Assignment-15

Dhanush V Nayak
ee23btech11015@iith.ac.in

1 Introduction

The core idea of learning through interaction with the environment is highlighted in this chapter. The concept revolves around how infants and humans gain knowledge by exploring their surroundings. This idea serves as the foundation for reinforcement learning (RL), a computational approach that focuses on goal-directed learning through trial and error. Rather than directly theorizing about how people or animals learn, we primarily explore various learning situations and evaluate the effectiveness of various learning methods.

2 Reinforcement Learning (RL)

Reinforcement Learning (RL) is about learning how to map situations to actions to maximize a reward signal. The agent is not told which actions to take but must discover the best actions through trial and error. Two critical aspects that differentiate RL are:

- **Trial-and-error search**
- **Delayed reward**

In RL, actions can influence not only immediate rewards but also future states and rewards. This means the agent must learn to balance short-term and long-term outcomes. RL differs from supervised learning, where a system is trained on labeled examples with the correct actions provided. In RL, the agent must figure out the best actions on its own through interaction with the environment. It also differs from unsupervised learning, which is about finding patterns in data. RL focuses on maximizing a reward signal instead of discovering hidden structures.

A key challenge in RL is the exploration-exploitation trade-off, which is explained in the next subsection. In summary, RL is a powerful framework for learning from interaction, focusing on maximizing cumulative rewards over time.

2.1 The Exploration-Exploitation Trade-off

A key dilemma in RL is the trade-off between exploration and exploitation. The agent must balance trying new actions (exploration) to discover their rewards and choosing actions that have already proven successful (exploitation). If an agent only exploits known actions, it might miss out on better strategies that could yield more rewards. Conversely, too much exploration without exploitation can lead to poor performance. Mathematically, this is an unresolved challenge in reinforcement learning theory.

3 Examples of Reinforcement Learning

Reinforcement Learning (RL) can be understood better through practical examples where agents interact with their environment to achieve goals. These examples illustrate various real-world applications of RL:

- **Chess Player:** A master chess player makes decisions informed by both planning and intuition, balancing immediate judgments with long-term consequences.
- **Adaptive Controller:** A controller optimizes a petroleum refinery's operation by adjusting parameters in real time, focusing on cost, yield, and quality trade-offs.
- **Gazelle Calf:** Minutes after birth, a gazelle calf struggles to stand and within a short time learns to run at high speed, adapting through experience.
- **Mobile Robot:** A robot must decide whether to continue searching for trash or return to its charging station based on its current battery level and past experiences.
- **Phil's Breakfast:** Preparing breakfast involves a complex sequence of interrelated actions, such as fetching items, and reacting to the environment and at the same time watching everything goes right in proportion . (preventing over-pouring of milk.)

These examples demonstrate key RL features such as interaction with the environment, decision-making, and the need for foresight, as the agent's actions affect not only the immediate outcome but also future opportunities and states.

4 Elements of Reinforcement Learning

Four key elements define RL systems:

1. **Policy**: A policy is the strategy the agent follows to choose actions based on the current state.
2. **Reward signal**: This defines the immediate feedback the agent receives after taking an action, indicating the success of that action.
3. **Value function**: This estimates the long-term success (future rewards) of being in a specific state. It helps the agent decide actions not just based on immediate rewards but the overall expected outcome.
4. **Model of the environment**: This helps the agent predict the next state and reward based on the current state and action, which can be used for planning.

5 Limitations and Scope of RL

A significant limitation of RL is its reliance on the state signal, which the agent uses to make decisions. While this approach is effective, the complexity of real-world environments often leads to challenges in state representation. Most RL methods estimate value functions to make decisions, but alternative methods such as evolutionary algorithms can also be used.

6 An Extended Example: Tic-Tac-Toe

Tic-Tac-Toe is a great example to illustrate reinforcement learning (RL). The goal is to learn the best strategy to win against an imperfect opponent. The RL agent starts with no knowledge of the game and learns by playing many games.

Each game state is represented by Xs and Os on a 3x3 grid. The agent assigns a value to each state, indicating the chance of winning from that state. Winning states (three Xs in a row) have a value of 1, losing or drawn states have a value of 0, while other states are initialized at 0.5, indicating a 50% chance of winning.

During gameplay, the agent explores different actions, primarily choosing the one that leads to the highest-valued state, but it occasionally makes random moves to discover new strategies. After each game, the agent updates its value function using temporal-difference learning, where the value of the current state gets closer to the value of the next state. The update rule is:

$$V(S_t) \leftarrow V(S_t) + \alpha \left(V(S_{t+1}) - V(S_t) \right)$$

where α is the learning rate, and $V(S_t)$ is the value of the current state.

This process allows the agent to improve its strategy over time. As it plays more games, it better predicts the value of each move and converges to an optimal strategy, highlighting key RL aspects like learning through interaction and balancing exploration with exploitation.