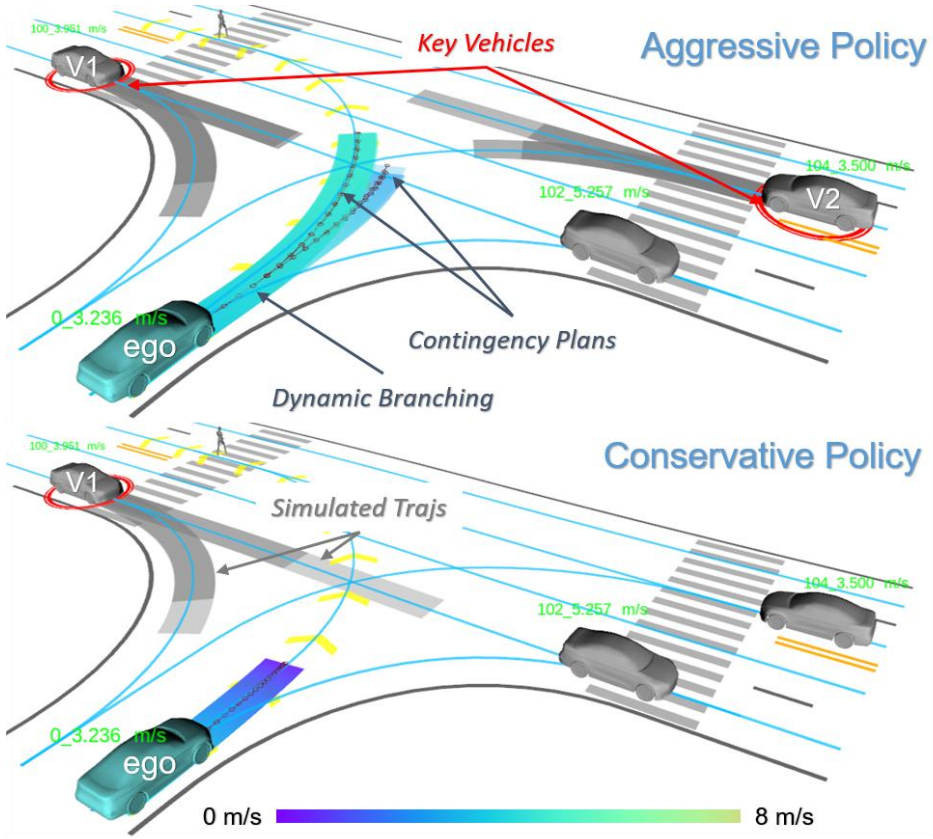
自动驾驶决策规划现存问题

- 现有的决策缺乏同时处理多智能体之间的多模态交互的能力
  - 由于其他交通参与者的行为具有随机性和交互性，只处理单一交互模态难以保证决策的一致性
- 现有的规划方法不能很好地处理多智能体的多模态预测
  - 单轨迹优化方法无法同时处理多模态预测轨迹
  - 多轨迹优化方法对多模态预测轨迹有强假设（固定的模态数量、分叉的时间等）
- 现有的决策规划没有考虑到安全风险的影响
  - 对于同一个驾驶场景，不同的驾驶员会根据不同的风险偏好做出不同的决策

MARC 解读

该系统主要贡献

- 提出了一个可以处理多智能体多模态交互的决策规划系统
- 该系统可以生成与自车决策有关的场景树，该场景树考虑了关键交互场景以及场景之间差异。
- 引入了具有风险意识的树状轨迹优化，可同时处理多模态的预测轨迹及不同的风险容忍程度。
- 在复杂的仿真环境下与EPSILON进行了对比测试验证

