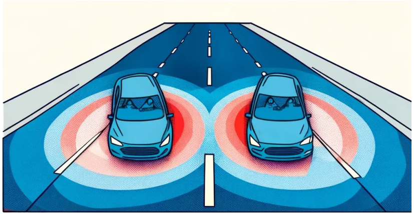
Explaining the Behavior of POMDP-based Agents Through the Impact of Counterfactual Information

AAMAS-2024, Auckland, New Zealand

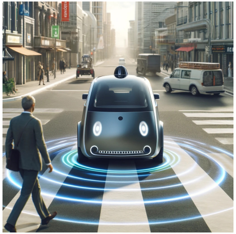
Saaduddin Mahmud, UMass Amherst, USA
Marcell Vazquez-Chanlatte, Nissan North America, USA
Stefan Witwicki, Nissan North America, USA
Shlomo Zilberstein, UMass Amherst, USA

University *of* Massachusetts Amherst

Motivation



(a) Uncertainty over nearby vehicle positions.



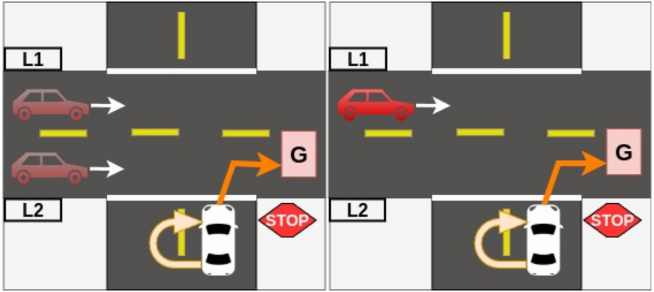
(b) Uncertainty over human intention.

- Real-world decision-making scenarios contain uncertainty over features/variables.
- Uncertainty due to measurement error (perception sensor inaccuracy [right]).
- Uncertainty due to partially observed variables e.g. human intention [left].
- Under uncertainty agents can exhibit different types of behavior; such as goal-driven behavior and information-seeking behavior.

Motivation

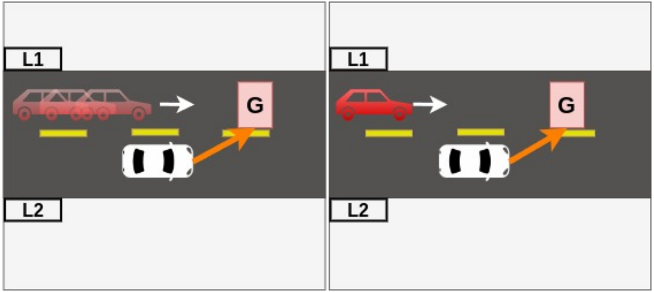
- POMDPs are often used to model state feature uncertainty in sequential decision-making problems.
- While ubiquitous no prior works attempted to explain how feature uncertainty affects an agent's behavior or tried to distinguish between goal-driven vs information-seeking behavior.
- The sequential nature of the problem makes it difficult to apply existing feature attribution methods.

Example 1



- Left: White car waits due to uncertainty about the correct lane position of the oncoming red vehicle.
- Right: The agent could go to goal 'G' quicker if the correct lane position was provided.

Example 2

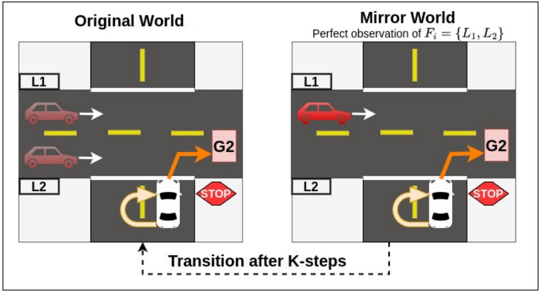


- Left: White car waits to change its current lane due to uncertainty about the correct position and speed of the vehicle on the L1 (top) lane.
- Right: Sensor accumulates errors over time. Therefore, a single step of information probing might not be enough to change behavior.

Sequential Information Probing: Key Ideas

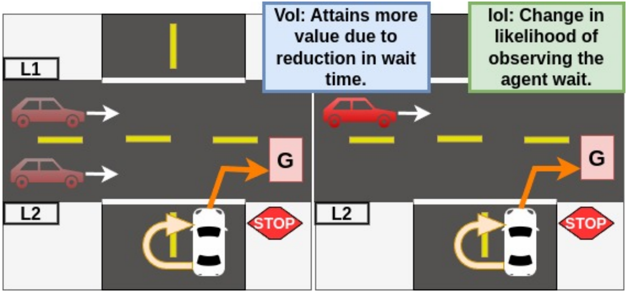
- Key Idea 1: Sequential Information Probing.
- Key Idea 2: Quantification of Change in Behavior.
- Key Idea 3: Marginal Attribution of Information.

