Last time:

- ☐ HJ Safety Analysis!
- ☐ safety filters

Lecture 7
IR, Spring '24
Andrea Bajcsy

Today:

- ☐ human behavior prediction
  - ☐ physics-based
  - ☐ planning-based

So far, we have understood a rigorous safety formalism which can compute (automatically!) unsafe states & the safety controller, even under the presence of other agents.

What does this have to do with human behavior prediction?

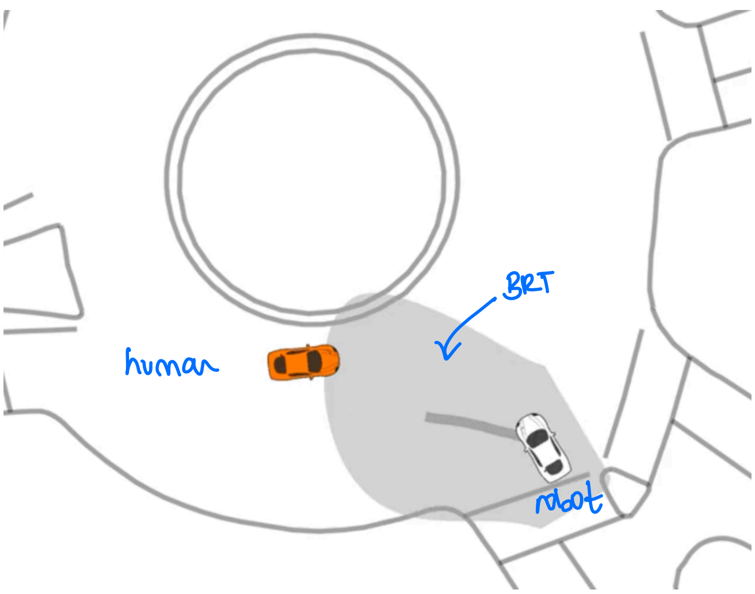
Well, baked into our safety analysis is already one human prediction model!

Hamilton-Jacobi Variational Inequality (HJI-VI)

$$\min\left\{ \ell(x) - V(x,t),\ \frac{\partial V}{\partial t} + \max_{u} \min_{d} \frac{\partial V(x,t)}{\partial x} \cdot f(x,u,d) \right\} = 0$$

$$V(x,T) = \ell(x)$$

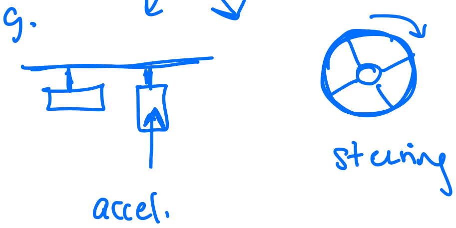
Here, we modelled the other agent (d) who could be the human, as a worst-case adversary. This agent chooses their action $d \in D$



Two key assumptions about human behavior:

1) adversary (maximin)
2) willing to do anything ($d \in D$)

e.g.



A <u>predictive human model</u> is a map $\mathcal{P}$ from history of observations $\theta^{0:t}$ to the future behavior of the human:

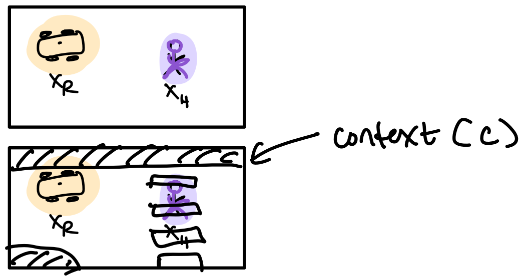
$$\mathcal{P}(\theta^{0:t}) \longmapsto \underbrace{x_H^{t+1:T}}_{\text{human phys. state}} \quad \boxed{\text{OR}} \quad \mathcal{P}(\theta^{0:t}) \longmapsto \underbrace{u_H^{t+1:T}}_{\text{human actions}}$$

(*) 1) "observations" can be:

