How to Represent and Learn Human Utility

Zhuoying Li
Yuanpei College
Peking University
joy@stu.pku.edu.cn

Abstract

Utility is a frequently used concept to model worth and value. t is intrinsically linked with human preferences, thereby offering a potential explanation for human behavior. Many efforts across various fields have been made to modeling and learning about utility. In this essay, we will first introduce Expected Utility Theory (EUT), an important theory for understanding and modeling preferences and decision-making in economics. Then we will discuss some computational methods in RL to represent and learn human utility. Finally, we concludes that modeling utility can be a promising way towards AGI.

1 Introduction

The concept of utility represents a foundational element in economic theory, encapsulating the satisfaction or value derived from the consumption of goods or experiences. Our preferences are an outward manifestation of this utility. Distinct from the concept of *value*, which holds a more objective stance, utility is inherently subjective and varies depending on context and circumstances.[6].

The relevance of utility extends well beyond economic theory, playing an important role in the realm of artificial intelligence, particularly reinforcement learning (RL). Traditional RL algorithms integrate utility into their architecture by defining reward functions intended to guide agents towards optimal behaviors. However, this approach depends on the prior knowledge to shape reward functions. Therefore, we proceed to explore Preference-Based Reinforcement Learning (PbRL), a paradigm that introduces a more nuanced understanding of utility by prioritizing subjective preferences over fixed reward structures. Then in the final section of this essay, we argue that studying human utility and embedding it within AI systems might be a promising way to reach AGI.

2 Expected Utility Theory (EUT)

EUT is the basis of many methods to learn and represent utility. Therefore, before delving into computational methods for modeling utility, we will first introduce this theory from economy. EUT offers a framework for making rational choices under uncertainty, guiding individuals towards actions that maximize their potential benefits. The guiding principle of this theory is straightforward: select the action that yields the highest expected utility.

2.1 Definition

In a decision problem, there are three kinds of important entities: the outcomes o, the possible states s, and the acts a. Expected utility theory provides a way of ranking the acts according to how choice-worthy they are. Given these three pieces of information, the expected utility of action a is defined as:

$$EU(A) = \sum_{o \in O} P_A(o)U(o)$$

where O is is the set of outcomes, $P_A(o)$ is the probability of outcome o conditional on A, and $U(o)$ is the utility of o.

2.2 Why EUT?

Then there is the question: why it is best to choose acts that maximize expected utility? While this concept may appear intuitive, it still requires a degree of mathematical proof or theoretical derivation. In this discussion, we will present two key arguments: the Long-Run Arguments and the Representation Theorems.

2.3 Long-Run Arguments

The *strong and weak laws of large numbers* indicate that for sequences of independent, identically distributed trials, over the long run, the average amount of utility gained per trial is overwhelmingly likely to be close to the expected value of an individual trial. Therefore it's best to choose action that maximize expected utility in the long-run. However, it's important to note that these laws presuppose the independence of trials. For scenarios where trials are not mutually independent, these laws are not applicable. Additionally, many decisions do not lend themselves to repetition over a series of similar trials (e.g. decisions about whom to marry). The relevance of long-term considerations, which apply to repeated gambles, to these one-off choices is not immediately evident.

2.4 Representation Theorems

A second type of argument for expected utility theory is based on Representation Theorems. According to [1], the argument has three premises:

- **The Rationality Condition:** The axioms of expected utility theory are the axioms of rational preference.
- **Representability:** If a person's preferences obey the axioms of expected utility theory, then she can be represented as having degrees of belief that obey the laws of the probability calculus and a utility function such that she prefers acts with higher expected utility.
- **The Reality Condition:** If a person can be represented as having degrees of belief that obey the probability calculus and a utility function such that she prefers acts with higher expected utility, then the person really has degrees of belief that obey the laws of the probability calculus and really does prefer acts with higher expected utility.

From these premises, we can entail the following conclusion:

If a person fails to prefer acts with higher expected utility, then that person violates at least one of the axioms of rational preference.

