Utility Representations

Hulingxiao He
Wangxuan Institute of Computer Technology
Peking University
hehulingxiao@stu.pku.edu.cn

Abstract

The concept of utility is commonly employed to quantify worth and value, closely tied to human preferences and capable of offering insights into human behaviors. Utility functions have been crafted by researchers to articulate decision-making challenges, particularly within the domain of reinforcement learning. This essay aims to explore the distinctions between utility functions and reward functions, delving into the representation of utility in the context of reinforcement learning. Finally, it emphasizes the future representation of utility through curiosity-driven approaches and the like.

1 Introduction

Utility functions serve as abstract representations of human utility, unveiling preferences and satisfaction. Despite their generalized design, they play a crucial role in the decision-making process, providing an internal estimation of human value for choices. Originally devised to model people's decisions, utility uniquely captures the expected value derived from individuals' preferences rather than the actual observed value in the physical world. This inherent property introduces challenges in measuring human utility, compounded by its subjective nature, varying for each individual. For instance, someone with a passion for outdoor activities may assign higher utility to a hiking trip than spending a day indoors, while a person who enjoys a quiet day at home might have the opposite preference. This diversity in preferences raises questions about how AI's utility should align with human values, adding complexity to the value alignment problem and highlighting the inherent difficulties in establishing a universal representation of human utility.

2 Reward and Utility Functions

The reward function serves to encapsulate the immediate outcomes of specific behaviors, whereas utility functions provide a more enduring representation. To clarify, the utility function represents the anticipated weighted sum of both immediate and long-term rewards, considering the optimal policy.

$$\mathrm{U}\left(s_t, a\right) = E\left\{\mathrm{R}\left(s_t, a\right) + \max_{\mathrm{P}} \sum_{i=1}^{N-1} \mathrm{R}_{t+i}\right\} \tag{1}$$

where U, R, P represent utility, reward and policy, respectively. In accordance with Bellman equations, the following equation expresses the utility in the present state in relation to the utility of the subsequent state.

$$\mathrm{U}\left(s_t, a\right) = E\left\{\mathrm{R}\left(s_t, a\right) + \max_{b} \mathrm{U}\left(s_{t+1,b}\right)\right\} \tag{2}$$

With a simple linear updating transition function, we can update the utility function.

$$\mathrm{U}\left(s_t, a\right) = \left(1 - \alpha\right) \mathrm{U}_t\left(s_t, a_t\right) + \alpha\left(\mathrm{R}\left(s_t, a_t\right) + \max_{b} \mathrm{U}_t\left(s_{t+1,b}\right)\right) \tag{3}$$

where α is a hyper-parameter for momentum update.

3 Utility in Reinforcement Learning

Reinforcement learning (RL) constitutes a subfield of artificial intelligence dedicated to acquiring knowledge through interactions with the environment. Within RL, utility values serve as a means to depict the desirability associated with specific states or actions. In its simplest manifestation, an agent receives a reward signal for executing particular actions within a given environment. Subsequently, this reward signal guides adjustments to the agent's policies, aimed at optimizing future rewards.

Different from reward, utility functions offer a broader representation of desirability. For instance, in the game of Super Mario, the utility function considers several factors to evaluate utilities, including:

- Score: The cumulative points earned during gameplay.
- Remaining Lives: The number of lives Mario has, influencing risk management.
- Time Remaining: The time left to finish the level, encouraging efficiency.
- Coins Collected: The total number of coins acquired during the level.
- Progress in the Level: Mario's position and advancement within the level.
- Enemies Defeated: The count of defeated enemies, impacting both score and safety.
- Obstacles Avoided: Successful navigation and avoidance of hazards.
- Level Completion: Whether Mario successfully completes the level.

These components collectively contribute to the utility function, shaping the agent's decision-making in Super Mario. The key distinction between utility function and reward function lies in their focus and purpose. The utility function represents a more comprehensive, long-term measure of desirability, considering multiple factors, while the reward function typically provides an immediate feedback signal to reinforce specific actions during learning. The utility function guides the agent's overall strategy, encompassing a broader perspective on desirability beyond immediate rewards.

In the context of AlphaGo [4, 5], a system devised by DeepMind for the game of Go, explicit utility functions implemented through neural networks play a pivotal role. These utility functions assess each position on the board, informing strategic move decisions.

Crucially, the utilization of utility functions in reinforcement learning extends beyond the realm of gaming, demonstrating significance in diverse domains such as robotic control and autonomous driving. Through encapsulating a more comprehensive understanding of desirability, utility functions facilitate the development of policies better attuned to the intricacies of the environment [6].

4 Future Direction

Most studies learn to represent human utility by the prior knowledge, human feedback, or interaction with environments. However, utility is merely studied from intrinsic perspective. For instance, recent approaches in deep reinforcement learning are addressing the challenge of enabling machines to mimic the exploratory play processes observed in children or to exhibit curiosity-driven behavior. This is achieved through innovative goal and reward design strategies. The underlying idea is to develop reinforcement learning agents that are not heavily reliant on external rewards. For instance, a surprise incentive, quantified as the KL-divergence between true Markov Decision Process (MDP) transition probabilities and the learned ones is introduced [1]. Employing this surprise incentive for efficient exploration, their model demonstrated exceptional performance in continuous control tasks. Subsequent research has continued to explore the prediction error paradigm, treating curiosity as the discrepancy in an agent's ability to predict the outcomes of its actions, as learned by an inverse dynamics model [3]. Scaling up the application of pure intrinsic curiosity-based rewards has shown promise in diverse environments, including Atari games [2].

This line of research validates the feasibility of reinforcement learning without relying on extrinsic rewards, offering support for the embodiment hypothesis. Nevertheless, debates persist regarding its limitations. For example, there is speculation that certain games may be intentionally designed to promote curiosity-driven exploration. The question of whether representing utility functions from intrinsic rewards can generalize to a broader spectrum of tasks is still under investigation.

References

[1] Joshua Achiam and Shankar Sastry. Surprise-based intrinsic motivation for deep reinforcement learning. *arXiv preprint arXiv:1703.01732*, 2017. 2

[2] Yuri Burda, Harri Edwards, Deepak Pathak, Amos Storkey, Trevor Darrell, and Alexei A Efros. Large-scale study of curiosity-driven learning. *arXiv preprint arXiv:1808.04355*, 2018. 2

[3] Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *International conference on machine learning*, pages 2778–2787. PMLR, 2017. 2

[4] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016. 2

[5] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017. 2

[6] David Silver, Satinder Singh, Doina Precup, and Richard S Sutton. Reward is enough. *Artificial Intelligence*, 299:103535, 2021. 2