Learning to Reason about Contextual Knowledge for Planning under Uncertainty

Cheng Cui[1,2], Saeid Amiri[1], Yan Ding[1], Xingyue Zhan[1], Shiqi Zhang[1]
[1]SUNY Binghamton [2]Cognex Corporation

Published at The 39th Conference on Uncertainty in Artificial Intelligence (UAI), 2023

Motivation

Recently, sequential decision making algorithms have been enabled to reason with contextual knowledge. They have the following limitations:

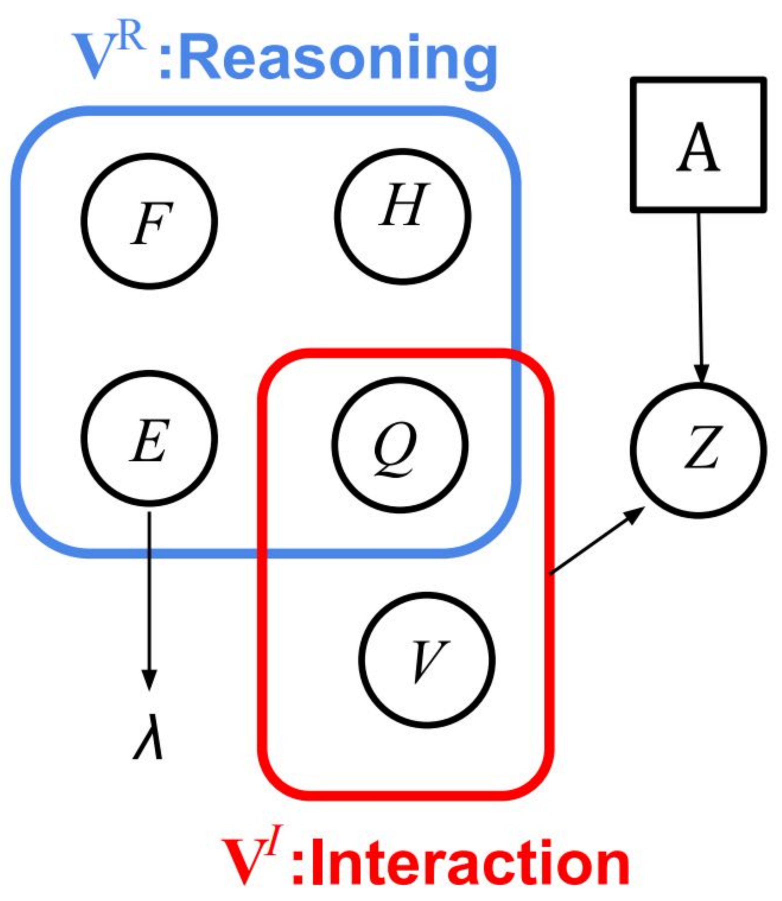
- Contextual knowledge is hardly comprehensive
- Contextual knowledge may include inaccurate information
- Agents need significant efforts to recover from incomplete and inaccurate contextual knowledge

Objectives

We developed a new algorithm called PERIL (perceptual reasoning and interactive learning) to achieve the following goals:

- Learning to improve contextual knowledge base (relational learning)
- Data gathering at runtime (supervised learning)
- Closing the perceive-reason-act loop, which identifies the main contribution of this research

