SAFETY THROUGH DISRUPTION

LEARNING ACTIVE INFERENCE MODELS OF PERCEPTION AND CONTROL: APPLICATION TO CAR FOLLOWING TASK

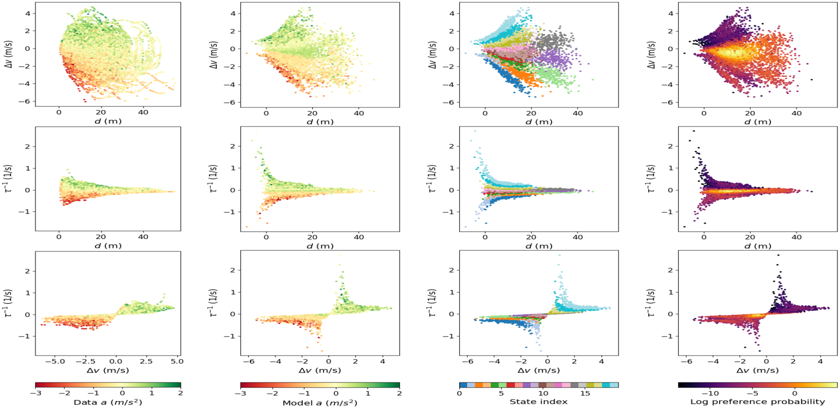
Alfredo Garcia

Texas A&M University

Joint work with:

Ran Wei (Verses), A. McDonald (Wisconsin), G. Markkula (Leeds U.), J. Engstrom (Waymo), M. O'Kelly (Waymo)

A Preview of Results



(a) Data (b) Model (c) Perception (d) Preferences

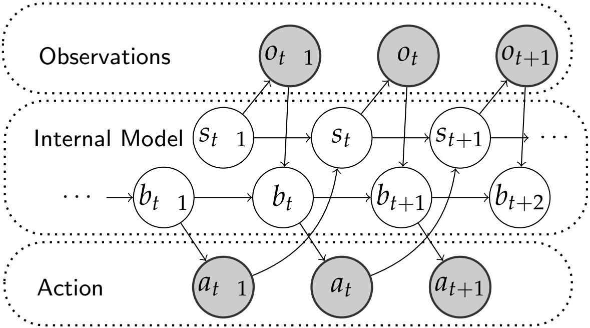
Figure 1: Visualizations of active inference model.

Outline

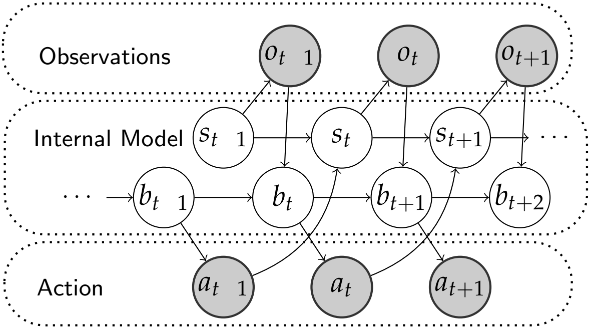
Outline

A POMDP Model

- A generative model of observations $\mathbb{T}(o_t|s_t)$.

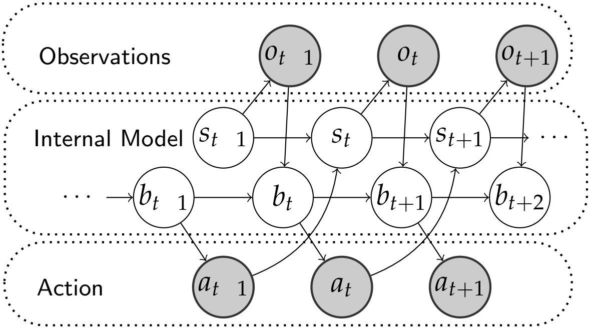


A POMDP Model



- A generative model of observations $\mathbb{T}(o_t|s_t)$.
- A belief distribution about the hidden state $b_t(s) = \mathbb{P}(s_t = s|h_t)$

