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

Introduction

► **What?** We introduce a novel framework called SIP: Sequential Information Probing to explain how uncertainty about state features affect agents behavior.

► **Why?** Most real-world decision-making scenarios contain partially observed features.
  ► While ubiquitous scenario no prior works has attempted to explain this class of problems.
  ► The sequential nature of POMDPs makes it difficult to apply existing feature attribution methods.

► **How?** We probe agent with a sequence of perfect counterfactual information about state features. Then we quantify the effect on agents' behavior.

► **Works?** We both theoretically and quantitatively analyze the properties of our method. We also run a case study on a working AV system for demonstration.

Value and Influence of Information (VoI and IoI)

► **VoI: Value of Information**
  ► 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.

► **IoI: Influence of Information**
  ► 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 agent with lane position information (Example 1).
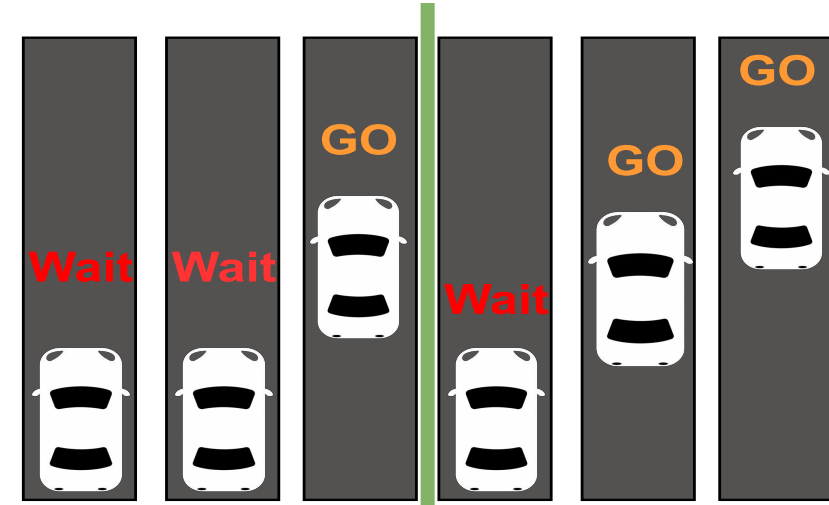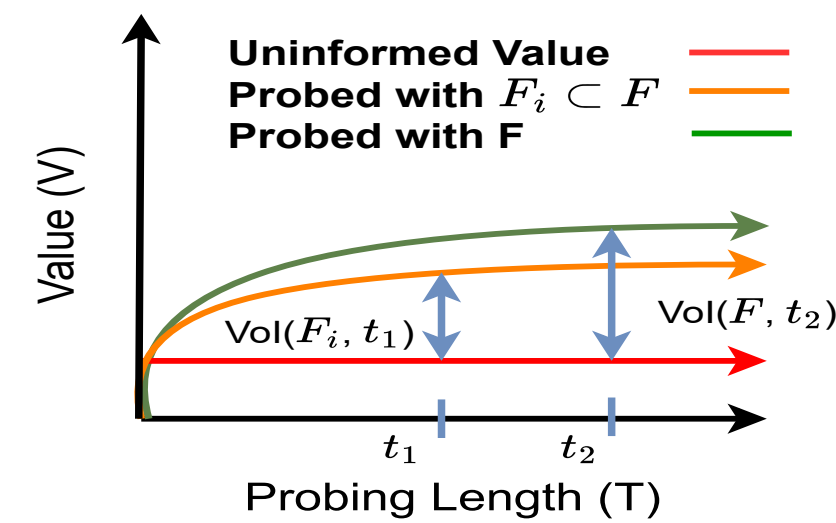
► **Probing Strategy**
  ► **[KS]**: The agent stays exactly K steps in the mirror world.
  ► **[GE]**: The agent stays K steps in expectation in the mirror world where $K \sim Geometric(1 - \lambda)$
  ► **[MY]**: The agent remains in the mirror world for K steps but the agent is not aware of the probing.

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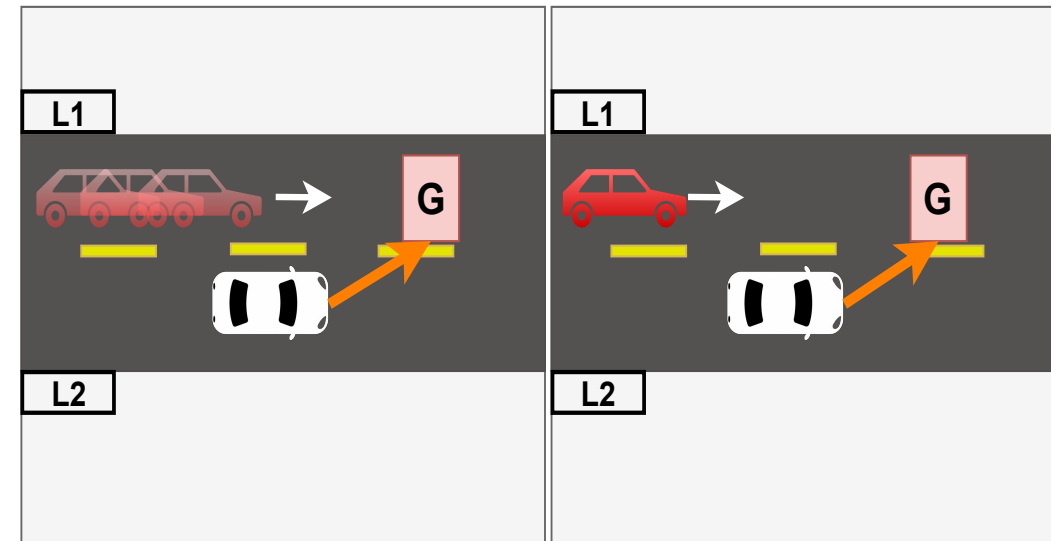
Example

