LEARNING "WHAT-IF" EXPLANATIONS FOR SEQUENTIAL DECISION-MAKING

IE708 course Presentation
Hanish Dhanwalkar - 210100060

Purpose of the Paper:

- Modeling expert behavior based on the idea of "what-if" scenarios, which ask what would happen if different actions were taken in the past.
- The paper introduces a method to explain decision-making in complex environments by learning from the actions of experts.
- The method creates interpretable reward functions that explain why experts make specific choices.

Applications

- In areas like healthcare, experts (e.g., doctors) make critical decisions based on patients' past medical history, but experimenting on patients is impossible.
- CIRL helps understand how these decisions are made without needing new experiments by creating interpretable reward functions that explain why experts make specific choices.
- CIRL is able to explain the trade-offs experts make, such as prioritizing treatment effectiveness over potential side effects.

Overview

1. 'What-If' Explanations
2. Counterfactual Reasoning in Decision-Making
3. Inverse Reinforcement Learning (IRL)
4. What-If Reward Parameterization
5. Counterfactual Inverse Reinforcement Learning (CIRL)

Sequential Decision-Making

- Involves a series of actions over time, aiming to optimize some long-term objective.
- Example used in the paper: A doctor making treatment decisions for a patient over multiple visits.
- Key Challenge: Understanding the rationale behind expert decisions, especially in real-world environments where active experimentation is impossible.

Focus of the Paper: Explaining expert behavior using counterfactual reasoning and learning from past data in offline or batch settings.

What is and Why 'What-If' Explanations?

- Aim: Explain expert behavior based on observed trajectories.
- Idea revolves around predicting the outcome if different possible actions was taken.
- Given an action a_t at time t, a 'what-if' explanation estimates what would have happened if an alternative action a_t' had been taken.
- Example: In healthcare, we might ask, "What would have happened if the doctor had prescribed a different medication?"

Counterfactual Reasoning in Decision-Making

- Counterfactual reasoning involves estimating the potential outcomes for actions that weren't taken
- In decision-making, this helps evaluate the unobserved effects of alternative actions.
- Mathematically,
  - Let Y_t(a_t) denote the outcome of action a_t at time t.
  - For each action ata_tat, the goal is to estimate the counterfactual outcome E[Y_t+1[a_t]∣h_t], where h_t is the history up to time t.
- In medical decision-making, understanding how alternative treatments (actions) could have affected patient outcomes helps explain why certain treatments were preferred

Inverse Reinforcement Learning (IRL)

- **Standard RL:** In Reinforcement Learning, an agent learns a policy π to maximize a reward function $R(s_t,a_t)$
- IRL: Instead of learning a policy, IRL works backwards—it observes the actions of an expert and tries to recover the reward function that the expert was trying to optimize.
- Mathematically,
  - In IRL, given expert trajectories $D=\{(s_t,a_t)\}$, we aim to learn the expert's reward function $R(h_t,a_t)$
  - The goal is to minimize the difference between the expert's behavior and the learned policy.
  - Why IRL?: In settings like healthcare, it's essential to understand why an expert (e.g., a doctor) chooses a particular treatment path, based on implicit trade-offs they make

'What-If' Reward Parameterization

- Parameterizing rewards based on **counterfactual outcomes** helps explain how experts make decisions.
- Mathematically,
  - Suppose we have two outcomes for a patient: **disease progression** U and **side effects** Z.
  - The expert's decision is based on the trade-off between these outcomes. The reward function is modeled as a weighted sum of counterfactuals:
    - $R(h_t,a_t)=w_U E[U_{(t+1)}[a_t]|h_t]+w_Z E[Z_{(t+1)}[a_t]|h_t]$
    - Where w_U and w_Z represent the expert's relative preference weights for reducing disease progression vs. minimizing side effects
  - Interpretation:
    - If $|w_U|>|w_Z|$, it indicates the expert is treating **aggressively**, prioritizing disease reduction over side effects.
    - If $|w_U|<|w_Z|$, it indicates the expert is more **conservative**, prioritizing minimizing side effects

Counterfactual Inverse Reinforcement Learning (CIRL)

- CIRL integrates **counterfactual reasoning** into **batch IRL**, making it suitable for environments where the expert's behavior depends on the **history** of actions and observations, and active experimentation is impossible.
- CIRL learns a reward function that explains expert behavior in terms of 'what-if' scenarios—what would have happened if different actions had been taken at various points in the history

Counterfactual Inverse Reinforcement Learning (CIRL)

Method Overview:

- Input: A batch dataset $D=\{(h_t,a_t,x_t)\}$ of expert trajectories (observations, actions, outcomes).
- **Counterfactual Estimation**: Estimate the potential outcomes $E[Y_{(t+1)}[a_t] \mid h_t]$ for all possible actions.
- **Reward Function:** weighted sum of these counterfactual outcomes:
  - $R(h_t, a_t) = w \cdot E[Y_{(t+1)}[a_t] | h_t]$
- **Max-Margin IRL**: Learn the reward weights w using max-margin inverse reinforcement learning, minimizing the distance between expert and learned policies.
- **Policy Learning**: Use **deep recurrent Q-learning** to find the optimal policy under the learned reward function.
- **Outcome**: CIRL outputs a reward function that explains expert preferences over **counterfactual outcomes**, providing an interpretable decision model.

Comparison with Other Methods

- **CIRL** was compared with other batch IRL methods such as MB-IRL (model-based) and DSFN (successor feature networks).

Advantages:

- Accuracy: CIRL outperformed other methods in matching expert actions.
- Interpretability: Unlike other methods, CIRL provides interpretable reward functions that explain expert decisions based on counterfactuals

Thank You