A POMDP Model



- A generative model of observations $\mathbb{T}(o_t|s_t)$.
- A belief distribution about the hidden state $b_t(s) = \mathbb{P}(s_t = s|h_t)$
- A representation of state dynamics, i.e. a transition to a new state s_{t+1} takes place with probability $\mathbb{P}(s_{t+1}|s_t, a_t)$

A POMDP Model

- After $t > 0$ time periods, the observable history of observations and actions is denoted by

$$h_t := \{o_t, ..., o_0, a_{t\ \ 1}, ..., a_0\} \in H_t$$

A POMDP Model

- After $t > 0$ time periods, the observable history of observations and actions is denoted by

$$h_t := \{o_t, ..., o_0, a_{t-1}, ..., a_0\} \in H_t$$

- Denoting control policies (possibly random) by $\pi(\cdot|h_t)$, the POMDP model is the solution to:

$$\max_{\pi} \mathbb{E}\Big[\sum_{t \geq 0} \gamma^t [r(s_t, a_t) - c(\pi(\cdot|h_t))]\Big]$$

where $r(s_t, a_t)$ is the reward and $c(\pi(\cdot|h_t))$ information processing cost.

A Bayesian Agent

- A Bayesian agent forms beliefs b_t about the state of the environment:

$$b_t(s) = \mathbb{P}(s_t = s | h_t)$$

A Bayesian Agent

- A Bayesian agent forms beliefs b_t about the state of the environment:

$$b_t(s) = \mathbb{P}(s_t = s | h_t)$$

- When implementing action a_t under beliefs b_t, the agent expects:
  - a reward

$$r(b_t, a_t) := \sum_s r(s, a_t) b_t(s)$$

  - observation o_{t+1} with probability:

$$\sigma(o_{t+1}|b_t, a_t) := \sum_{s_{t+1}} \sum_{s_t} \mathbb{T}(o_{t+1}|s_{t+1}) \mathbb{P}(s_{t+1}|s_t, a_t) b_t(s_t)$$

A Bayesian Agent

- With Markovian dynamics and additive reward the model of optimal behavior has recursive structure:

$$V^*(b) = \max_{\pi(\cdot|b)} \Big\{ \sum_s \sum_a r(s,a)\pi(a|b)b(s) \quad c(\pi(\cdot|b)) \\ + \gamma \sum_a \sum_{o'} \sigma(o'|b,a)\pi(a|b)V^*(b') \Big\}$$

where b' is the resulting belief when observation o' is recorded after implementing action a.

A Bayesian Agent

- With the information processing cost as Kullback-Leibler divergence between the control policy and a default policy π^0, i.e.

$$c(\pi(\cdot|b)) = \mathcal{D}_{KL}(\pi(\cdot|b)||\pi^0(\cdot|b))$$

A Bayesian Agent

- With the information processing cost as Kullback-Leibler divergence between the control policy and a default policy π^0, i.e.

$$c(\pi(\cdot|b)) = \mathcal{D}_{KL}(\pi(\cdot|b)||\pi^0(\cdot|b))$$

- The model is of the form:

$$\pi^*(a|b) = \frac{\pi^0(a|b)\exp \; Q^*(b,a)}{\sum_{a' \in A} \pi^0(a'|b)\exp \; Q^*(b,a')} \tag{1}$$

where

$$Q^*(b,a) := r(b,a) + \gamma \sum_{o'} \sigma(o'|b,a) V^*(b') \tag{2}$$

Outline

Learning a Model of Perception and Action

Based upon data $\mathcal{D}$ (i.e sequences of observations and implemented actions say τ) estimate the primitives of the perception & control model:

Learning a Model of Perception and Action

Based upon data $\mathcal{D}$ (i.e sequences of observations and implemented actions say τ) estimate the primitives of the perception & control model:

- Perception The agent's internal representation: $\mathbb{P}_{\theta_1}(s'|s,a)$ and $\mathbb{T}_{\theta_1}(o'|s')$ parametrized by $\theta_1 \in \mathbb{R}_1^p$.

