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

Saaduddin Mahmud[1], Marcell Vazquez-Chanlatte[2], Stefan Witwicki[2] and Shlomo Zilberstein[1]

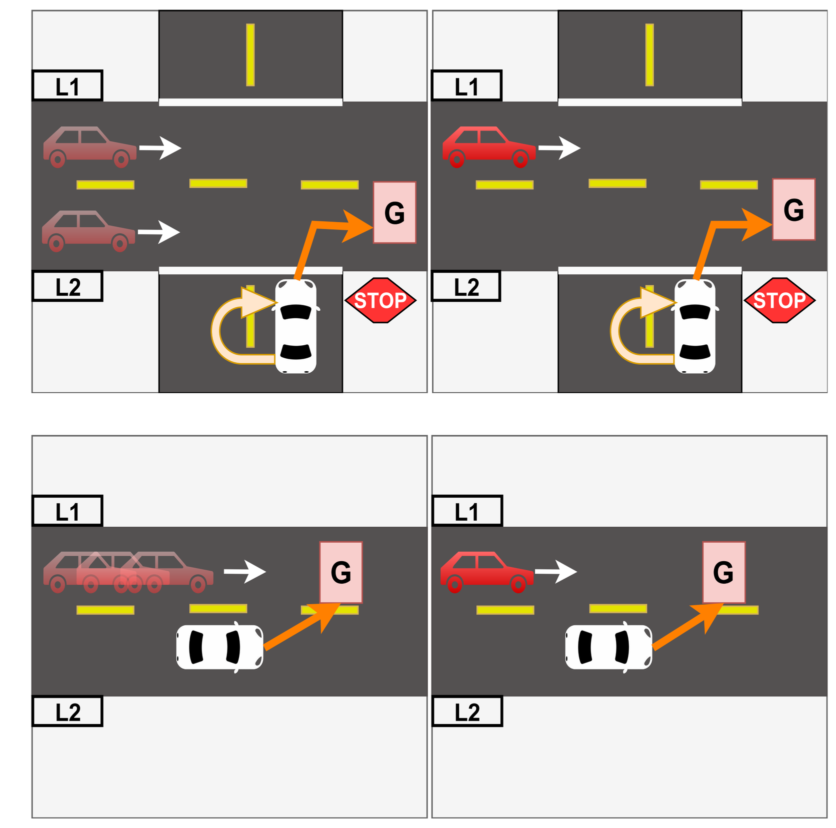
[1] University of Massachusetts Amherst, [2] Nissan Advanced Technology Center Silicon Valley

