Learning and Representing Human Utility

Rui Yang
Yuanpei College
Peking University
ypyangrui@pku.edu.cn

Abstract

This essay investigates the critical role of utility both in the process of human decision-making and artificial intelligence system design. We start by examining the studies in behavioural psychology, which demonstrate an early understanding of utility, and list the discussion and applications from philosophy, economy, and sociology. The essay then explores Preference-based Reinforcement Learning (PbRL) as a key computational modeling method of utility. Details of PbRL are carefully discussed, as well as strengths and challenges in existing and future research.

1 Introduction

We make thousands of decisions every day. Some of them are made without second thoughts, while others require careful considerations. Some decisions are made on a daily basis and will only have minor influences, such as the choice for lunch. Others could be life-changing, and only occur for very few times throughout the entire lifetime. In a word, decisions shape our lives profoundly.

In the realm of cognitive reasoning and artificial intelligence, understanding human decision-making is paramount. The concept of utility [4], which is an abstract measure of the preferences and values assigned by individuals to different scenarios or outcomes, is central to this understanding. Derived from philosophy [6, 2] and intuitive psychology [7], it evolves into a computational modeling for the weighing of pros and cons in subjective human mind.

In this essay, we delve into the complexities of learning and representing human utility within AI systems. We first start with cases and analysis from behavioral psychology, and then advance to computational techniques. Drawing insights from a range of academic literature, including studies on preference-based reinforcement learning and the inference of goals from actions, we hope to explore the nature of human utility and its implications for AI development.

2 Understanding human utility

It is amazing that humans perceive the concept of utility since a very young age. Psychologists carry out experiments with infants and explore how they understand the goals of others and infer their value based on the actions' costs [8].

The underlying computational framework for this line of research is called "naive utility calculus" [7], which rests on three nested assumptions. First, agents act to maximize the utility U under constraints. Second, $U(S, A) = R(S) - C(A)$, which means that utility separates into rewards and costs. Third, the cost of an action is jointly determined by the agent and the situation.

Based on these assumptions, the authors conducted three experiments, each using different physical challenges (height, width, and incline of paths) to represent action costs. A total number of 80 ten-month-old infants participated. In each experiment, infants observed an agent choosing between two goals, each associated with different costs. Bayesian inferences and utility-theoretic calculations are used to model infants' expectations about the agent's preferences.

The results showed that infants expected the agent to prefer the goal achieved through costlier actions. Their expectations held across different types of physical challenges, suggesting they inferred this through an abstract understanding of concepts like "force", "work", or "effort". This further suggests that infants use cost and reward as interconnected abstract variables in understanding actions, which supports the view that a system grounded in cost-reward trade-offs guides action understanding from an early age.

The above study on infants is only a small piece of the abundant research conducted by psychologists, economists and sociologists. An utilitarian view can be adopted to capture optimal social choice functions [3], to establish the mechanism of fairness, competition and cooperation [5], and extend to animal activities similarly [9]. The proliferation in this field demonstrates the expressiveness of utility, which may be critical in endowing AI systems with human-like decision-making capabilities.

3 Computational modeling methods

In this section, we go over some of the prominent ways for learning and representing human utility. We will introduce the methods briefly and discuss their advantages and disadvantages in data collection, generalization, and efficiency. Hopefully this will provide an view on what stage we are at in computational modeling of human utility.

3.1 Challenges and obstacles

Although analysis from behavioural studies provides an insight into the intuitive psychology of human utility, successfully modeling it still faces complicated challenges. The request involves understanding how humans assign value to different outcomes and actions. Since those utility functions are internal to humans and vary between individuals, it is difficult to collect extensive and effective data. Moreover, as the behavioural experiments suggest, the perception of cost and reward is abstract, therefore lacking a meaningful unit of measure for utility.

