EVERCOT

Bachelor Thesis

Evercot AI

ask@evercot.ai

Title: Agent Actions Evaluation and Policy Optimization using Causal Reinforcement Learning

Introduction:

The traditional Reinforcement Learning (RL) problem is formalized using Markov Decision Process (MDP). On the other hand, theories from causal inference is beneficial in obtaining a better understanding of the behavior and characteristics of a system or object. A combination of causal inference and reinforcement learning has shown to enhance traditional RL results in a research area known as Causal Reinforcement Learning (CRL).

Specifically, causal reinforcement learning (CRL) is also formalized using MDP as follows. Given an environment and an agent, the environment selects a state at time *t.* Based on this state, an agent uses theories in causal inference to decide which action to *intervene* with at time *t*. After an action has been chosen by the agent, the environment in turns awards a reward depending on the causal action selected by the agent at time *t*. In addition, the environment samples a next state at time *t+1*. The agent intervenes again with a new causal action at time *t+1* and receives a reward and next state at time *t+2* as a response from the environment. This loop continuous until a terminal state at time *t+n* is reached. CRL comes in two forms. Namely, CRL when causal information is known beforehand. The second form is CRL with unknown causal information.

Thesis Objectives:

In traditional RL, one of the prime goals is to find an optimal policy for a task or system. The usage of known (or unknown) contextual causal information about a task in CRL can be depicted as a causal graph structure. Such causal graphs can be exploited by an agent within its environment to build better policy to improve the performance of traditional RL.

Off-policy learning is a problem in RL. Causal inference can be explored to detect and evaluate the effect of actions taken by the agent. Such evaluation can be used in off-policy learning to improve the policy of RL in different RL environment.

The main objective of this thesis is to investigate and propose a new causal RL technique with known causal information that can be utilized to:

1. Compute the influence of an agent's actions
2. Derive a causal policy that performs better that traditional RL policies
3. Utilize the new causal policy in several RL tasks to demonstrate that it delivers better learning results

Contacts

Are you interested or need more Information?

Contact us at ask@evercot.ai

Evercot AI GmbH
Landsberger Str. 302
80687 Munich

ask@evercot.ai