Learning a Model of Perception and Action

Based upon data $\mathcal{D}$ (i.e sequences of observations and implemented actions say τ) estimate the primitives of the perception & control model:

- Perception The agent's internal representation: $\mathbb{P}_{\theta_1}(s'|s,a)$ and $\mathbb{T}_{\theta_1}(o'|s')$ parametrized by $\theta_1 \in \mathbb{R}_1^p$.
- Preferences A reward function $r_{\theta_2}(b,a)$ which is parametrized by θ_2

Learning a Model of Perception and Action

The log-likelihood of dataset $\mathcal{D}$ can be written as:

$$
\begin{aligned}
\log \mathbb{P}(\mathcal{D}|\theta) &= \log \prod_{\tau \in \mathcal{D}} \mathbb{P}(\tau|\theta) \\
&= \mathbb{E}_{\tau \sim \mathcal{D}} \Big[\sum_{t=0}^{T} \log \Big(\pi^*_\theta(a_t | b_{\theta_1,t}) \mathbb{P} \;\; o_{t+1} | h_t \cup \{a_t\} \;\; \Big) \Big] |\mathcal{D}| \\
&= \mathbb{E}_{\tau \sim \mathcal{D}} \Big[\sum_{t=0}^{T} \log \pi^*_\theta(a_t | b_{\theta_1,t}) \Big] |\mathcal{D}| + \text{constant}
\end{aligned}
$$

Learning a Model of Perception and Action

Assumption 1: $P(\theta) = P(\theta_1)P(\theta_2)$, where:

$$P(\theta_1) \propto \exp\left(\lambda \mathbb{E}_{\tau\sim\mathcal{D}}\left[\prod_{t=0}^{T} \sigma_{\theta_1}(o_{t+1}|b_{\theta_1,t}, a_t)\right]|\mathcal{D}|\right)$$

for some $\lambda > 0$.

Assuming a uniform prior $P(\theta_2)$ on a compact subset $\Theta_2 \subset \mathbb{R}_2^p$, the log of the posterior distribution can be written as:

$$\begin{aligned}\log P(\theta|\mathcal{D}) &= \log P(\mathcal{D}|\theta) + \log P(\theta_1) + \text{constant} \\ &= \mathbb{E}_{\mathcal{D}}\Big[\log \sum_{t=0}^{T} \pi_\theta^*(a_t|b_{\theta_1,t}) + \lambda \sum_{t=0}^{T} \log \sigma_{\theta_1}(o_{t+1}|b_{\theta_1,t}, a_t)\Big]|\mathcal{D}| \\ &\qquad + \text{constant}\end{aligned}$$

Learning a Model of Perception and Action

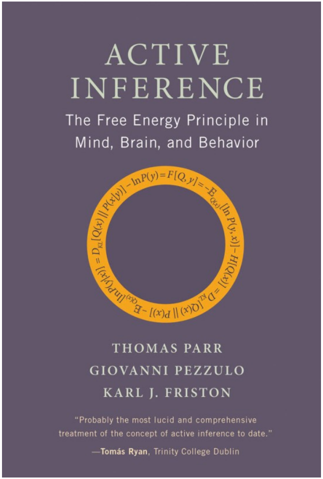
The estimation problem as the following bi-level optimization problem:

$$\max_{(\theta_1,\theta_2)} \mathbb{E}_{\mathcal{D}}\Big[\log \sum_{t=0}^{T} \pi^*_\theta(a_t|b_{\theta_1,t}) + \lambda \sum_{t=0}^{T} \log \sigma_{\theta_1}(o_{t+1}|b_{\theta_1,t}, a_t)\Big]$$

$$\text{s.t.} \quad \pi^*_\theta = \arg\max_{\pi\in\Pi^H} \mathbb{E}\Big[\sum_{h\leq H} [r_\theta(b_h, a_h) \quad \log \pi(\cdot|b_h)]\Big]$$