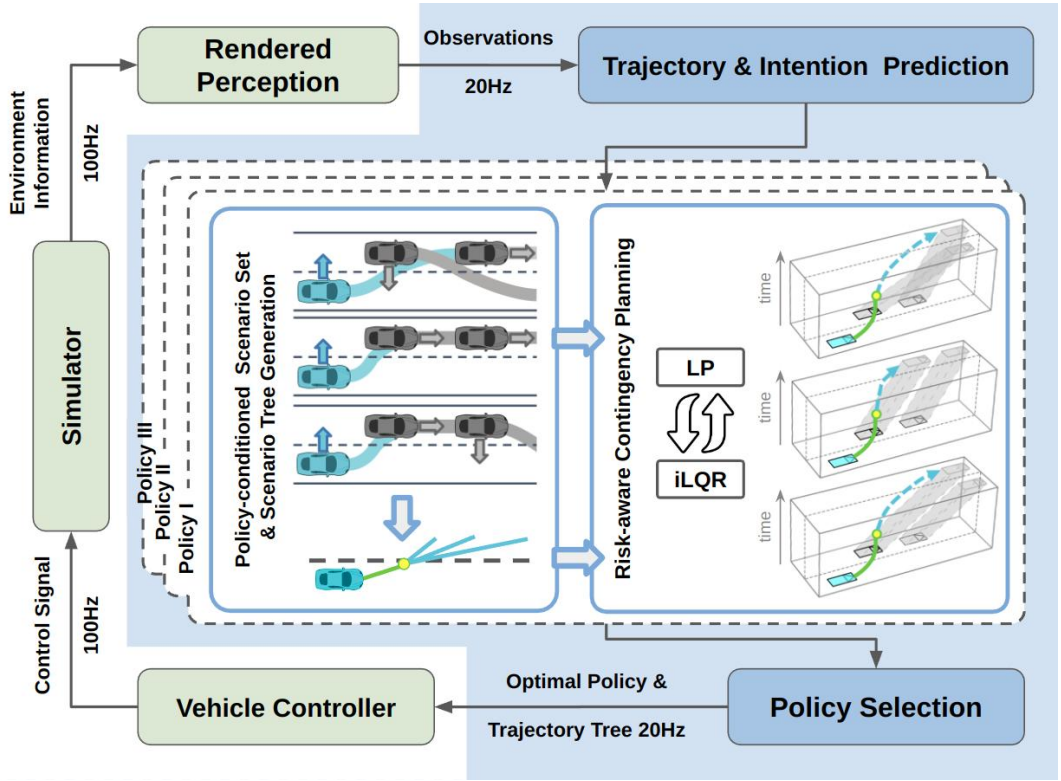


MARC 解读

系统概述

MARC核心模块由场景及场景树生成模块（行为规划层）和运动规划层的分层结构组成：

- 场景及场景树生成模块采用考虑自车决策的前向仿真生成关键场景集合并根据场景的差异生成场景树；
- 运动规划采用迭代优化的方法生成考虑风险并兼顾多模态预测的轨迹树。

