Value-based Decision System: A Top-Down Reflection

Jingze Zhang
Department of Automation
Tsinghua University
jz-zhang21@mails.tsinghua.edu.cn

Abstract

The exploration of *Utility* and *Human Value* is an inherently challenging endeavor. The task of appropriately representing value in computers and aligning the value systems of machines with human values has been widely studied in recent years. However, through a review of the literature, we discern that this representation of value is still in its infancy. Currently, computational human value can only address relatively straightforward tasks associated with human preferences, and there is room for improvement in the effectiveness of these solutions.

Moreover, when faced with intricate value judgments and decision-making problems, there is currently no unified model to address such complexities. In this essay, we will focus on rethinking the forms of representation for human value and value from a top-down perspective. Following the *U-V theory*, we will explore what constitutes a meaningful representation of Utility and Human Value, considering among social value attributes, individual needs, and survival requirements.

Figure 1: Building intelligent systems aligned with human values. The intelligent robot *Baymax* in the *Big Hero 6* is what we want artificial intelligence to be like in the future.

1 Introduction

Human decision-making operates across multiple dimensions. On a societal level, entities ranging from governments and businesses to households need to make decisions in policy or production based on the values of the social community. On an individual level, there is a dual requirement: individuals need a personal value system that aligns with the societal value framework to fulfill the value-related needs of integration into society. Simultaneously, they also need a fundamental judgment and decision-making system to meet basic survival needs. Decisions at various levels are intertwined with complex value attributes, forming the foundation of human societal production and life.

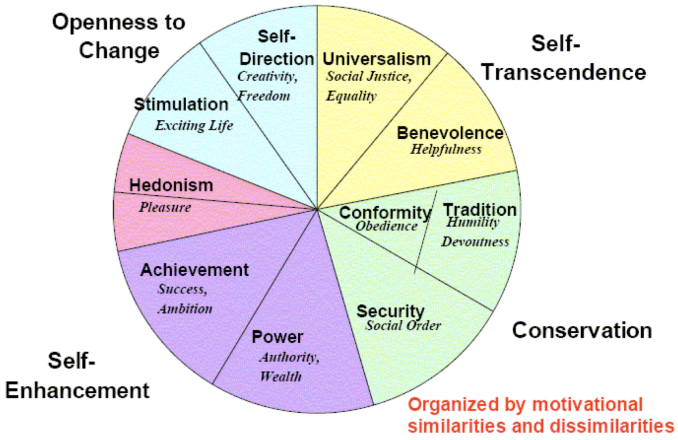

Figure 2: Theoretical model of relations among ten motivational types of values. The image is obtained from [6].

As for a single human or even a primate, the dimension of value is a diverse problem. Many of us decision is powered by curiosity, self-enhancement or other attributes. However, these attributes of human value is hard to be modeled in a computational system. The interaction of human with its environment is not suitable to represent as a markov decision process or MDP in many cases[5], while the internal value of human or a social community is also hard to be abstracted as a score or a reward, mainly due to the multi-dimension attribute of human value system[6].

In this essay, we first review some current explorations of Computational Value System. These explorations about computational value system mainly focus on the learning of agents' preference in the decision process[2, 8]. Furthermore, we will summarize the current research about decision system in human brain structure[3, 4]. Finally, we will propose some ideas and insights about human value system.

2 Early Exploration of Computational Value System

In many cases, human makes choices based on their preference, and the preference among different states and different actions can be quantified into a utility function $U(s, a)$. A basic definition of the *utility function* was provided in [10], where preferences for state and action can be modeled as a *Utility Function* that follows a partial order relationship.

With a utility function, a computational model can conduct rational decisions in different scenario. But how can we obtain the utility function after observing other agents' actions and inferring their preference?

2.1 Preference-based Reinforcement Learning

Preference-based Reinforcement Learning or PbRL is a framework to model the utility and learn the preference of the expert. A brief definition of PbRL is depicted in [8] as follows:

> *Preference-based reinforcement learning (PbRL) is a paradigm for learning from non- numerical feedback in sequential domains. Its key idea is that the*

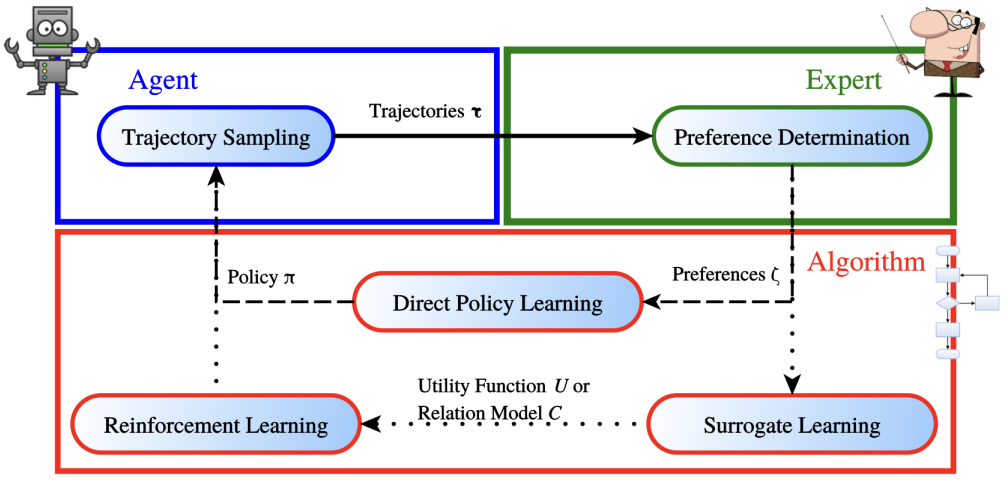

Figure 3: PbRL: Learning policies from preferences via direct (dashed path) and surrogate- based (dotted path) approaches. The image is obtained from [8].