Quick Overview

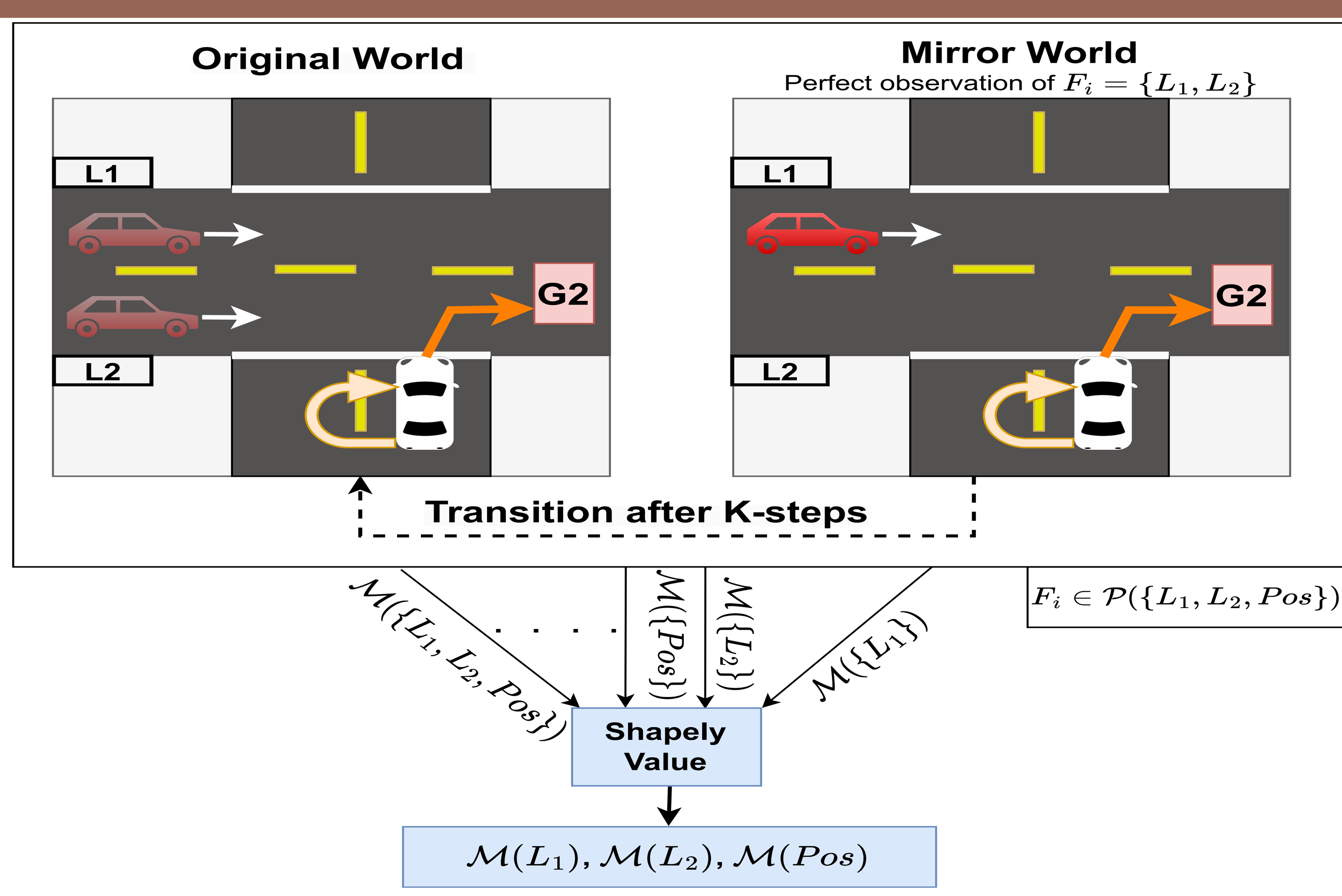
- **Objective:** We present SIP: Sequential Information Probing, a novel framework designed to explain how uncertainty in state features influences agent behavior in sequential decision-making.
- **Significance:** In real-world scenarios, decision-making is often hampered by partially observed features. Despite the prevalence:
  - Existing studies have largely overlooked this specific challenge.
  - The dynamics of sequential decision-making complicate the use of traditional feature attribution methods designed for supervised machine learning tasks.
- **Methodology:** Our approach involves probing the agent with a carefully designed sequence of perfect counterfactual information regarding state features, followed by a quantification of the impact on agent's behavior by contrasting with the uninformed agent.
- **Validation:** We conduct both theoretical and empirical analyses to validate our framework. A case study involving a functioning autonomous vehicle (AV) system demonstrates the practical applicability and effectiveness of our method.

Example

- **One-Step Information Probing**
  - The White car on the left waits to make a more accurate estimation of the lane position of the oncoming red vehicle before making a right turn.
  - If we probe with white care with the lane position information it can avoid waiting.
- **Sequential Information Probing**
  - Estimation of both position and speed needed when safe lane change.
  - Each step the white car moves it degrades its estimation of the other car. Therefore, A single step of information probing might not be sufficient to elicit lane-changing behavior.