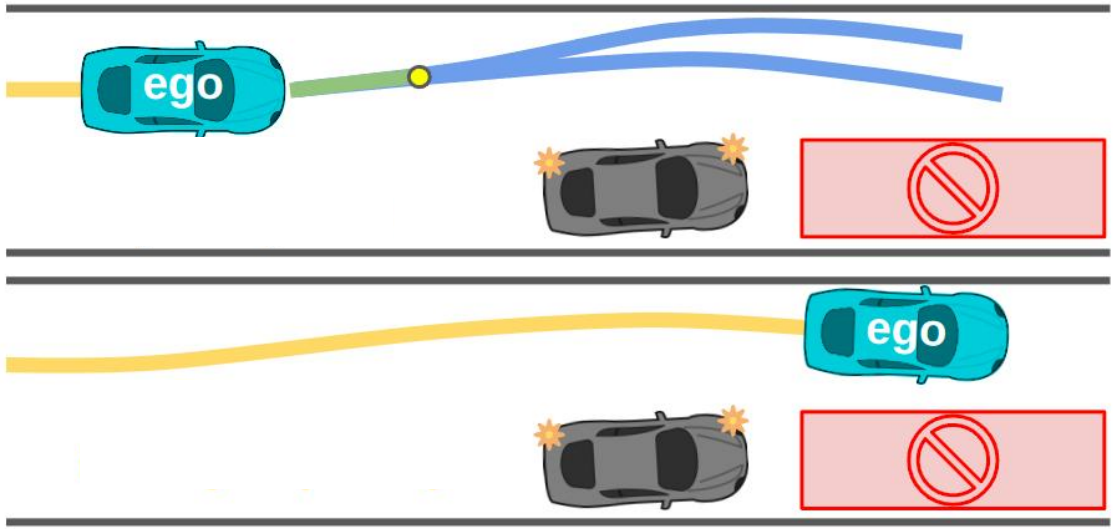


MARC 模块图（蓝色）

MARC 解读

Policy-conditioned Scenario Tree Generation -- Motivating Example

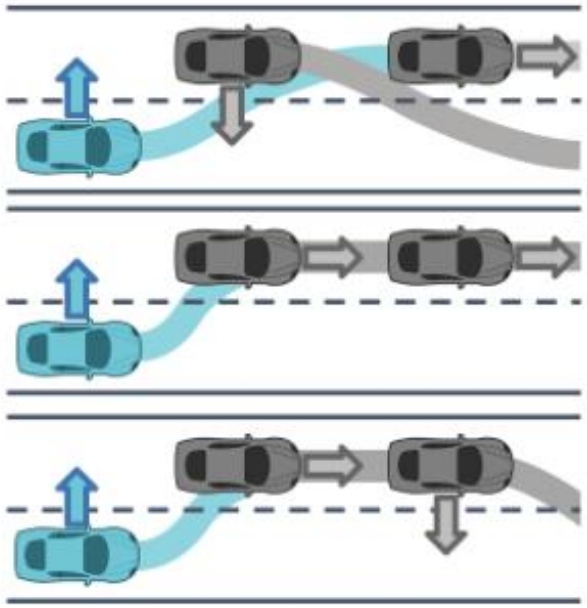
- 有经验的人类司机会倾向于执行稳定且不需要频繁切换的驾驶行为来保证乘坐舒适性;
- 防御性驾驶通常选择兼顾考虑不同潜在可能场景的行为来保证安全性。



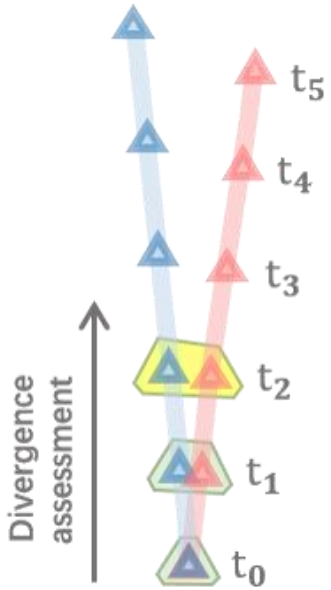
MARC 解读

Policy-conditioned Scenario Tree Generation via Closed-loop Forward Simulation

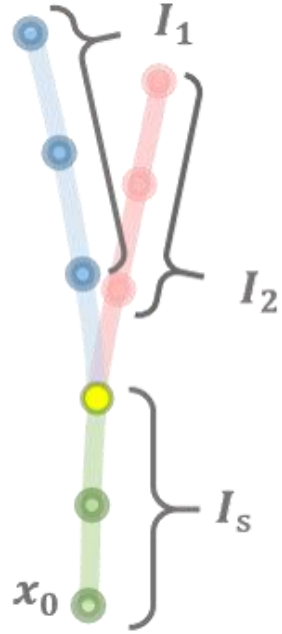
- 基于同样的自车行为决策及不同的他车意图组合，通过闭环前向仿真将意图耦合到行为规划中，生成不同的关键场景；
- 对场景间的差异进行评估，将符合差异阈值的场景节点进行合并，保留超出阈值的节点，生成场景树。



同一自车行为决策下的关键场景集



关键场景差异评估



场景树

MARC 解读

Policy-conditioned Scenario Tree Generation via Closed-loop Forward Simulation

- 关键场景差异评估

问题：不同场景中的多智能体交互的差异缺乏显示的指标表征。

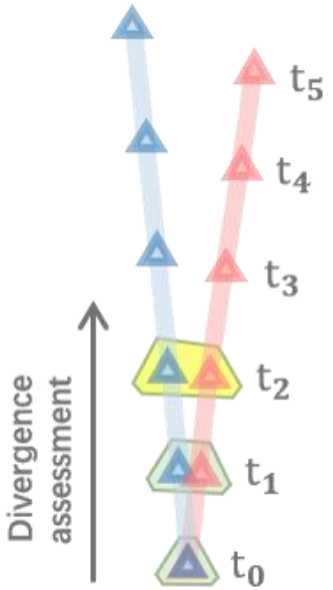
解决思路：基于*POMDP/MPDM*的前向仿真可以得到给定自车决策下对他车不同交互的粗粒度响应。仿真生成的自车行为是与多智能体交互的结果。因此可以将不同场景自车行为的差异作为一种表征场景差异的间接指标。