$\theta = [x_H]$

$\theta = [x_H, x_R] \longleftrightarrow$

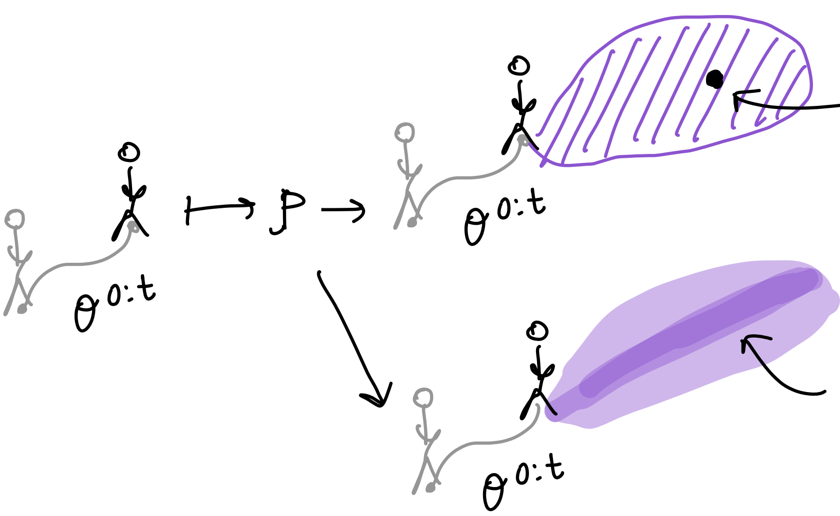
$\theta = [x_H, x_R, c] \longleftrightarrow$



$\theta = [x_H^1, x_H^2, x_H^3, \dots, x_R, c]$

$\theta = [\underbrace{\mu_H, \sigma_H}_{\text{estimate of human } x_H}, x_R, c]$

2) the predicted states or actions could be represented non-deterministically or probabilistically



$x^\tau \in \mathcal{P}(\theta^{0:t})$
$\forall \tau \in [t+1, T]$

<u>NON-DETERMINISTIC</u>

$x^\tau \sim \mathcal{P}(\theta^{0:t})$

<u>PROBABILISTIC</u>

3) frequently, you will see actions output by $\mathcal{P}$ instead of states. What are pros + cons of each?

states

$$\mathcal{P}(o^{0:t}) \longmapsto x_H^{t+1:T}$$

⊕ "baking-in" the dynamics or transition func.

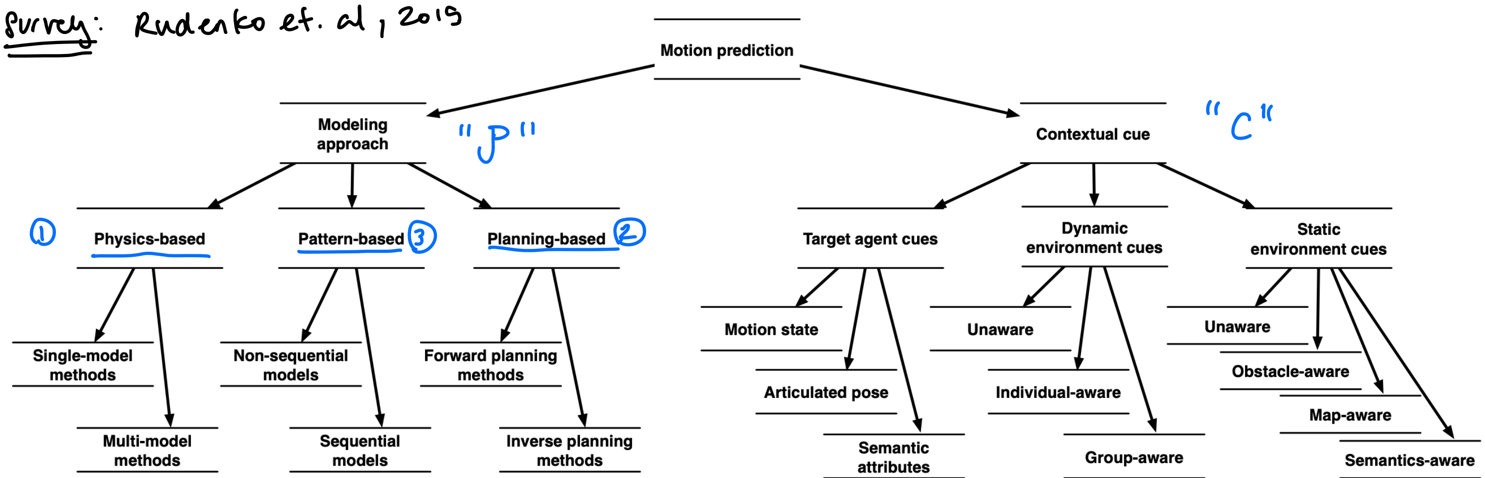
⊕ compatible w/ robot planners "directly"