Reinforcement learning (RL) is a representative case where the optimization of reward function, similar to quantitative utility, guides the learning process and behaviour of agents [10]. One obstacle associated with traditional reinforcement learning is designing a reward function, which often requires significant task-specific prior knowledge. Moreover, since the learning process and learned policy could be sensitive to small changes of the reward, the choice of the reward function may have a crucial impact on the success and must be handled with caution. Tasks such as robotics require a lot of reward engineering, facing problems such as reward hacking [1], reward shaping [11], infinite rewards [13], and multi-objective trade-offs.

3.2 Proposed strategies

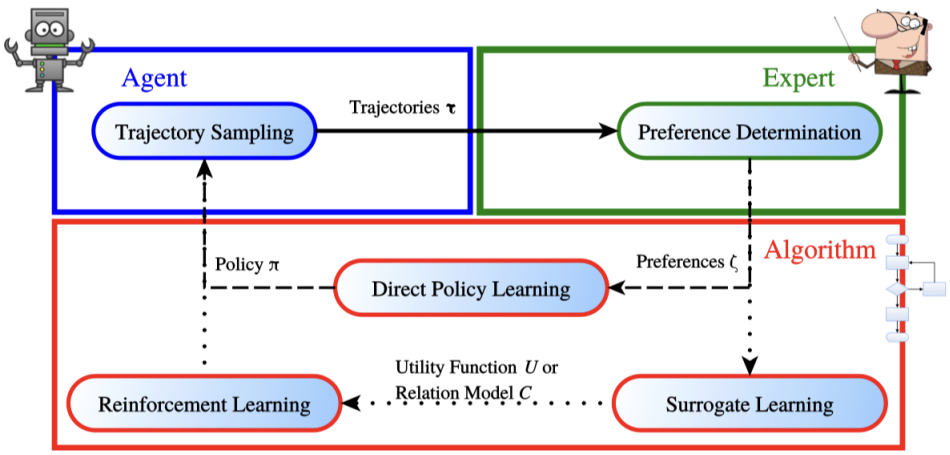

Figure 1: PbRL: Learning policies from preferences via direct (dashed path) and surrogatebased (dotted path) approaches [12]

Preference-based Reinforcement Learning (PbRL), different from standard RL, is proposed to directly learn from an expert's preferences rather than a hand-designed numeric reward. This strategy deals with the reward shaping problem without the dependence on prior expert knowledge.

The workflow of preference-based reinforcement learning and its difference from standard reinforcement learning is illustrate in Figure 1. The algorithm involves two actors: an agent which acts according to a given policy and an expert which evaluate the agent's behavior. There are three categories of approaches from the literature, including learning a policy, learning a preference model, and learning a utility function. These approaches work in different scenarios and complement each other in subtleties.

The design principles of preference-based reinforcement learning include the following aspects:

- Types of feedback: The type of feedback that assumed includes action preferences, state preferences, and trajectory preferences. A major issue for this aspect is how to distinguish between short-term optimality and long-term ones. The former two poses high demands for the expert actor, while the latter requires the discrimination of states or actions responsible for the encountered preferences. Both are difficult problems to tackle.

- Defining the learning problem: As mentioned before, the learning approaches involve options like learning a policy, a preference model, or a utility function. Derict learning of policy is to find a parametrization which maximizes the correspondence with the observed preferences in the parametric policy space. It requires a deliberate approximation of policy distribution and comparison / ranking method. The learning of preference model and utility function, on the other hand, can be used to derive a policy and therefore work in an indirect way.

- Temporal credit assignment problem: This problem is to determine which states or actions are responsible for the obtained preference. Explicit solutions give birth to different types of inferred utility functions, such as value-based, return-based, and reward based.

There are also other techniques, such as trajectory preference elicitation, policy optimization, and modeling of the transition dynamics. Detailed discussions will be omitted due to the limited scope.

3.3 Strengths and weaknesses

The PbRL method has been successfully applied to many practical tasks. Most of the applications rely in robot teaching and board game domains, since they are subject to some of the most sophisticated reward shaping problems.