► **One-Step Information**
  ► 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.
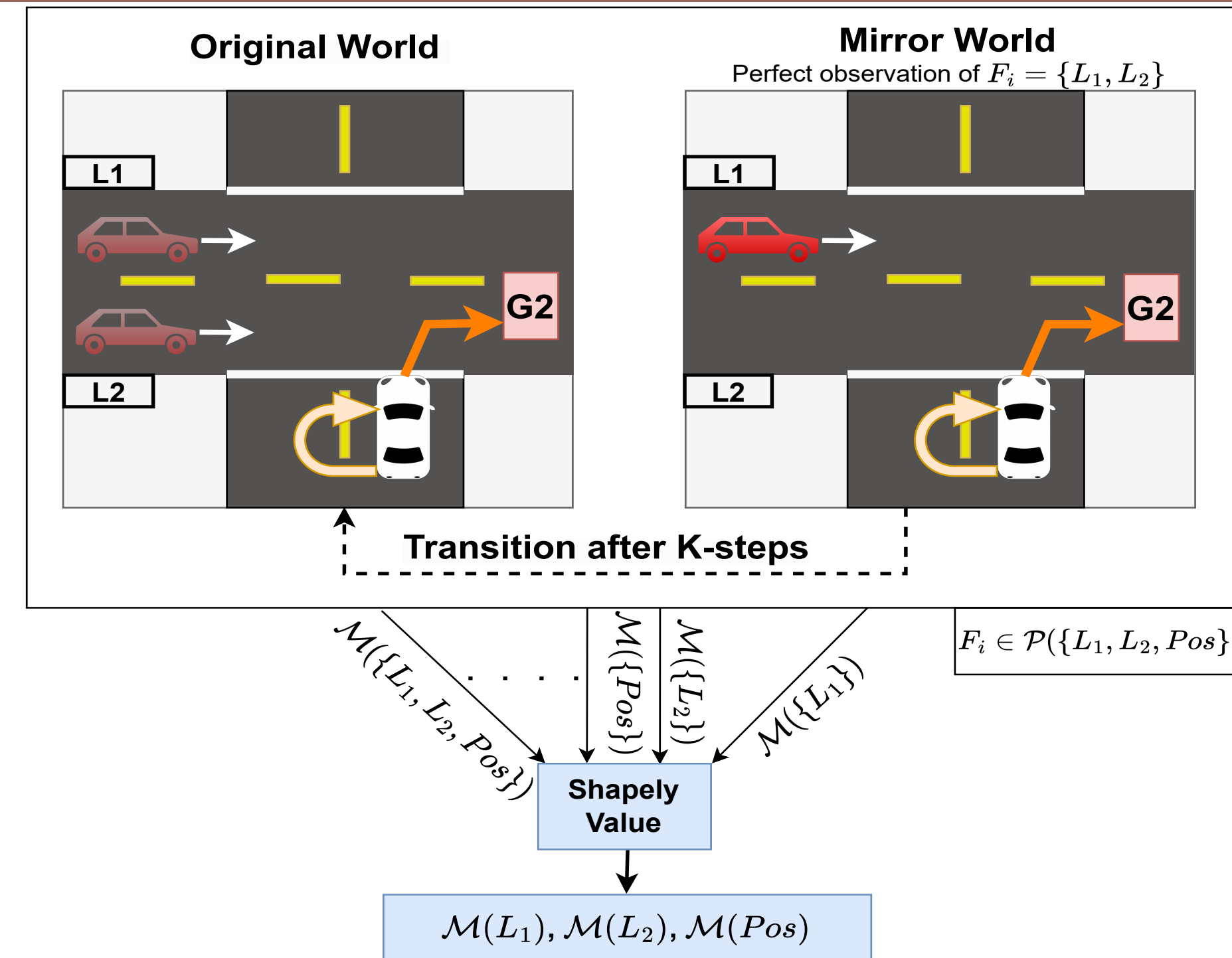 ► If we probe with white care with the lane position information it can avoid waiting.
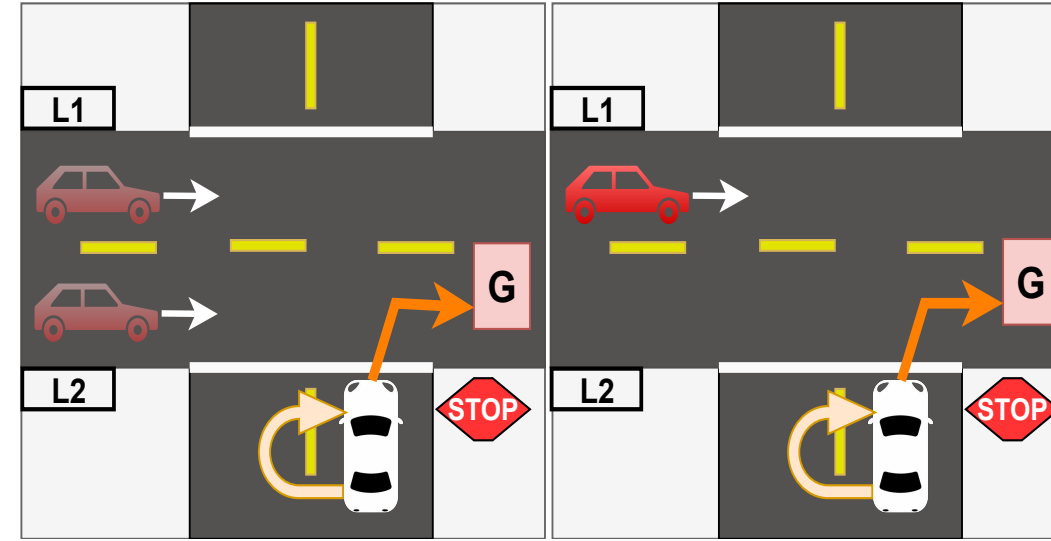
► **Sequential Information**
  ► Estimation of both position and speed needed when changing lane. Further, each step degrades estimation.
  ► Therefore, a single step of information probing might not be sufficient to elicit lane-changing behavior.