actions

$$\mathcal{P}(o^{0:t}) \longmapsto u_H^{t+1:T}$$

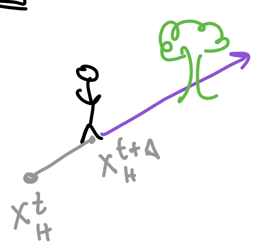
⊕ if $\mathcal{P}$ is NN then this is an easier way to get dynamically feasible predictions (since you will just "roll out" $u_H^{t+1:T}$ w/ model instead of NN having to learn dynamics as well)

<u>survey</u>: Rudenko et. al, 2019



① <u>Physics-based Models</u>: primary guidance for $\mathcal{P}$ is just $\dot{x} = f(x, u)$ and ANY u ∈ U or heuristic u.

$\boxed{A}$ constant velocity prediction: this is THE baseline to beat :)



$$\underbrace{\mathcal{P}(o^{0:t})}\longmapsto a_H^{t+1:T}$$

$$\frac{x_H^{t+\Delta} - x_H^t}{\Delta} = \hat{v} \longmapsto a_H^{t+1:T} = \hat{v} \quad \forall \tau \in \{t+1, \dots T\}$$

⊕ simple!

⊕ can be reasonable on short t-horiz.

⊖ not good @ long-horizon behavior

⊖ no multi-agent interaction

⊖ not safe / robust (all the time)

B "robust" physics-based prediction: obtain ALL states that human could ever get to!

$$\underbrace{P(\theta^{0:t})}_{\text{thinking of all } u \in \mathcal{U}} \longmapsto \underbrace{\mathcal{X}_H^{t+1:T}}_{\text{set of future states}}$$

FORWARD REACHABLE SET (FRS): set of all states a system $\dot{x} = f(x,u)$ can reach starting from an initial state in $\mathcal{L}$ in T timesteps:

reachable state

$$FRS(t) := \{ y : \exists u(\cdot) \in \mathbb{U}_t^T,\ x_0 \in \mathcal{L},\ \mathbf{x}^u_{x_0,t}(T) = y \}$$

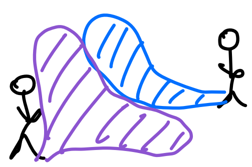
⊕ robust ⇒ strong assurances!

⊖ computationally challenging ⇒ dual to BRT

⊕ in theory, can handle static env ✓

⊖ typically doesn't consider multi-agent interaction

⊖ quite pessimistic (human could do ANYTHING!)



② Planning-based prediction: Idea is that agents - like ppl - are goal / objective driven. If we know what the person's goal / objective is and their optimization process, we can predict them:

$$\underbrace{P(\theta^{0:t})}_{} \longmapsto u_H^{t+1:T}$$

e.g. $\arg\max_{u_H^{t+1:T}} R_H(x^t, u_H^{t+1:T})$

(!) Rationality assumption: human optimizes + acts under their objective, thinking about the impact of $\vec{u}_H$ on the future.

Two big Q's:

Q1) what "optimizer" is human using? ← (e.g. "argmax")

Q2) what objective are they optimizing? ← (e.g. R_H)

We will take perspective of RL / optimal ctrl. for Q1).