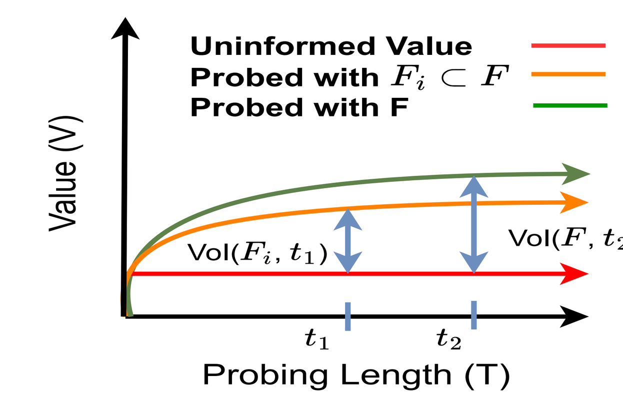
SIP: Sequential Information Probing



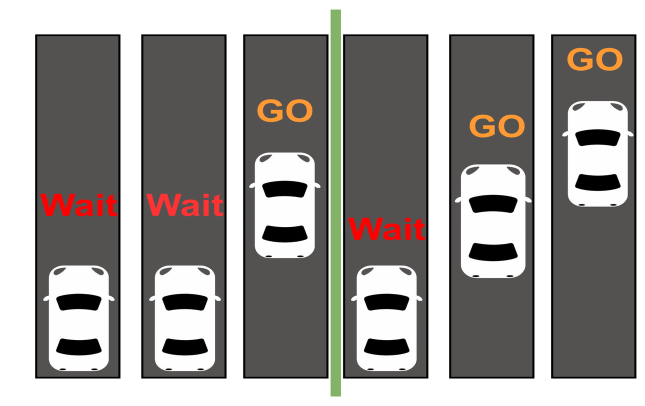
- Sequential Probing: Prob the agent with a sequence of perfect information about different subsets of the state features.
- Quantification: Calculate metrics such as VoI and IoI by contrasting the probed agents with the uninformed agent.
- Attribution: Compute and present the marginal contribution of each feature using the Shapley value method.

Value and Influence of Information (VoI and IoI)

- **VoI: Value of Information**
  - Expected difference between the value attained by the uninformed agent and information probed agent.
  - As we increase the probing length and provide information about a larger subset of the features the agent attains more value.

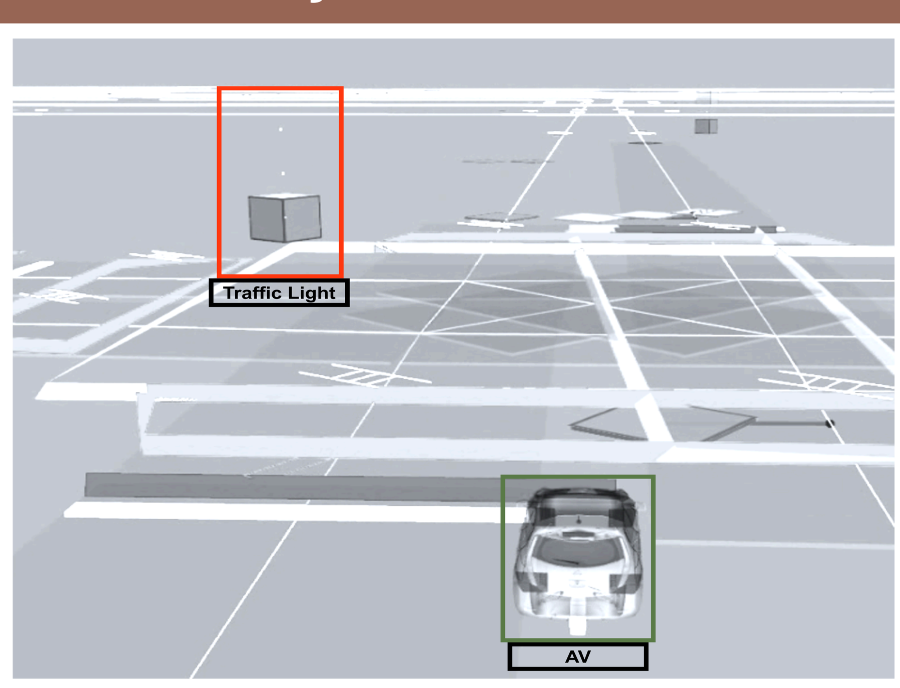


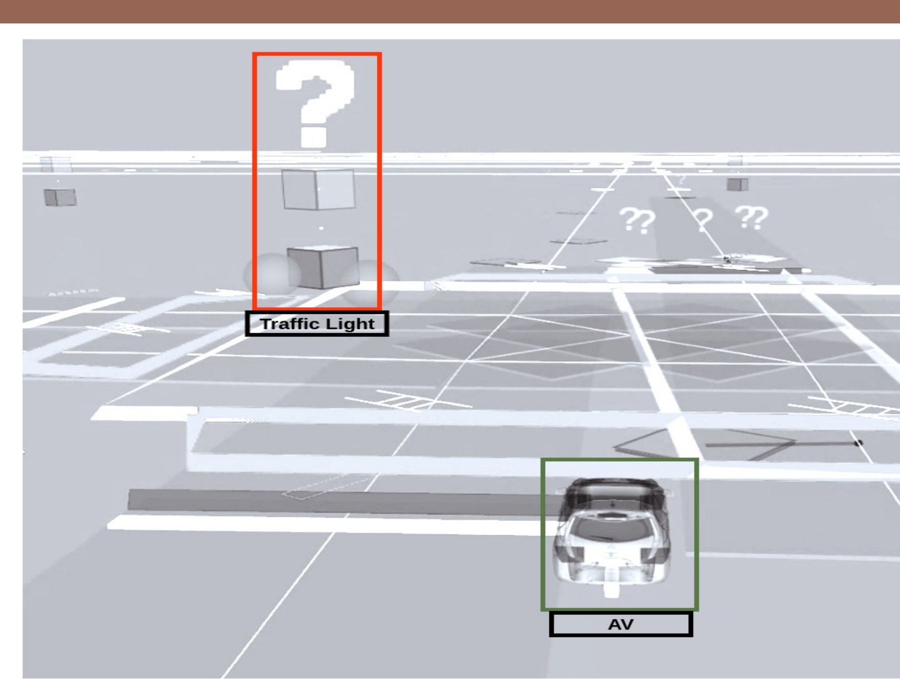
- **IoI: Influence of Information**
  - Expected change in the likelihood of observing a particular trajectory given probed information.
  - The likelihood of observing ***WAIT* − *WAIT* − *GO*** goes down and ***WAIT* − *GO* − *GO*** goes up when we probe the agent with lane position information (Example 1).