- 场景节点的合并/分叉时间点的选择

问题：对于存在多个可合并的场景节点，应如何选择合并的数量/分叉时间点。

解决思路：场景树决定了之后的规划层的轨迹树结构。合并数量少/分叉时间点早则意味着轨迹树会有更短的公共轨迹，更早的应急规划切换时间点，更少的响应时间，使得整体轨迹规划不够鲁棒。为了使得生成的决策规划可以稳定地切换到不同的场景，我们希望合并尽可能多的场景节点，生成尽可能晚的分叉时间点。

$$\begin{aligned} & \max \tau \in \{0, \ldots, \tau_{max}\} \\ \text{s.t.} \quad & \forall \mathcal{M}(s_i, s_j, \tau) < \theta, \quad (s_i, s_j) \in \{\mathcal{S} \times \mathcal{S}\}, \end{aligned}$$



关键场景差异评估

MARC 解读

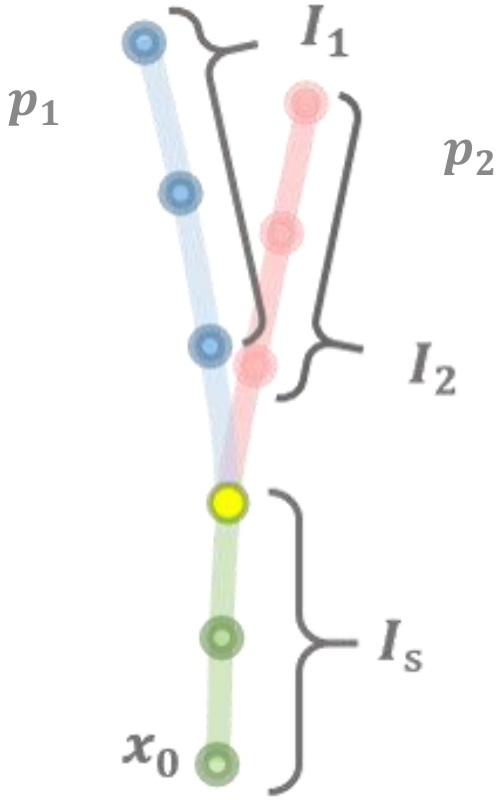
考虑风险兼顾多模态预测的轨迹树生成/风险敏感的应急规划

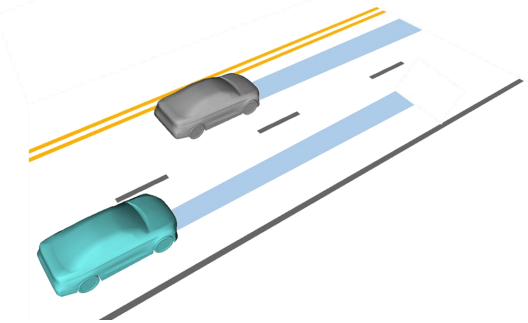
- 典型的轨迹树生成/应急规划。

$$\min_{U} \sum_{j \in I_s} l_j(x_j, u_j) + \sum_{k \in K} \sum_{j \in I_k} l_j(x_j, u_j)$$