The key premise here is the representability, and mathematical proofs of representability are called *representation theorems*. There are several influential representation theorems in the field of economics and decision theory. Each theorem is underpinned by its own set of axioms and differs in the permissible transformations of probability and utility functions. Take for example the Von Neumann-Morgenstern (VNM) utility theorem [4]. In this theorem, preferences are defined over a domain of *lotteries*, which are essentially scenarios where each potential outcome occurs with a certain probability, all of which collectively add up to one. The VNM utility theorem includes four axioms and it's proved that every preference relation obeying these axioms can be represented by the probabilities used to define the lotteries, together with a utility function which is unique up to positive linear transformation.

3 Utility in RL

3.1 Traditional RL

In traditional RL, typically there is a value function V which denotes the expected return starting with state s_t given the policy π:

$$V_\pi(s) = E[R|s_t]$$

Based on EUT, the agent's goal is to maximize the value function:

$$V^*(s_t) = \max_\pi V_\pi(s_t) = \max_\pi E[r_t(s_t, a_t) + \sum_{j=t+1}^{\infty} \gamma^j r_j]$$

According to Bellman equations, we then have:

$$V^*(s_t) = \max_{\pi} V_\pi(s_t) = \max_{\pi} E[r_t(s_t, a_t) + \sum_{j=t+1}^{\infty} V_\pi(s_j)]$$

3.2 Preference-Based Reinforcement Learning (PbRL)

RL has achieved remarkable success and made significant advancements in various domains, including robotics and gaming [2]. However, the effectiveness of traditional RL often heavily depends on the prior knowledge put into the definition of the reward function, which requires a high amount of reward-engineering and faces the problems of reward shaping and reward hacking.

To address this problem, Preference-based reinforcement learning (PbRL) has been proposed to learn from non-numerical feedback in sequential domains. Its key idea is to replace the numerical feedback signal with a preference-based feedback signal that indicates relative instead of absolute utility values. As shown in Fig. 1, the learning usually starts with a set of trajectories, either predefined or sampled from a given policy. An expert evaluates one or more trajectory pairs (τ_{i1}, τ_{i2}) and indicates his/her preference.

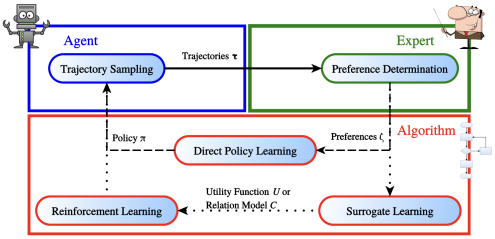

Figure 1: Learning policies from preferences via direct (dashed path) and surrogate-based (dotted path) approaches. Surrogate-based (dotted path) approaches mainly includes learning a preference model and learning a utility function [5].

Learning a utility function is a surrogate-based method to compute the policy π. In this method, we compute the utility difference between the trajectories made by the agent and the preferences given by the human to formulate a loss function. Fig. 2 is a summary of commonly used loss functions for linear utility functions.