Key Idea 1: Sequential Information Probing



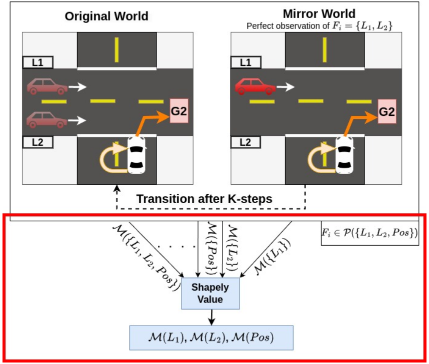
- Mirror World: Agent is provided with perfect information about different subsets of the state features.
- Sequential Probing: Agent is provided information sequentially.

Key Idea 2: Quantification of Change in Behavior



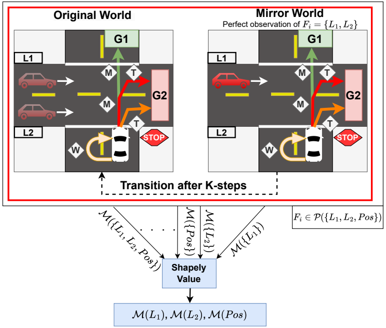
- Value of Information (VoI): Increase in Value (i.e. attained utility/returns) due to provided information
- Influence of Information (IoI): Change in the likelihood of observing a behavior (i.e. trajectory) due to provided information.

Key Idea 3: Marginal Attribution of Information



- Marginal Attribution: The difference in the value or likelihood is marginalized using the Shapley value method to highlight important information.

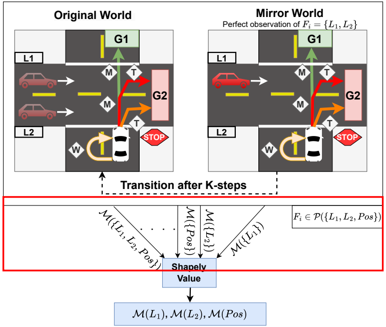
Method (Key Idea 1): Sequential Information Probing



- Mirror Worlds: Different mirror worlds provide perfect information about different subsets of the features. Including the original world there are 2^N mirror worlds.
- Probing Strategy:
  - [*KS*]: The agent stays exactly K steps in the mirror world.
  - [*GE*]: The agent stays K steps in expectation in the mirror world where $K \sim Geometric(1 - \lambda)$
  - [*MY*]: The agent remains in the mirror world for K steps but the agent is not aware of the probing.

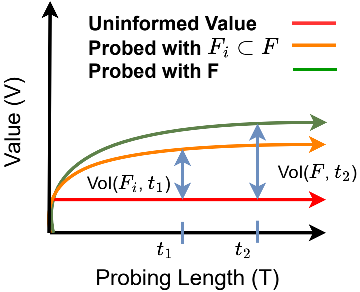
Method (Key Idea 2): Quantification of Change in Behavior

2 Proposed Method



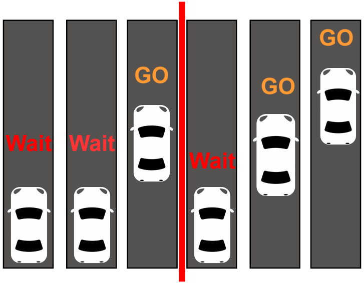
- We want to quantify the difference in behavior between an uninformed and an information-probed agent.
- Value of Information (VoI).
- Influence of Information (IoI).

Method (Key Idea 2): Value of Information (VoI)



- Expected difference between uninformed value function and information probed value function.
- As we increase the probing length and provide information about a larger subset of the features we gain more value.
- When probed infinitely long and with perfect information about all the features VoI is just a gap between QMDP upper bound.

Method (Key Idea 2): Influence of Information (IoI)



- Expected change in negative log-likelihood of observing a given trajectory.
- The likelihood of observing $WAIT - WAIT - GO$ goes down and $WAIT - GO - GO$ goes up when we probe the agent with lane position information (Example 1).