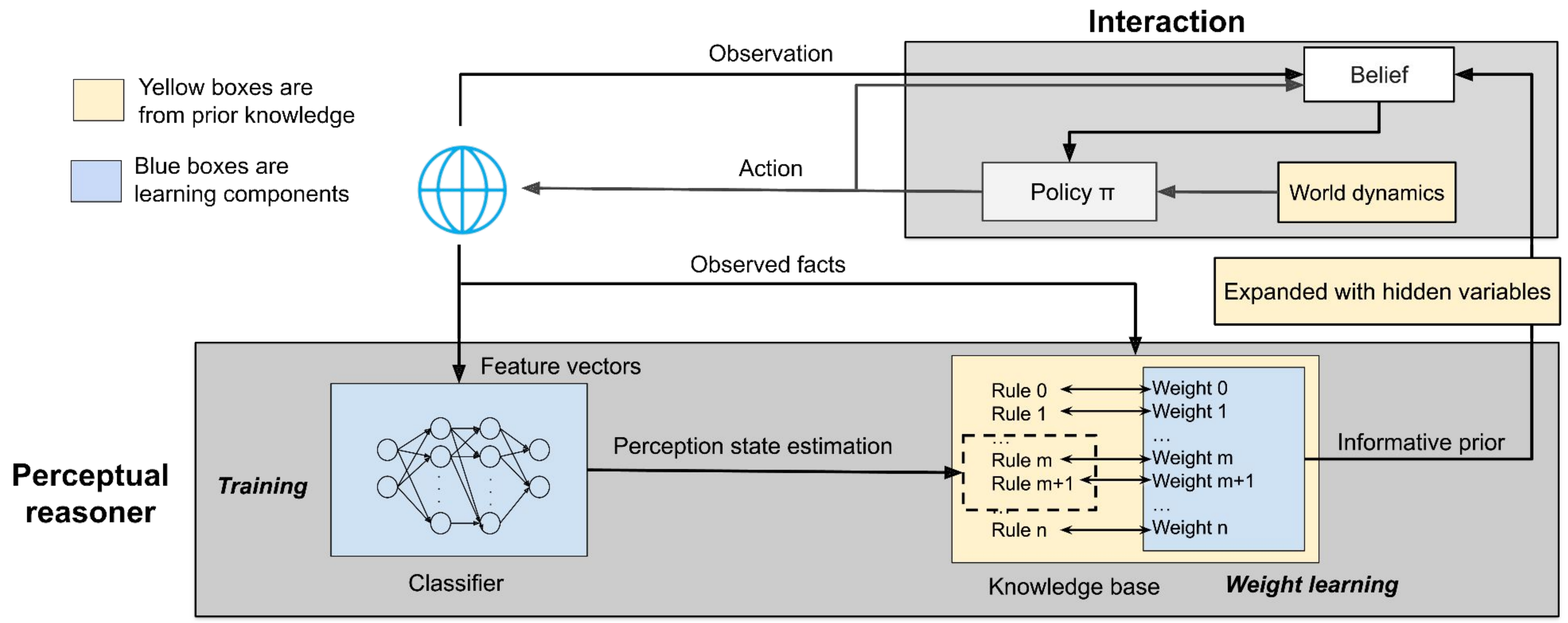
Problem Statement



Reasoning:
- Input:
  - F : fact variables
  - E : evidence variables)
  - H : hidden variables
- Output:
  - Q : query variables

Interaction:
- Input:
  - Q : query variables
  - Z : observation
  - V : unobservable variables
- Output:
  - A: actions

PERIL Overview

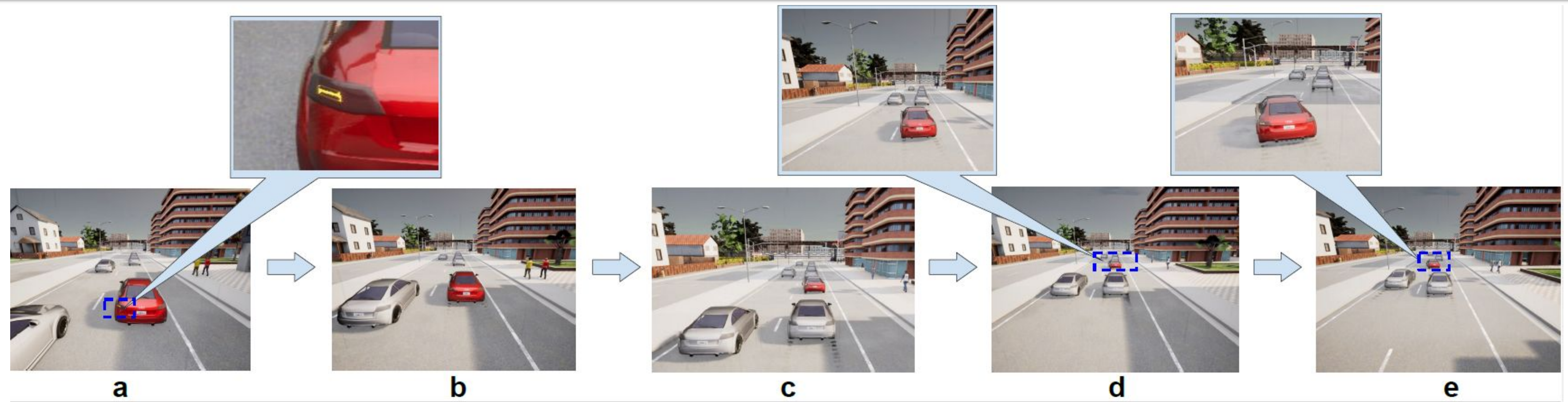


Instantiation

We use following models to instantiate each component of our PERIL framework:

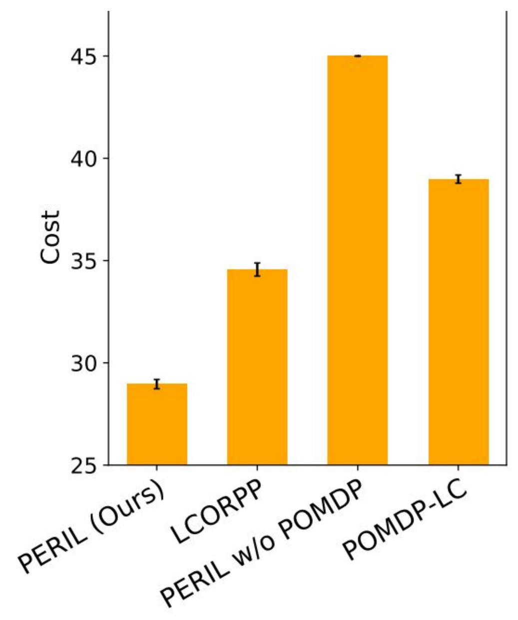
- Convolutional neural network as our classifier:
  - Input: feature vectors
  - Output: classification
- Markov logic network for logical probabilistic reasoning:
  - Input: first order logic rules with weights, value of fact and evidence variables
  - Output: value of query variables
- POMDPs (Partially observable Markov decision process) for probabilistic planning
  - Input: world dynamics, initial belief
  - Output: actions

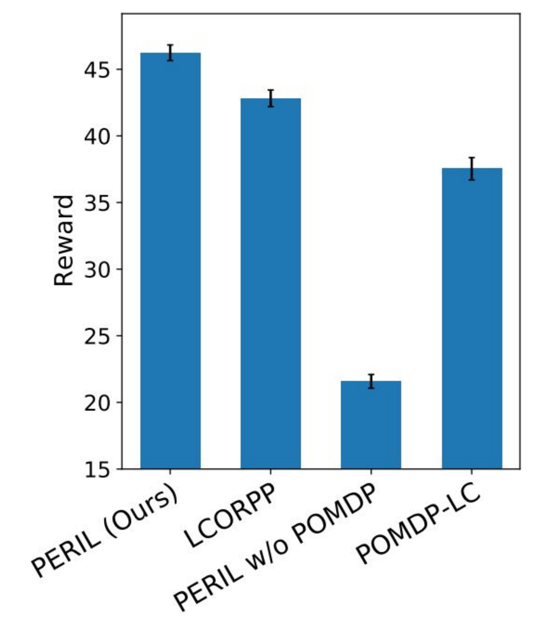
Illustrative Example



The ego vehicle took a sequence of actions in the interaction process to successfully merge left. (a) The ego vehicle intended to merge left. It turned on the left signal. (b) The surrounding vehicle on the left was not cooperative at first. The ego vehicle kept left blinking. (c) The surrounding vehicle on the left became cooperative, and the ego vehicle started to move left. (d) The ego vehicle kept moving left and found room in the left lane. (e) The ego vehicle successfully merged left.

Experimental Results





We compared PERIL with three baseline methods and we see that PERIL achieved the highest cumulative reward on average, and required the lowest interaction cost on average.

- LCORPP (a method that cannot learn to reason about knowledge).
- PERIL w/o POMDP is the same as PERIL except that the action policy is manually crafted.
- POMDP-LC is a classic POMDP-based approach for planning lane changing behaviors.

Conclusion

- We develop an algorithm called PERIL that learns to reason with contextual knowledge for sequential decision making.
- PERIL outperformed competitive baselines at 0.05 significance level, as well as its own ablations, in both overall reward and interaction cost.

Acknowledgement

This work has taken place at the Autonomous Intelligent Robotics (AIR) Group, SUNY Binghamton. AIR research is supported in part by grants from the National Science Foundation (NRI-1925044), Ford Motor Company (URP Award 2019-2022), OPPO (Faculty Research Award 2020), and SUNY Research Foundation.