Name	Loss Equation	Reference
ranking SVM	$L^{\succ}(\boldsymbol{\theta}, \zeta_i) = \max(0, 1 - d(\boldsymbol{\theta}, \zeta_i))$	Akrour et al. (2011, 2012) Zucker et al. (2010) Wirth and Fürnkranz (2012, 2015) Runarsson and Lucas (2012, 2014)
IRL	$L^{\succ}(\boldsymbol{\theta}, \zeta_i) = c(d(\boldsymbol{\theta}, \zeta_i))$ $c(x) = \begin{cases} -2x & \textbf{if } x \leq 0 \\ -x & \textbf{otherwise} \end{cases}$	Wirth et al. (2016)
cumulative	$p_{\boldsymbol{\theta}}(\zeta_i) = \Phi\left(\frac{d(\boldsymbol{\theta}, \zeta_i)}{\sqrt{(2)}\sigma_p}\right)$	Kupcsik et al. (2015)
sigmoid	$p_{\text{sig}\,\boldsymbol{\theta}}(\zeta_i) = \frac{1}{1+\exp(-m \cdot d(\boldsymbol{\theta}, \zeta_i))}$	Christiano et al. (2017)
0/1	$p_{01\,\boldsymbol{\theta}}(\zeta_i) = \mathbb{I}(d(\boldsymbol{\theta}, \zeta_i) < 0)$	Sugiyama et al. (2012)
combined	$p_{\boldsymbol{\theta}}(\zeta_i) = \frac{\lvert\zeta\rvert - 1}{\lvert\zeta\rvert} p_{01\,\boldsymbol{\theta}}(\zeta_i) + \frac{1}{\lvert\zeta\rvert} p_{\text{sig}\,\boldsymbol{\theta}}(\zeta_i)$	Wirth et al. (2016)
integrated piecewise	$p_{\boldsymbol{\theta}}(\zeta_i) = \int_0^{\epsilon_{\max}} c(d(\boldsymbol{\theta}, \zeta_i), \epsilon)$ $c(x, \epsilon) = \begin{cases} 0 & \textbf{if } x < -\epsilon \\ 1 & \textbf{if } x > \epsilon \\ \frac{\epsilon + x}{2\epsilon} & \textbf{else,} \end{cases}$	Akrour et al. (2013, 2014)

Figure 2: A summary of loss functions for linear utility functions [5].

PbRL excels in learning from non-numeric rewards and reduces the need for prior knowledge to craft the reward function. However, PbRL still faces many challenges. For instance, utility-based methods within PbRL are currently incapable of generating a collection of Pareto-optimal policies in case of incomparabilities. Besides, PbRL has not yet established a standardized framework for evaluation. Often, the test environments adapted from traditional RL problems, and different publications have their own specific setups, modifying the basic configuration by adding noise or removing the terminal conditions [5].

4 Discussions and future directions

Utility is a complicated concept in the realm of human psychology, and the journey to accurately model it remains challenging. Nevertheless, the pursuit of modeling utility is worthwhile as it might be a potential pathway to achieving AGI. Currently, we meticulously design loss functions to enhance a model's ability to address issues within specific domains. But perhaps utility is just enough. Typically, we humans have goals (e.g. get a good grade, find a girlfriend, etc.) and we engage in complex actions to accomplish these aims, which exhibits a wide variety of abilities associated with intelligence. Silver et al. [3] shows that the hypothesis of the maximisation of total reward associated with goals could provide a basis for understanding many daily abilities. In a similar vein, forging AI systems that are goal-driven, capable of forming a concept of utility, and tailoring their behaviors to maximize this utility might be a promising approach towards AGI.

References

[1] R. A. Briggs. Normative Theories of Rational Choice: Expected Utility. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2023 edition, 2023. 2

[2] Jens Kober, J Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11):1238–1274, 2013. 3

[3] David Silver, Satinder Singh, Doina Precup, and Richard S Sutton. Reward is enough. *Artificial Intelligence*, 299:103535, 2021. 4

[4] John von Neumann, Oskar Morgenstern, and Ariel Rubinstein. *Theory of Games and Economic Behavior (60th Anniversary Commemorative Edition)*. Princeton University Press, 1944. ISBN 9780691130613. URL http://www.jstor.org/stable/j.ctt1r2gkx. 2

[5] Christian Wirth, Riad Akrour, Gerhard Neumann, and Johannes Fürnkranz. A survey of preference-based reinforcement learning methods. *Journal of Machine Learning Research*, 18(136):1–46, 2017. URL http://jmlr.org/papers/v18/16-634.html. 3

[6] Yixin Zhu, Tao Gao, Lifeng Fan, Siyuan Huang, Mark Edmonds, Hangxin Liu, Feng Gao, Chi Zhang, Siyuan Qi, Ying Nian Wu, et al. Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense. *Engineering*, 6(3):310–345, 2020. 1