Method (Key Idea 2): Calculating VoI and IoI

$$P_{F_i} = \langle S, A, O, T, \Omega, R, \gamma \rangle$$

$$\downarrow$$

$$\overline{P_{F_i}} = \left\langle \overline{S}, A, \overline{O}, \overline{T}, \overline{\Omega}, R, \gamma \right\rangle$$

- Discrete State Space: We construct a new probing augmented POMDP.
- Augmented state: remaining probing steps.
- Augmented observation: If being probed add feature information.
- Solving this augmented POMDP allows $O(S)$ calculation of VoI, and $O(ST)$ of IoI.

Method (Key Idea 2): Calculating VoI and IoI

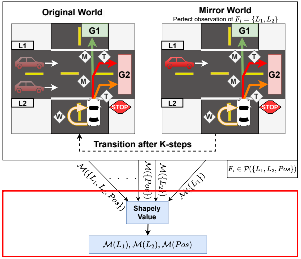
$$P_{F_i} = \langle S, A, O, T, \Omega, R, \gamma \rangle$$

$$\downarrow$$

$$\overline{P_{F_i}} = \left\langle \overline{S}, A, \overline{O}, \overline{T}, \overline{\Omega}, R, \gamma \right\rangle$$

- Continues State Space: We propose Meta-CDQL algorithm, a deep Q-learning-based algorithm for estimating counterfactual information probed values.
- Solves the different augmented-POMDP as a Multi-task problem.

Method (Key Idea 3): Marginal Attribution of Information.



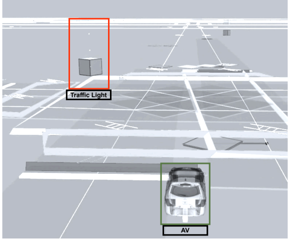
- We compute the marginal contribution of each feature using the Shapley value method.
- Interpretation: Marginal VoI will divide the expected K-step QMDP bound gap among features: "Value being lost due to uncertainty about a feature"
- Interpretation: Marginal IoI will divide the expected gap in the likelihood of observing a given trajectory under the K-step QMDP policy and original policy: "Likelihood increased or decreased due to uncertainty about a feature"

Key Ideas Recap

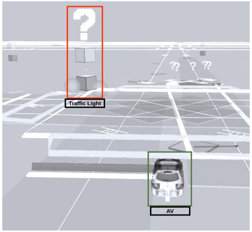
- Key Idea 1: Sequential Information Probing.
- Key Idea 2: Quantification of Change in Behavior.
- Key Idea 3: Marginal Attribution of Information.

Case Study

3 Results



(a) Waiting at a controlled interaction



(b) Sudden change in light state.

- We run a case study on an Autonomous vehicle in an urban environment. The vehicle models the world with a set of POMDPs.
- We put a '?' on object corresponding to each feature. We control the opacity based on the normalized magnitude of VoI.
- Sudden change of the traffic light state causes uncertainty and '?' lights up on top of the traffic light (right).

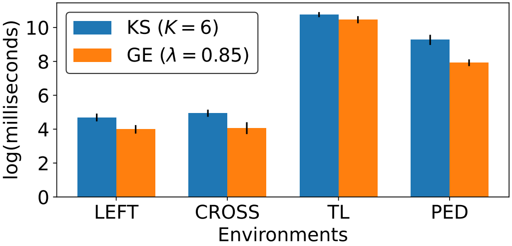
- We discussed several theoretical properties and bounds on the value of VoI and IoI.
- We established a direct relation between VoI and IoI:

$$|IoI(\tau) - \sum_{t=0}^{T}[VoI(b_t) - QoI(b_t)]| \leq log(|A|)$$

.

Quantitative Results

- Correlation: We finds that VoI and IoI is highly correlated. This indicates that the behavior of POMDP agents is driven by feature uncertainties that have high values.
- Consistence: We find that for discrete state space, the produced explanations are highly self-consistent. In continuous cases, due to higher approximation error, the consistency is comparatively lower.
- Runtime: We also compare computational requirement of different probing methods.



Key Ideas Recap

- Key Idea 1: Sequential Information Probing.
- Key Idea 2: Quantification of Change in Behavior.
- Key Idea 3: Marginal Attribution of Information.

Conclusions and Future work

- We propose a novel method for explaining how uncertainty affects a sequential decision-making agent.
- We propose two metrics VoI and IoI to quantify the long-term effect of uncertainty.
- We propose an efficient method to calculate VoI and IoI and apply the Shapley value method to calculate the marginal effect of feature uncertainty.
- In the future, we want to extend our work to offline RL settings.

Explaining the Behavior of POMDP-based Agents Through the Impact of Counterfactual Information

Thank you for watching this presentation!