The Representations of Human Utility

Hengli Li
Yuanpei College
Peking University
2000017754@stu.pku.edu.cn

Abstract

This paper explores the concept of utility in decision-making, rooted in human cognition and social choices. It introduces two methodologies for acquiring human utility: leveraging visual information and affordances, and applying deep reinforcement learning from human preferences. The first method enhances understanding through visual cues but faces challenges in generalizability. The second method, though innovative, encounters difficulties in obtaining preference data.

1 Introduction

The concept of utility emerges in the realm of human cognition [1] and bears conceptual resemblance to affordances [3]. In the context of decision-making and social choice, utility pertains to the subjective satisfaction individuals derive from the consequences of their choices. This foundational concept finds its roots in diverse academic disciplines, encompassing economics, ethics, and artificial intelligence. Within the domain of artificial intelligence and decision theory, a comprehensive examination of utility functions is imperative for the development of systems capable of executing rational and socially optimal decisions. Boutilier's scholarly contribution to social choice functions [1] intricately explores the dynamics of decision-making in a social milieu, wherein the preferences and utilities of multiple agents are considered. The comprehension of utility assumes pivotal significance in the optimization of social outcomes, thereby harmonizing the diverse needs and aspirations of individuals within a collective framework.

However, the concept of utility lacks a unified representation, thereby rendering its predictability challenging across diverse situations. An examination of the research conducted by Hassanin et al. [3] accentuates the substantial influence of environmental visual features on human behavioral patterns. The affordances emanating from objects within our surroundings exert a noteworthy impact on the utility derived from our actions. This intricate interplay between visual affordances and utility contributes to a nuanced understanding of decision-making processes, underscoring the complex interrelationship among perception, cognition, and social choices. Furthermore, Liu et al. [4] demonstrates that the influence of environmental visual features on human utility extends to the early stages of cognitive development. The study, focused on ten-month-old infants, reveals that even in the absence of advanced verbal communication skills, infants can deduce the value of goals from the associated costs of actions. This implies that foundational elements of utility, such as the assessment of outcome desirability relative to the effort or cost involved in achieving them, may be ingrained in human cognition from a very early age. The implications of these findings align with a broader understanding of utility as a multifaceted concept. Utility encompasses not only rational decision-making based on explicit preferences but also implicit and often subconscious evaluations of outcomes and costs.

Hence, this paper introduces two prospective methodologies for the acquisition and representation of human utility. Additionally, we will succinctly analyze their merits and shortcomings concerning data collection, generalization, and efficiency:

1. Leveraging Visual Information and Affordances
2. Adopting Deep Reinforcement Learning from Human Preferences

2 Visual Information and Affordances

This approach entails the incorporation of visual information and affordances in the process of learning and representing human utility. The utilization of environmental visual cues and the examination of affordances presented by objects are intended to offer insights into the nuanced dimensions of decision-making processes. The systematic collection of such data assumes a pivotal role in comprehending how individuals perceive and engage with their surroundings, particularly in the realm of goal inference. Extending this exploration into the domain of machine learning and computer vision, the research on inferring forces and learning human utilities from videos by Wang et al. [5] augments our understanding. Through the utilization of visual information, algorithms can discern the underlying forces and utilities associated with human actions. This not only enhances our comprehension of decision-making processes but also paves the way for the development of artificial intelligence systems capable of learning and emulating human preferences based on visual cues.

While the integration of visual information and affordances enhances our understanding of decision-making processes, addressing challenges in data collection is imperative. The diverse origins of visual data give rise to considerations related to consent, privacy, and potential biases inherent in the collected information. Achieving a delicate balance between the depth of data richness and the ethical ramifications of its acquisition is pivotal for upholding responsible research practices. Notably, in comparison to the subsequent proposed method, the accessibility of data in this approach is relatively more straightforward. Beyond the realm of data collection, the generalization of acquired utilities from visual data introduces complexities in ensuring the adaptability of AI systems to diverse and dynamic real-world scenarios. The presence of biases in data collection, whether inadvertent or systemic, poses a significant risk of distorting representations of human preferences. This distortion, in turn, can have consequential impacts on the performance and fairness of AI models in practical applications. Researchers and practitioners must remain vigilant in the ongoing effort to mitigate biases and actively promote diversity in data sources to enhance the overall generalizability of learned utilities.

3 Deep Reinforcement Learning from Human Preferences

This methodology involves the application of deep reinforcement learning techniques to glean insights from human preferences and, consequently, to learn utility [2]. By assimilating knowledge from human choices and preferences, this approach endeavors to encapsulate the nuanced facets of utility in decision-making. However, a notable drawback arises in this context, primarily pertaining to the formidable challenge of data collection. The inclusion of humans in the loop is not only economically demanding but also time-intensive, while the acquisition of preference data remains concealed and arduous to procure in real-life scenarios. As for generalizability, the effectiveness of deep reinforcement learning from human preferences hinges on the ability of the model to extrapolate insights from limited datasets to broader contexts. Achieving robust generalization is a complex task, as human preferences can be highly context-dependent and subject to change over time. Nonetheless, recent advancements in neural network architectures and training methodologies have shown promise in enhancing the generalization capabilities of these models. Despite the challenges, the integration of deep reinforcement learning with human preferences represents a pioneering approach that has the potential to significantly improve the adaptability and robustness of decision-making systems.

4 Conclusion

In conclusion, this paper thoroughly investigates two distinct methodologies designed to elucidate and depict human utility in decision-making processes: leveraging visual information and affordances, and adopting deep reinforcement learning from human preferences. Each approach manifests its distinctive array of advantages and challenges. While the inclusion of visual cues contributes to a deeper understanding, it reveals limitations in terms of generalizability. Conversely, deep reinforcement learning from human preferences, although innovative, confronts impediments in the realm of data collection.

References

[1] Craig Boutilier, Ioannis Caragiannis, Simi Haber, Tyler Lu, Ariel D. Procaccia, and Or Sheffet. Optimal social choice functions: A utilitarian view. *Artificial Intelligence*, 227:190–213, 2015. ISSN 0004-3702. doi: https://doi.org/10.1016/j.artint.2015.06.003. URL `https://www.sciencedirect.com/science/article/pii/S0004370215000892`. 1

[2] Paul Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences, 2023. 2

[3] Mohammed Hassanin, Salman Khan, and Murat Tahtali. Visual affordance and function understanding: A survey, 2018. 1

[4] 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. doi: 10.1126/science.aag2132. URL `https://www.science.org/doi/abs/10.1126/science.aag2132`. 1

[5] Yixin Zhu, Chenfanfu Jiang, Yibiao Zhao, Demetri Terzopoulos, and Song-Chun Zhu. Inferring forces and learning human utilities from videos. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3823–3833, 2016. doi: 10.1109/CVPR.2016.415. 2