- **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.
- **Calculating VoI and IoI**
  - Discrete State Space: We construct a new probing augmented POMDP. Solving this augmented POMDP allows ***O***(|***S***|) and ***O***(|***S***||***T***|) calculation of VoI, and IoI respectively for any given input.
  - Continues State Space: We propose Meta-CDQL algorithm, a deep Q-learning-based algorithm for estimating counterfactual information probed values functions.

Case Study





- **AV System**
  - We run a case study on a functional AV prototype in an urban environment supervised by a safety driver.
  - The AV models the world with a set of POMDP. Each POMDP models a particular aspect of the AV's interaction with the world.
  - The Above picture is a simplified information-augmented view of the raw videos taken by the AV's camera. It was created for the developers to enable debugging.
- **VoI Visualization**
  - We put a '?' sign on top of objects corresponding to each feature. We control the opacity based on the normalized magnitude of VoI.
  - The AV is waiting in the intersection for the traffic light to change. When suddenly the traffic light changes state (red to green) uncertainty about the state goes up.
  - Due to the uncertainty the AV chooses to wait even though it could go. Therefore, '?' appears bright on top of the traffic light as having this information will allow the agent to gain more value.

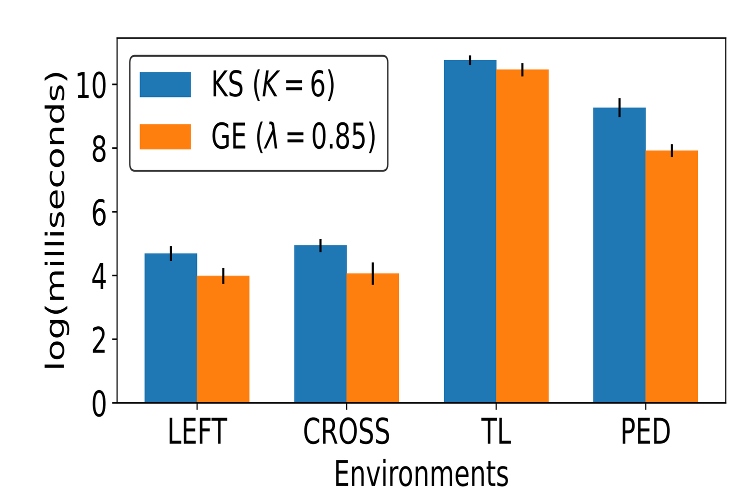
Additional Results and Takeaway Messages

- **Theoretical Analysis**
  - Present several theoretical properties of VoI and IoI, including a direct relation between the two metrics:

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

.

- **Quantitative Analysis**
  - Correlation: We evaluate the similarity among VoI and IoI and different probing mechanisms. We find a high correlation between these methods.
  - Consistence: We analyze the consistency of the generated explanations. We find comparatively high consistency in discrete state space. We propose applying an ensemble as a remedy in continuous state space.
  - Runtime: We compare the computation requirements of different probing strategies. We find that GE strategy requires less computation.



- **Takeaway Messages**
  - SIP sheds light on how uncertainty affects an agent's behavior.
  - SIP is versatile, providing a flexible probing strategy, and can be used to quantify different aspects of change in behavior.
  - Explanations can be generated quickly via efficient pre-computation.
  - Extension to offline RL settings will be interesting.

SCAN ME

University of Massachusetts Amherst

Acknowledgement: Nissan Advanced Technology Center Silicon Valley, NSF Grants 1954782 and 2321786.