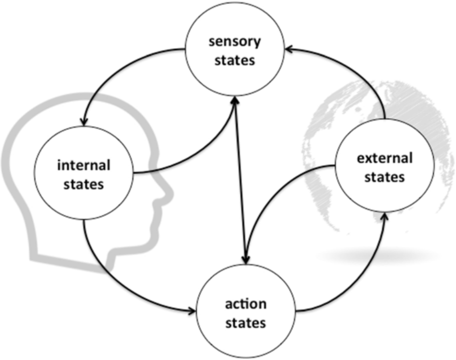
Outline

Active Inference and Free Energy



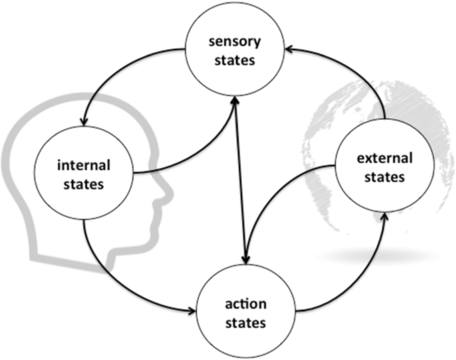
Active inference is a novel framework for cognition and behavior according to which the agent jointly perceives and acts upon the world so as to maximize the match between perceived vs preferred states of the world.

Active Inference and Free Energy



A principle of *free energy minimization*:

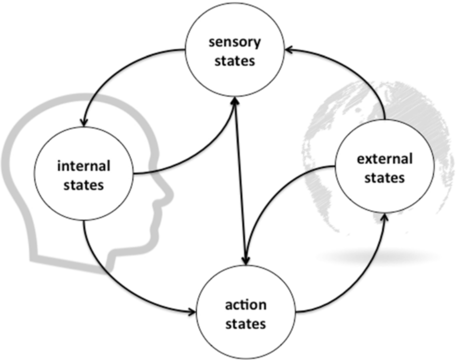
Active Inference and Free Energy



A principle of *free energy minimization*:

- (backward) free energy is minimized when the agent's belief distribution b_t corresponds to the Bayes updated belief distribution on the state s_t.

Active Inference and Free Energy



A principle of *free energy minimization*:

- (backward) free energy is minimized when the agent's belief distribution b_t corresponds to the Bayes updated belief distribution on the state s_t.
- (forward) surprise is measured with respect to a *preferred* distribution $\tilde{P}(s_{t+1})$ over states of the environment.

The immediate "surprise" associated with action a_t when current beliefs are b_t is quantified by the *expected free energy* defined as:

$$EFE(b_t, a_t) = \mathbb{E}\left[D_{KL}\ \ b_{t+1}||\tilde{P})\right] + \mathbb{E}\left[\mathcal{H}(\mathbb{T}(\cdot|s_{t+1}))\right]$$

where

$$b_{t+1}(s) = \mathbb{P}(s_{t+1} = s|h_t \cup \{a_t, o_{t+1}\})$$

and $\mathcal{H}(\mathbb{T}(\cdot|s_{t+1}))$ is the entropy of the resulting generative model of observations, i.e.:

$$\mathcal{H}(\mathbb{T}(\cdot|s_{t+1})) := \quad \sum_{o'} \mathbb{T}(o'|s_{t+1}) \log\Big(\mathbb{T}(o'|s_{t+1})\Big).$$

Outline

4 Application to Car Following Task

Application to Car Following Task, Ran et al. (2023)

- We use the active inference specification (reward equal to negative free energy).
- We use the INTERACTION dataset: a set of time-indexed trajectories of the positions, velocities, and headings of each vehicle in the scene in the map's coordinate system at a sampling frequency of 10 Hz.