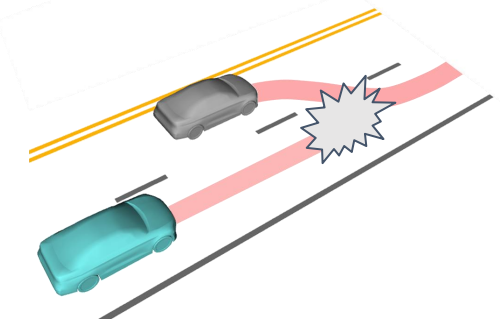
$$\text{s.t.} \quad x_i = f(pre(x_i), u_i), \quad \forall i \in I \backslash \{0\}$$

$$h_i(x_i, u_i) \leq 0, \quad \forall i \in I,$$





prob: 0.9 loss: 0



prob: 0.1 loss: -100

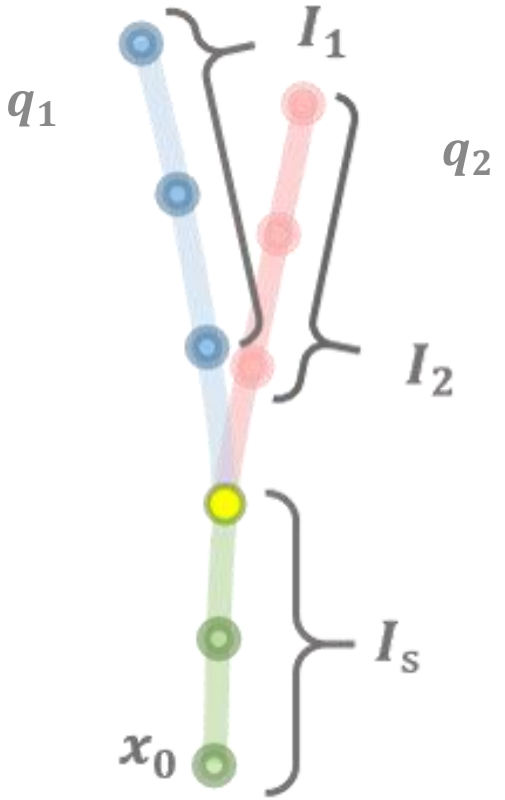
MARC 解读

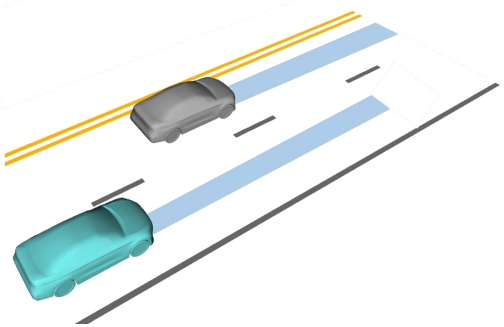
考虑风险兼顾多模态预测的轨迹树生成/风险敏感的应急规划

- 风险价值评估：条件风险价值（conditional value at risk，CVaR），熵风险价值（entropic value at risk）

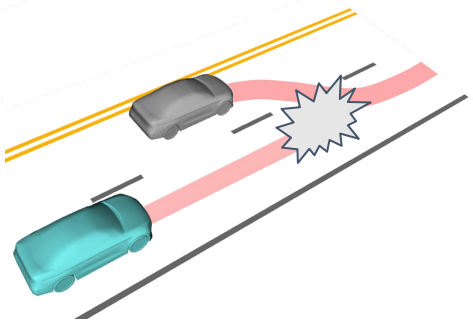
$$\begin{aligned}\mathrm{CVaR}_\alpha = \max_Q & \sum q_k p_k \xi_k \\ \text{s.t.} \quad & 0 \le q_k \le (1-\alpha)^{-1}, k \in K \\ & \sum q_k p_k = 1,\end{aligned}$$

$$\begin{aligned}\max_Q \min_U & \sum_{j \in I_s} l_j(x_j, u_j) + \\ & \sum_{k \in K} \sum_{j \in I_k} \left(p_k q_k l_j^{safe}(x_j, u_j) + l_j^{-safe}(x_j, u_j) \right) \\ \text{s.t.} \quad & x_i = f(pre(x_i), u_i), \quad \forall i \in I \backslash \{0\} \\ & h_i(x_i, u_i) \le 0, \forall i \in I \\ & 0 \le q_k \le (1-\alpha)^{-1}, k \in K \\ & \sum q_k p_k = 1,\end{aligned}$$