> *requirement for a numerical feedback signal is replaced with the assumption of a preference-based feedback signal that indicates relative instead of absolute utility values.*

By the tool of preference-base RL, agents can infer the experts' preference among different choices. It can solve 3 most challenging drawbacks in traditional RL algorithms, which are Reward Hacking[1], Multi-objective trade-off, and infinite rewards.

2.2 Model Utility via Scored-based Model

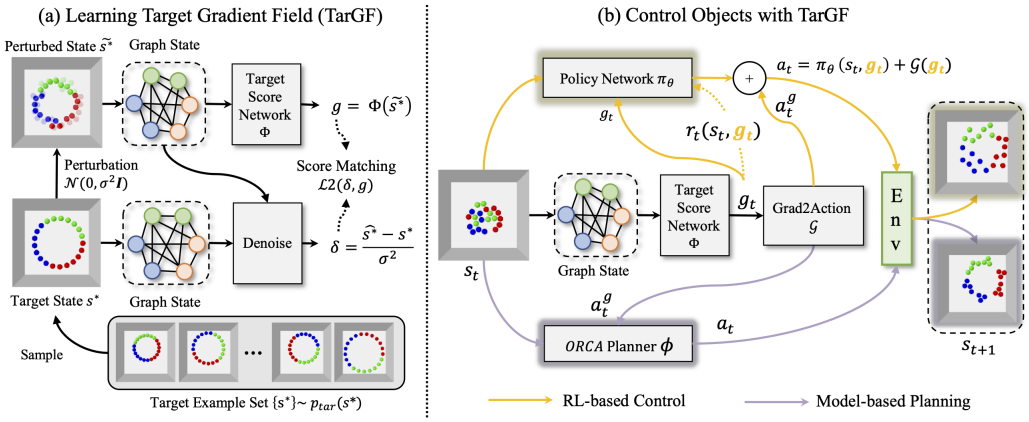

Figure 4: The pipeline of *Target Gradient Field*. The image is obtained from [9].

The intuition of *Utility Function* is similar to the energy-based model and score-based model. Take the room clean-up tasks as an example, we can model the utility function implicitly via a scored-based model, where the score function is a log gradient of the probabilistic function.

$$s_\theta(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} \log p_\theta(\boldsymbol{x}) \tag{1}$$

The cleaning process of the room is similar to stochastic gradient descent (SGD) and the level of cleanliness can be viewed as a utility function or the probabilistic function. The pipeline of the score-based utility method is showcased in Figure 3. The method was first proposed by [9]

Furthermore, we argue that the preference of a user can be represented by a latent feature and the feature can served as a condition of the score-based diffusion model. In such settings, the preference

of a user may be able to learn in several iterations in a few-shot way. We will explore the idea in the future research, during our course project.

The limitation of the scored-based utility function method is an implicit model in modeling the human value or human preference. However, the limitations of the method is also a high time consuming and non-robust inference. The generalizable feature of the model in real world scenario hasn't been fully investigated either.

3 Discussion

Based on the literature review, we know that the development of computation value model is still in its infancy. As for real world settings, the preference-based choosing of us humans can be classified into 3 categories[7].

1. Value-first decision making.
2. Comparison-based decision making with value computation.
3. Comparison-based decision making without value computation.

These methods of decision making can both be found in the decision making process in our brains, where the second and third method are widely obtained in our brains. However, only the first method is widely explored in the computational utility system and we argue that the second and third computing paradigms may be another breakthrough point for next generation decision making system.

We also argue that some metrics should be put forward to evaluate whether the agents' utility and value are in alignment with humans. And Furthermore, looking backward to figure 1, the value system of human and society is an open-vocab and high-dim system, and it is hard to evaluate the system only a simple scalar. This direction is especially challenging.

4 Summary

In this essay, we explores challenges in representing human values via computational systems, advocating a top-down approach following the U-V theory. We notes limitations in current computational models, emphasizing the need for sophistication in addressing complex human preferences. And we calls for metrics to evaluate alignment with human values.

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. 3

[2] 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

[3] Philippe Domenech, Jérôme Redouté, Etienne Koechlin, and Jean-Claude Dreher. The neuro-computational architecture of value-based selection in the human brain. *Cerebral Cortex*, 28(2): 585–601, 2018. 2

[4] Marios G Philiastides, Guido Biele, and Hauke R Heekeren. A mechanistic account of value computation in the human brain. *Proceedings of the National Academy of Sciences*, 107(20): 9430–9435, 2010. 2

[5] Aoyang Qin, Feng Gao, Qing Li, Song-Chun Zhu, and Sirui Xie. Learning non-markovian decision-making from state-only sequences. *arXiv preprint arXiv:2306.15156*, 2023. 2

[6] Shalom H Schwartz. Basic human values: An overview. 2006. 2

[7] Ivo Vlaev, Nick Chater, Neil Stewart, and Gordon DA Brown. Does the brain calculate value? *Trends in cognitive sciences*, 15(11):546–554, 2011. 4

[8] 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, 3

[9] Mingdong Wu, Fangwei Zhong, Yulong Xia, and Hao Dong. Targf: Learning target gradient field to rearrange objects without explicit goal specification. *Advances in Neural Information Processing Systems*, 35:31986–31999, 2022. 3

[10] 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. 2