Application to Car Following Task

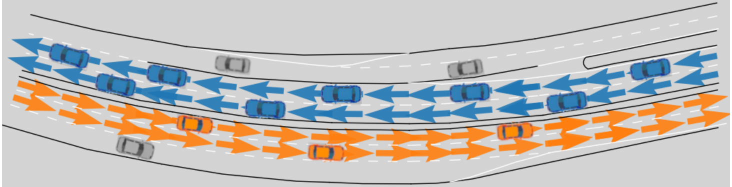


Figure 2: Top down view of the roadway in Dataset

Application to Car Following Task

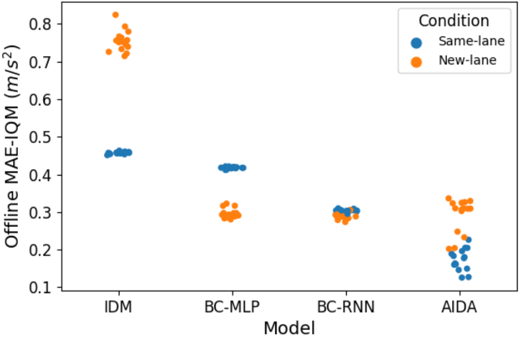


Figure 3: Offline evaluation MAE-IQM. Each point corresponds to a random seed used to initialize model training and its color corresponds to the testing condition of either same-lane or new-lane.

Application to Car Following Task

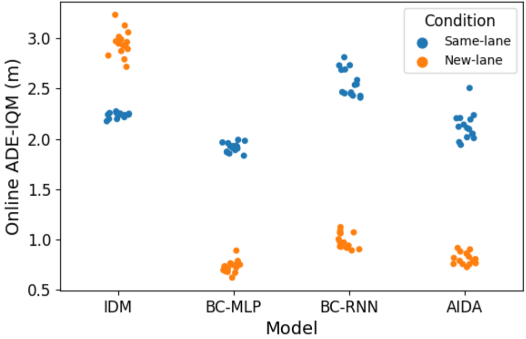


Figure 4: Online evaluation ADE-IQM. Each point corresponds to a random seed used to initialize model training and its color corresponds to the testing condition of either same-lane or new-lane.

Application to Car Following Task

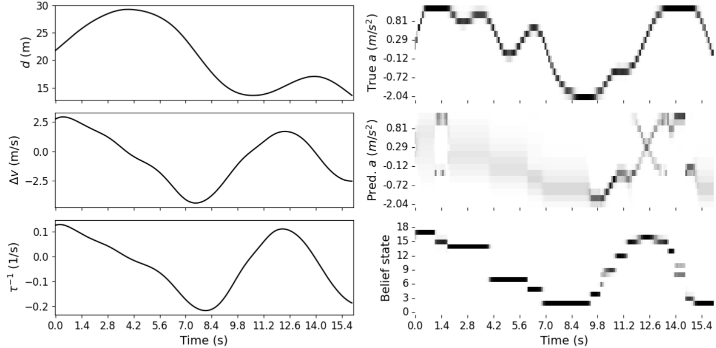


Figure 5: Visualizations of a same-lane offline evaluation trajectory

Application to Car Following Task

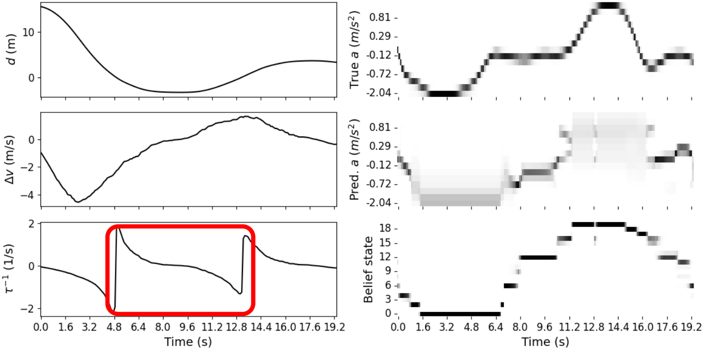


Figure 6: Visualizations of a same-lane online evaluation trajectory where the AIDA generated a rear-end collision with the lead vehicle.

Outline

5 Conclusions

Conclusions

- We proposed a novel model of driver behavior using active inference (AIDA).
- Using car following data, we showed that the AIDA significantly outperformed the rule-based IDM on all metrics and performed comparably with the data-driven neural network benchmarks.
- We showed that the structure of the AIDA provides superior interpretability of its input-output mechanics than the neural network models.
- Future work should focus on training with data from more diverse driving environments and examining model extensions that can capture heterogeneity across drivers