prob: 0.9 loss: 0



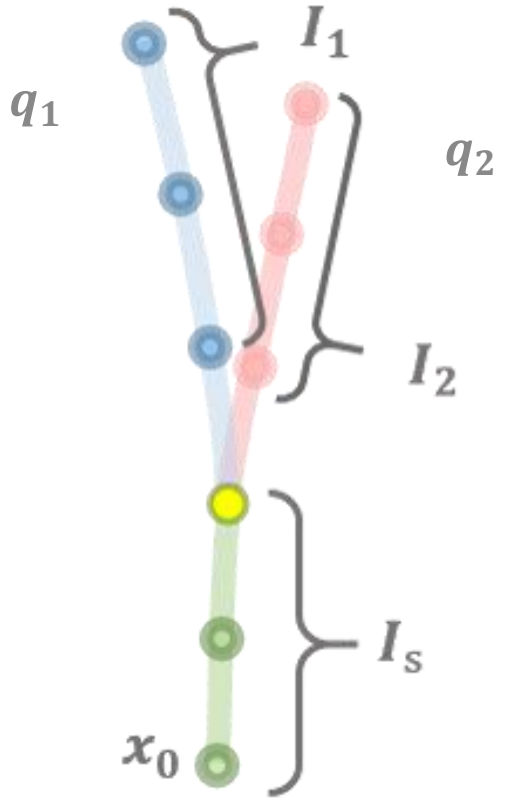
prob: 0.1 loss: -100

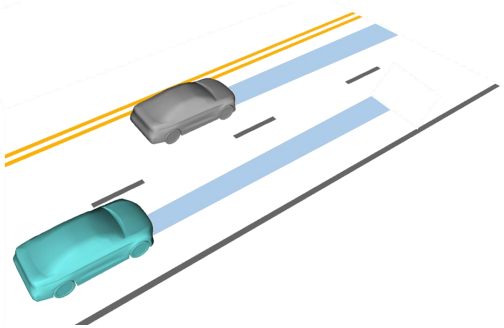
MARC 解读

- 考虑风险兼顾多模态预测的轨迹树生成/风险敏感的应急规划

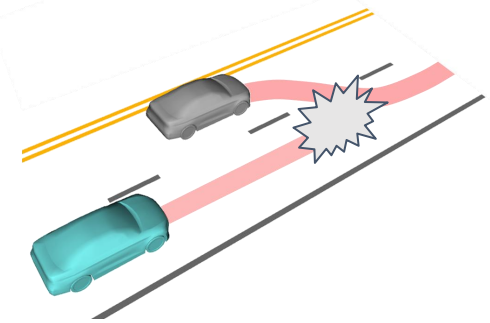
$$
\begin{aligned}
&\max_{Q}\min_{U} \sum_{j\in I_s} l_j(x_j,u_j)+ \\
&\qquad \sum_{k\in K}\sum_{j\in I_k}\left(p_k q_k l_j^{safe}(x_j,u_j)+l_j^{-safe}(x_j,u_j)\right) \\
&\text{s.t.} \quad x_i = f(pre(x_i),u_i), \quad \forall i\in I\backslash\{0\} \\
&\qquad h_i(x_i,u_i)\le 0, \forall i\in I \\
&\qquad 0\le q_k\le (1-\alpha)^{-1}, k\in K \\
&\qquad \sum q_k p_k = 1,
\end{aligned}
$$

$$
\begin{aligned}
U^{k+1} &= \underset{U}{\text{argmin}} \ \text{iLQR}\left(X^k,U^k,Q^k\right) \\
Q^{k+1} &= \underset{Q}{\text{argmax}} \ \text{LP}\left(X^{k+1},U^{k+1},Q^k\right)
\end{aligned}
$$