"Dynamic Prog. Human"

$$\begin{cases} V_t(x) := \max_{u_H^t} Q_t(x, u_H) \\ Q_t(x, u_H) := r_H(x, u_H) + \mathbb{E}_{x^{t+1} \sim f(x^t, u_H^t)}\left[V_{t+1}(x^{t+1}) \right] \end{cases}$$

$$\boxed{u_H^*(x^t) := \arg\max Q_t(x, u_H)}$$

"Traj Opt Human"

$$u_H^{t+1:T} = \arg\max_{u_H^{t+1:T}} R_H(x, u_H^{t+1:T})$$

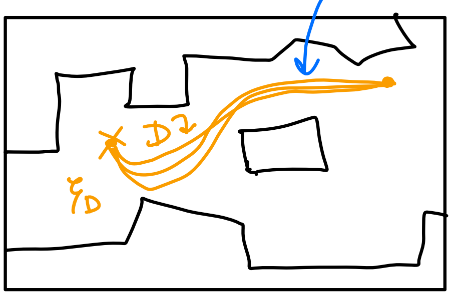
← (Pontryagin max Principle)

↑ planned traj, perhaps person replans @ some fixed freq.

↳ will see both in the class...

But who gives us $\boxed{R_H}$?? This is Q2). This is where IRL (inverse RL) comes into play!

IDEA: $\xi := \{(x_1, u_1), \ldots, (x_T, u_T)\}$ sequence of states/actions over time



Given observations / data of behavior, find the reward function R_H which best explains the data, $\xi_D \in D$.

PROBLEM: ill-posed b/c many R_H can yield the same behavior ξ_D additionally, ξ_D are noisy too...

2008, Ziebart et al.

$\boxed{\text{Max Ent IRL}}$ assume the demos are drawn from some distribution

$\xi_D \sim P(\cdot)$ instead of $\arg\max_{\xi_D}$. let's find this distribution which makes minimal assumptions about behavior, while still "matching" $\xi_D \in D$.

- recall : $\uparrow$ entropy $\Rightarrow$ $\uparrow$ uncertainty. Finding $P(\cdot)$ that maximizes entropy over ξ subject to "matching" the demo. data, we will avoid favoring any particular ξ traj over another one so long as they satisfy the "feature matching" constraints.

prevent overfitting

$$\boxed{\begin{aligned} &\max_{P} \int -P(\xi)\log P(\xi)\, d\xi \\ &\text{s.t.} \quad \mathbb{E}_{\xi\sim P(\xi)}\left[\phi(\xi)\right] = \phi_D = \frac{1}{|D|}\sum_{\xi_D \in D} \phi(\xi_D) \\ &\text{valid } P(\cdot) \text{ dist.} \begin{cases} \int P(\xi)\, d\xi = 1 \\ P(\xi) \geq 0 \quad \forall \xi \in \Xi \end{cases} \end{aligned}}$$

match empirical feature count of data D

space of all trajectories

☺ solve this as exercise

Solution:

$$P(\ldots) \equiv P^*(\xi;\theta) = \frac{e^{\theta^T\phi(\xi)}}{\int e^{\theta^T\phi(\xi)}\, d\xi}$$

unknown!

this is our pred. model!

$\equiv R_H(\xi;\theta)$ is linear combo of weights $\theta \in \mathbb{R}^k$ and features $\phi : \Xi \to \mathbb{R}^k$. Assume features known, we just need to learn θ!

Here, θ is the only param we need to learn! Can be found via

$$\theta^* = \arg\max_{\theta} \sum_{\xi_D \in D} \log P(\xi_D;\theta)$$

⊕ sample efficient b/c structure (small # params to learn)

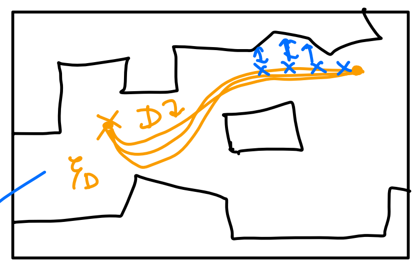
⊕ fairly comp.-efficient to learn θ in $|D|$

⊕ easy to pair w/ planning / decision-making

⊖ R_H representation is LINEAR combo of $\theta^T\phi(\xi)$

⊖ what are the right features ϕ?

⊖ what about history / non-markovian ...



$$\underset{\text{features}}{\phi(\xi)} = \begin{bmatrix} \text{distToObs}(\xi) \\ \text{speed}(\xi) \\ \vdots \end{bmatrix}$$

Good News: moves us away from treating humans as perfect optimizers, & towards noisily-rational optimizers!