SIP: Sequential Information Probing

Original World Mirror World
Perfect observation of $F_i = \{L_1, L_2\}$

Transition after K-steps

$\mathcal{M}(\{L_1, L_2, Pos\}) \cdots \mathcal{M}(\{Pos\}) \mathcal{M}(\{L_2\}) \mathcal{M}(\{L_1\})$

$F_i \in \mathcal{P}(\{L_1, L_2, Pos\})$

Shapely Value

$\mathcal{M}(L_1), \mathcal{M}(L_2), \mathcal{M}(Pos)$

► **Step 1:** Prob POMDP-agent with a sequence of information about different subsets of state features.
► **Step 2:** Calculate metrics such as VoI and IoI by contrasting the probed agents with the uninformed agent.
► **Step 3:** Compute and present the marginal contribution of each feature using Shapley value method.

Case Study

► **Details**
  ► We run a case study on an Autonomous vehicle in an urban environment.
  ► The vehicle models the world with a set of POMDP using the MODIA framework.
► **VoI Visualization**
  ► We put a '?' on object 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.
  ► Sudden change of the traffic light state causes uncertainty and lights up (right).

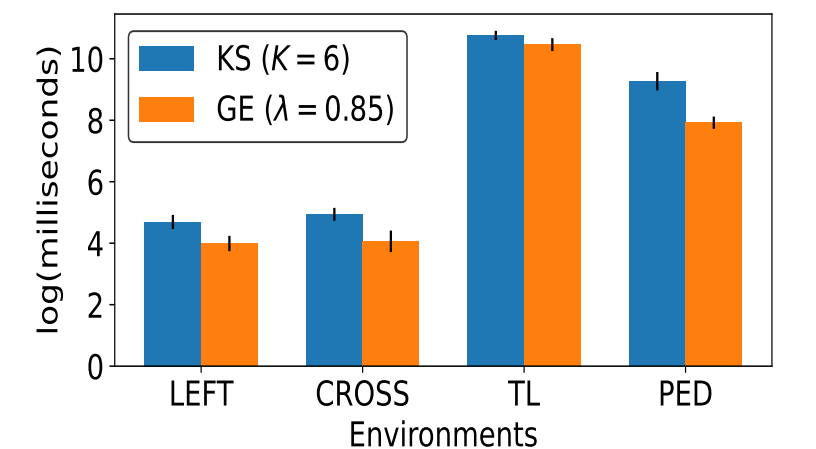## Additional Results and Takeaway Message

► **Theoretical Analysis**
  ► We present several theoretical properties of VoI and IoI including a direct relation between VoI and IoI:

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

► **Quantitative Analysis**
  ► We compare the computation requirements of different probing strategies.
  ► We evaluate the similarity among VoI and IoI and different probing

► **Takeaway Messages**
  ► 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.

mechanisms. We find a high correlation between these methods.
► We analyze the consistency of the generated explanations. We found high consistency in discrete state space. We propose an ensemble as a remedy in continuous sate space.

► In the future, we want to extend our work to offline RL settings.

SCAN ME