The strengths of PbRL method can be viewed in several aspects. First, just as its original purpose, PbRL reduces the need for extensive domain knowledge to define rewards. Second, it is particularly effective in environments with complex or unknown reward structures, and can utilize qualitative, non-numeric rewards, making it suitable for scenarios where numeric feedback is unavailable or difficult to quantify. Third, PbRL's ability to incorporate human preferences allows for more intuitive and interactive training processes, especially in settings where human expertise is valuable but hard to encode in standard reward functions. Last, some PbRL methods are reported to be capable of generalizing effectively from a relatively small set of preference data, making it resource efficient in data collection.

Despite the promising capabilities, PbRL also suffer from apparent shortcomings. For example, Pareto-optimal policies cannot be achieved by any of the utility-based approaches in case of incomparabilities. More catastrophic problems stand such as the struggle with high-dimensional policy spaces, hence the scalability issue, posing challenges for practical implementation in large-scale problems. In addition, current PbRL methods are not equipped to perform risk-averse optimization, which is essential in scenarios where certain outcomes must be avoided (e.g., robot safety). Sample efficiency is also a concern, considering that it can be impractical to obtain human evaluators under many circumstances.

There are still many future explorations to be done in this field. Finding principled ways for combining various types of feedback from the human expert and efficient methods for exploring the preference function space are still needed. A unified evaluation framework is also called for in order to compare different algorithms effectively.

4 Conclusion

In this essay, we have examined the critical role of utility and its link to human / AI decision-making. We observed how human utility understanding, evident even in infants, is vital for the perception of goals from actions.

Among the computational methods for modeling human utility, Preference-based Reinforcement Learning (PbRL) emerged as an important component, offering an innovative way to integrate human preferences directly into AI agents.

However, challenges persist, notably in the abstract nature of human utility and the variability of individual preferences. Technical obstacles such as scaling issues and sample efficiency also exist. Despite these hurdles, the exploration of human utility in AI is crucial for the development of aligned AI systems. Hopefully one day there will be a significant stride in the evolution.

References

[1] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*, 2016. 2

[2] Jeremy Bentham. *The collected works of Jeremy Bentham: An introduction to the principles of morals and legislation*. Clarendon Press, 1996. 1

[3] Craig Boutilier, Ioannis Caragiannis, Simi Haber, Tyler Lu, Ariel D Procaccia, and Or Sheffet. Optimal social choice functions: A utilitarian view. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 197–214, 2012. 2

[4] Rachael A Briggs. Normative theories of rational choice: Expected utility. 2014. 1

[5] Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868, 1999. 2

[6] Francis Hutcheson. *An inquiry into the original of our ideas of beauty and virtue: in two treatises*. R. Ware, 1753. 1

[7] Julian Jara-Ettinger, Hyowon Gweon, Laura E Schulz, and Joshua B Tenenbaum. The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in cognitive sciences*, 20(8):589–604, 2016. 1

[8] Shari Liu, Tomer D Ullman, Joshua B Tenenbaum, and Elizabeth S Spelke. Ten-month-old infants infer the value of goals from the costs of actions. *Science*, 358(6366):1038–1041, 2017. 1

[9] Alicia P Melis, Brian Hare, and Michael Tomasello. Chimpanzees recruit the best collaborators. *Science*, 311(5765):1297–1300, 2006. 2

[10] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015. 2

[11] Andrew Y Ng, Daishi Harada, and Stuart Russell. Policy invariance under reward transformations: Theory and application to reward shaping. In *Icml*, volume 99, pages 278–287. Citeseer, 1999. 2

[12] Christian Wirth, Riad Akrour, Gerhard Neumann, Johannes Fürnkranz, et al. A survey of preference-based reinforcement learning methods. *Journal of Machine Learning Research*, 18 (136):1–46, 2017. 2

[13] Yufan Zhao, Michael R Kosorok, and Donglin Zeng. Reinforcement learning design for cancer clinical trials. *Statistics in medicine*, 28(26):3294–3315, 2009. 2