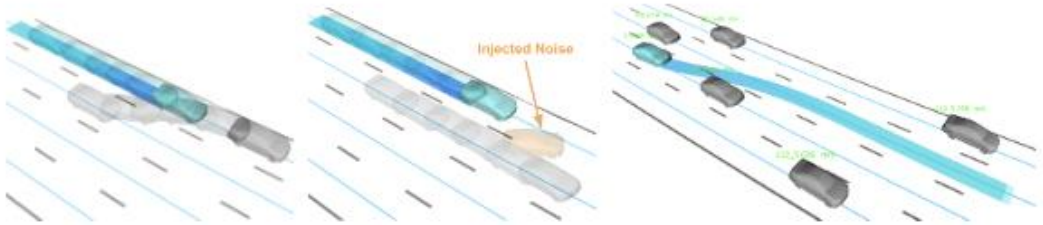
prob: 0.9 loss: 0



prob: 0.1 loss: -100

MARC 解读

实验



(a) cut-in (b) noise injection (c) dynamic traffic

- 消融实验

TABLE II: Ablation Study Results

Methods	Avg Max Dec (m/s^2)	Avg Min Dis (m)	Suc Rate $(\%)$
w/o branch	3.48	1.37	83
Fixed branch	2.52	2.53	93
Dyna branch	1.63	3.45	96
Dyna branch + Risk	**1.14**	**4.12**	**100**

- 对比实验

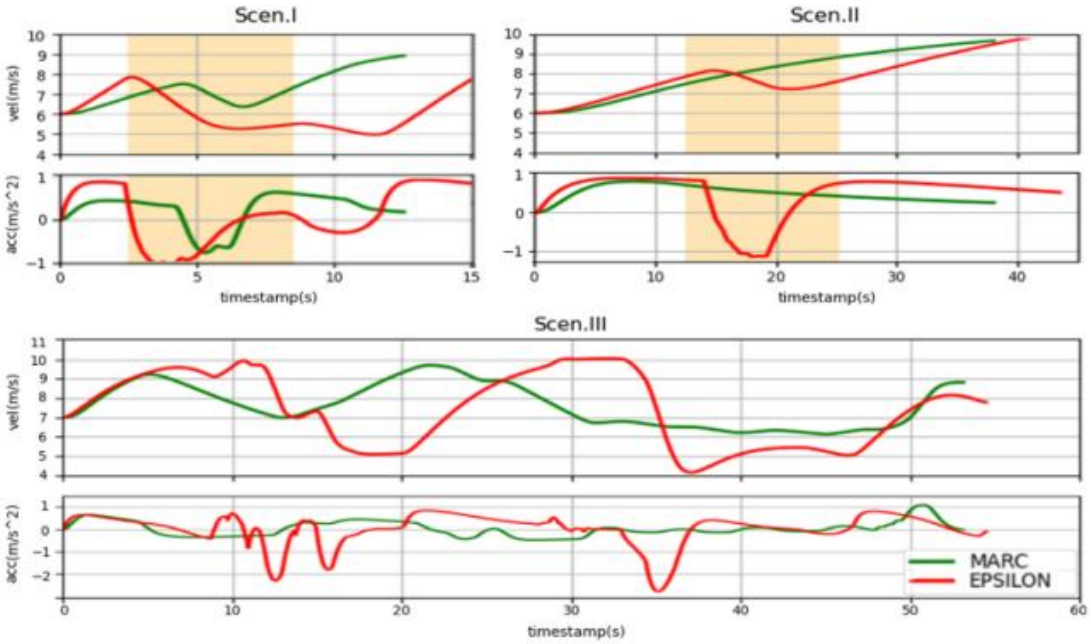


TABLE I: Comparison with EPSILON

Methods		Time (s)	Avg Spd (m/s)	RMS Acc (m/s^2)	Max Abs Acc (m/s^2)
Scen. I	EPSILON	19.90	6.87	0.63	1.03
	MARC	**12.56**	**7.26**	**0.45**	**0.75**
Scen. II	EPSILON	8.70	7.90	0.72	1.13
	MARC	**7.60**	**8.03**	**0.51**	**0.78**
Scen. III	EPSILON	54.46	7.41	0.70	2.78
	MARC	**53.13**	**7.61**	**0.34**	**1.09**

MARC 解读

存在的问题

- 基于意图预测得到场景组合的方式存在组合爆炸等问题。
- 基于模型的前向仿真处理复杂交互的能力有限。
- 基于模型产生的决策泛化性及多样性有限。

对于MARC这个工作感兴趣的话，大家可以在课程答疑群里与作者交流沟通。同时，丁文超老师也会在群内与大家进行直播交流。